

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Zuversicht im Mathematikunterricht –
Confidence-Based Learning und Assessment“

verfasst von / submitted by

Emily Oberhauser, BEd

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree
of

Master of Education (MEd)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 199 500 520 02

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Lehramt Sek (AB)
UF Bewegung und Sport UF Mathematik

Betreut von / Supervisor:

Mag. Mag. Dr. Olivia Vrabl

„Sometimes in maths, ‘I don’t know’
is exactly the right answer [...]“

Colin Foster, 2017

Kurzfassung

Confidence-Based Learning und Assessment wurden im Schulunterricht bislang sehr spärlich erforscht. Die vorliegende Studie beinhaltet eine dreiwöchige Intervention in einer Klasse der 7. Schulstufe (N=23), während dieser Schüler*innen Aufgaben im Mathematikunterricht nicht nur lösen, sondern auch angeben mussten, wie zuversichtlich sie sind, dass ihre Antwort richtig gelöst wurde. Dafür wurde eigens ein Instrument – Zuversichtsemojis – entwickelt. Ziel dieser Studie ist es herauszufinden wie Schüler*innen auf den Einbezug der Zuversicht in Form von Zuversichtsemojis reagieren und welcher Zusammenhang zwischen Zuversicht und Kompetenz im Bereich Prozentrechnung besteht. Die statistische Auswertung ergab, dass sich die Schüler*innen im Bereich der Prozentrechnung gut einschätzten und kein statistisch signifikanter Geschlechterunterschied bestand. Die Meinungen der Schüler*innen gingen auseinander, wobei zu erkennen war, dass sie den Nutzen des Einbezugs der Zuversicht in den Unterricht nicht erkannten. Einige Schüler*innen gaben an, sie könnten dadurch Wissenslücken erkennen und lernten sich besser einzuschätzen.

Schlüsselwörter: Confidence-Based Assessment, Confidence-Based Learning, Zuversicht, Prozentrechnung, Fachdidaktik, Mathematik

Abstract

Confidence-Based Learning and Assessment has been very sparsely researched in schools. This study involves a three-week intervention in a 7th grade class (N=23) in which students not only had to solve problems in math class, but also had to indicate how confident they were that their answer was solved correctly. A tool – confidence emojis – was specially developed for this purpose. The aim of this study is to find out how students react to the inclusion of confidence in the form of confidence emojis and what the connection is between confidence and competence in the calculation of percentages. The statistical evaluation showed that the students assessed themselves well in percentage calculation and that there was no statistically significant gender difference. Students' opinions varied, showing that they did not see the benefit of including confidence in the classroom. Some students stated that they could identify gaps in their knowledge and learn to assess themselves better.

Keywords: Confidence-Based Assessment, Confidence-Based Learning, Confidence, Calculation of Percentages, Subject-Specific Didactics, Mathematics

Inhaltsverzeichnis

1	Einleitung.....	1
2	Confidence-Based Learning.....	4
2.1	Begriffserklärung	4
2.2	Zusammenhang von Zuversicht und Wissen	4
2.3	Entwicklung von Confidence-Based Learning.....	8
2.4	Zuversicht und Gender.....	10
2.5	Zuversicht und Wissenserhalt	13
2.6	Prozess.....	15
2.7	Bedeutung für die Schule	16
3	Confidence-Based Assessment.....	18
3.1	Confidence-Based Assessment vs. Confidence-Based Marking	18
3.2	Angabe der Zuversicht	19
3.2.1	Skala.....	19
3.2.2	Zeitpunkt	20
3.3	Kalibrierung	21
3.4	Confidence-Based Marking	22
3.5	Vorteile von Confidence-Based Assessment und Marking	28
3.6	Confidence-Based Assessment im Mathematikunterricht	32
3.7	Einsatzmöglichkeiten im schulischen Kontext.....	38
4	Selbsteinschätzung	41
4.1	Selbstkonzept	41
4.1.1	Entwicklung der Selbstkonzeptforschung	41
4.1.2	Akademisches Selbstkonzept	43
4.2	Selbstüberschätzung.....	54
4.3	Selbstunterschätzung.....	56
4.4	Dunning-Kruger Effekt	58
4.5	Selbsteinschätzung in der Schule	59
5	Methodik	62
5.1	Setting und Teilnehmer*innen.....	62
5.2	Entwicklung Zuversichtsemojis	63
5.3	Studiendesign	66
5.3.1	Einführung der Zuversicht und der Zuversichtsemojis.....	67
5.3.2	Einbezug der Zuversichtsemojis in den Unterricht	68
5.3.3	Lernzielkontrolle	69
5.3.4	Evaluierung mittels Fragebogen	70

5.4	Feldnotizen	72
5.5	Auswertung der Daten	73
6	Ergebnisse	74
6.1	Feldnotizen und Ergebnisse der Schüler*innen	74
6.1.1	Einführung	74
6.1.2	Einbezug der Zuversichtsemojis in den Unterricht	79
6.2	Ergebnisse der Lernzielkontrolle	83
6.2.1	Richtigkeit und Zuversicht	83
6.2.2	Geschlechterunterschied	86
6.3	Ergebnisse der Evaluierung	87
7	Diskussion	94
7.1	Forschungsfrage 1	94
7.2	Forschungsfrage 2	96
8	Fazit und Ausblick	98
9	Zusammenfassung	99
10	Literaturverzeichnis	102
11	Abbildungsverzeichnis	114
12	Tabellenverzeichnis	116
13	Anhang	117
13.1	Anhang 1: Lernzielkontrolle	117

1 Einleitung

Schlagzeilen wie *Mädchen unterschätzen sich in Mathematik* und *Buben überschätzen sich*, finden sich immer wieder in den Medien. Auch die Autorin dieser Arbeit konnte in eigenen Unterrichtssituationen als Lehrerin beobachten, dass Schüler*innen sich immer wieder falsch einschätzen – sei es überschätzen oder unterschätzen. Die Beobachtungen reichten von Schüler*innen, die ganz überzeugt von einer Antwort waren, die leider falsch war, bis zu jenen, die es kaum wagten, ihre Antwort zu sagen, obwohl sie richtig war. Klar ist, dass beides ein Problem darstellt. Confidence-Based Assessment, der Einbezug von Zuversicht in Wissensüberprüfungen, bietet nun die Möglichkeit für Lehrpersonen und Schüler*innen zu erkennen, worin sich Schüler*innen gut einschätzen können und worin nicht. Ebenso können dadurch Wissenslücken und auch jenes Wissen, das fälschlicherweise als richtig wahrgenommen wird, identifiziert werden. Durch den Einbezug von Zuversicht in den Unterricht können aber nicht nur Lehrpersonen profitieren, sondern auch Schüler*innen. Es wurde festgestellt, dass es bereits Studien zu Confidence-Based Assessment gibt, jene beziehen sich aber meistens auf den Hochschulbereich mit Fokus auf Medizin. Gerade für den schulischen Bereich gibt es kaum Studien zu dieser Thematik. Vor allem fehlen aber Studien, bei denen Confidence-Based Assessment in den Unterricht über einen längeren Zeitraum implementiert wird. Jene Studien, die in Schulen durchgeführt wurden, beinhalteten lediglich das Implementieren von Confidence-Based Assessment bei einer Überprüfung für die Studie selbst. Bis dato gibt es keine Arbeit, in der die bisherigen Forschungen zu Confidence-Based Learning, Confidence-Based Assessment und Confidence-Based Marking gesammelt dargestellt und diskutiert wurden. Ebenso gab es auch keinerlei wirkliche Trennungen zwischen diesen Begriffen. Ein Versuch dessen wird in dieser Arbeit gestartet. Weiters gibt es vor allem im deutschsprachigen Raum keinerlei Literatur zu diesen Themen.

Zuversicht und Mathematikunterricht sind vermutlich zwei Wörter, die viele nicht sofort miteinander assoziieren würden. Mithilfe dieser Masterarbeit soll die Rolle von Zuversicht im Mathematikunterricht und generell im Unterricht ins Rampenlicht gerückt werden. Es ist ein Thema mit so viel Potenzial, das leider im Kontext der Schule, aber auch der Universitäten bislang viel zu wenig Aufmerksamkeit bekommen hat. Jedoch ist es gerade in einem nun leider angstbesetzten Fach wie Mathematik wichtig Zuversicht zu integrieren und somit Schüler*innen einerseits in ihrem Lernen zu fördern und andererseits diese auch bei möglicher Prüfungsangst zu unterstützen.

Mithilfe der theoretischen Darstellung und der im Rahmen dieser Arbeit durchgeführten Studie möchte die Autorin das Interesse an dieser Thematik wecken. Im weiteren Sinne möchte sie

Lehrer*innen dazu ermutigen, die Zuversicht in den eigenen Unterricht zu implementieren. Wie bereits erwähnt, gibt es zwar bereits Studien (z.B. Foster, 2016), bei denen die Zuversicht an Schulen erforscht wurde. Was jedoch fehlt, sind Studien von Lehrer*innen, die die Zuversicht aktiv in ihren Unterricht einbauen und im Rahmen dessen forschen. Es besteht die Vermutung, dass Confidence-Based Assessment aufgrund des Fehlens solcher Studien und der wenigen Publikationen mit Schulbezug bislang keinen Anklang an Schulen gefunden hat.

Aufgrund der geringen Forschungslage und hohen Bedeutung für den Unterricht ist diese Arbeit von hoher Relevanz. Die Forscherin stellte für diese Arbeit zuerst Literatur über Confidence-Based Assessment und ausgewählte angrenzende Gebiete zusammen dar, bevor sie sich mit der Beantwortung ihrer Forschungsfragen auseinandersetzte. Um dieses Ziel zu erreichen, wurde an einer Wiener Mittelschule eine dreiwöchige Studie durchgeführt, bei der die Zuversicht mithilfe von eigens dafür entwickelten Zuversichtsemojis in den Unterricht implementiert wurde. In dieser Zeit wurden Beobachtungen und Aussagen der Schüler*innen in Form von Feldnotizen festgehalten. Ebenso wurde eine Lernzielkontrolle durchgeführt sowie eine Evaluierung am Ende der dreiwöchigen Intervention.

Ziel dieser Masterarbeit war es, einerseits das Thema theoretisch darzulegen und andererseits mittels einer dreiwöchigen Intervention in einer Wiener Mittelschule folgende Forschungsfragen zu beantworten:

- Wie reagieren Schüler*innen auf eine dreiwöchige Intervention mit Zuversichtsemojis?
- Wie ist der Zusammenhang zwischen Zuversicht und Kompetenz von Schüler*innen im Bereich Prozentrechnung? Wie gut sind diese darin „kalibriert“?
 - Haben Schülerinnen anders angekreuzt als Schüler?¹

Die vorliegende Arbeit gliedert sich in mehrere Kapitel. Anfangs erfolgt die theoretische Fundierung, gefolgt von der empirischen Studie. Zu Beginn werden in Kapitel 2 theoretische Grundlagen des Confidence-Based Learning geklärt, bevor im Anschluss in Kapitel 3 Confidence-Based Assessment mit den verschiedenen Skalen und Einsatzmöglichkeiten im Mathematikunterricht nähergebracht und auch kritisch beleuchtet werden. Kapitel 4 befasst sich mit der zugrunde liegenden Selbsteinschätzung und dem Selbstkonzept. Stets wurde dabei versucht einen Bogen zur Schule zu spannen. Die darauffolgenden Kapitel behandeln die empirische Studie. In Kapitel 5 wird zunächst die Methodik dargestellt, die in dieser Studie verwendet wurde. Kapitel 6 liefert die

¹ Personen, die dem weiblichen/männlichen Geschlecht zugeschrieben werden

Darstellung der Ergebnisse, die in Kapitel 7 aus dem Blickwinkel der Forschungsfragen diskutiert werden. Das Fazit und ein Ausblick werden in Kapitel 8 gegeben. Zum Schluss findet sich in Kapitel 9 eine Zusammenfassung dieser Arbeit.

Vor dem Lesen dieser Masterarbeit muss auf eine Thematik hingewiesen werden. Die Forscherin möchte besonders betonen, dass es sich in den folgenden Kapiteln niemals um statische Zuschreibungen, sondern stets um einen momentanen Ist-Zustand handelt. Entscheiden sich Schüler*innen für eine Stufe der Zuversicht, dann muss immer klar sein, dass es sich dabei um eine Momentaufnahme mit dem Anspruch der Weiterentwicklung handelt. Es darf dabei keinesfalls zu Abstempelungen kommen (z.B. Person ist zuversichtlich oder nicht zuversichtlich), da die Zuversicht abhängig von der Aufgabe und der Thematik und demnach beispielsweise bei der Bruchrechnung anders sein kann als bei der Prozentrechnung. Es wird daher dringend darum gebeten dies während des Lesens dieser Masterarbeit im Hinterkopf zu behalten.

2 Confidence-Based Learning

2.1 Begriffserklärung

Confidence-Based Learning beschreibt ein Lehr- und Lernsystem, bei dem nicht nur das Wissen der Lernenden bewertet, sondern auch die Zuversicht bei der Erfüllung einer Aufgabe berücksichtigt wird. Dieses Lehr- und Lernsystem wurde ursprünglich als „Information Reference Testing“ bezeichnet, kurz IRT und geht auf die Forschung von Dr. James Bruno zurück.

Zuallererst wird der Begriff *confidence* geklärt, der laut dem Cambridge Dictionary wie folgt definiert wird: „the quality of being certain of your abilities or of having trust in people, plans, or the future“. Im Deutschen gibt es für das Wort *confidence* allerdings mehrere Übersetzungen, die zum Teil unterschiedliche Bedeutungen haben. Einerseits wird es mit Vertrauen, Zutrauen, aber auch mit Zuversicht übersetzt. In folgender Arbeit wird der Begriff „confidence“ mit Zuversicht übersetzt. Zuversicht wiederum wird definiert als: „festes Vertrauen auf eine positive Entwicklung in der Zukunft, auf die Erfüllung bestimmter Wünsche und Hoffnungen“ (Dudenredaktion, o.D.). Ist man zuversichtlich, hat man also festes Vertrauen auf eine positive Entwicklung im Jetzt. Im schulischen Kontext könnte dies bedeuten, dass man sich einer Antwort bei einer Aufgabe sehr sicher ist. Der*die Schüler*in ist also zuversichtlich, dass die gegebene Antwort richtig ist. Ist der*die Schüler*in nicht zuversichtlich, würde dies im Gegenteil nun bedeuten, dass die Person sich nicht sicher ist, ob die gegebene Antwort richtig ist. Wie sich im folgenden Kapitel zeigen wird, besteht ein Zusammenhang zwischen Wissen und Zuversicht, im weiteren Sinne auch zum Handeln einer Person.

2.2 Zusammenhang von Zuversicht und Wissen

Das Konzept von Wissen und der Versuch dieses empirisch zu messen, wurde schon seit Jahrtausenden diskutiert. Schon vor Christus äußerte sich Konfuzius zu dem Konzept von Wissen: "To know what you know and what you do not know, that is true knowledge."

In den vergangenen Jahren beschäftigen sich vor allem Wissenschaftler wie Darwin Hunt und James Bruno mit dem Einfluss von Zuversicht auf Wissen. Ihre Meinungen stimmen darin überein, dass Wissen mehr ist als nur das Wiedergeben korrekter Informationen. (Hunt, 2003; Bruno, 1995; Bruno; 1993)

Laut Bruno (1987) werden falsche Antworten bei Richtig/Falsch-Fragen einfach als Unwissenheit der Person interpretiert. Viel schwerwiegender ist allerdings, wenn diese Person die falsche Antwort wählt, allerdings fest davon überzeugt ist, dass es die richtige Antwort ist. Bruno

bezeichnet das als *confidently held misinformation*. Erkennbar ist dies auch in vielen derzeitigen verwendeten Überprüfungsmethoden. Eine Person, die eine Frage richtig beantwortet und die Antwort auch wirklich weiß, kann nicht unterschieden werden von einer Person, die die Frage durch Raten richtig beantwortet hat. Die Leistung beider Personen ist demnach also sehr unterschiedlich, auch wenn ihre Ergebnisse identisch sind. (Bruno, 1993; Hunt, 2003)

Auch Gvozdenko & Chambers (2007) halten in ihrem Artikel fest, dass sehr wohl unterschieden werden muss zwischen Personen, die richtig antworten und sich sehr sicher über ihre Antwort sind und denjenigen, die zwar richtig antworten, sich aber sehr unsicher sind oder sogar geraten haben. „On a regular test this difference is not evident to the learner, who will not realise the real status of his/her knowledge and whose further learning may be jeopardised by flawed foundations.“ (Gvozdenko & Chambers, 2007, S.206) Im Gegensatz zu Bruno (1993) und Hunt (2003) beziehen sich Gvozdenko & Chambers auch auf die Seite der Lernenden. Es geht dabei nicht um die Ergebnisse eines Tests per se, sondern viel mehr um das Lernen der Personen an sich. Eine Person ist nach Hunt und Furustig (1989) *unusably informed*, wenn sie sich unsicher ist über richtiges Wissen. Ist sie sich sicher über richtiges Wissen, so wird diese als *usably informed* bezeichnet.

Hunt (1993) führt diesen Ansatz weiter und erläutert, dass es *usable and unusable knowledge*, also nützliches beziehungsweise unnützes Wissen gibt. Voraussetzung für nützliches Wissen ist, dass die Antwort auch richtig gegeben werden kann. (Hunt, 2003) Wissen ist laut Hunt (1993, 2003) nur dann nützlich, wenn die Person wirklich zuversichtlich ist, dass dieses Wissen auch richtig ist. Nur dann wird dieses Wissen auch bei Entscheidungen oder Handlungen verwendet. Im Gegenzug ist jenes Wissen unnütz, das nicht mit Zuversicht genutzt wird, um Entscheidungen zu treffen. In Abbildung 2.1 wird dieses Prinzip anhand eines Beispiels verdeutlicht. Die Abbildung ist jeweils von oben nach unten zu lesen. Ein möglicher Pfad wird nun beispielhaft angeführt. Um einen Brand zu löschen, muss die Taste A gedrückt werden. Denkt eine Person nun, dass sie die Taste E drücken muss, ist das ein falscher Glaube. Im Anschluss wird unterschieden, ob sich die Person sicher ist oder nicht, was dazu führt, dass es entweder eine nützliche oder unnütze Annahme ist. Im Weiteren lässt sich daraus schließen, ob die Person falsch informiert oder uninformiert ist.

Criterion: Press Key A to extinguish an electrical fire.			
Person's belief is that Key E should be pressed to extinguish an electrical fire		Person's belief is that Key A should be pressed to extinguish an electrical fire	
Incorrect belief		Correct belief	
Sure of Correctness	Unsure		Sure of Correctness
Usable belief	Unusable belief		KNOWLEDGE
Misinformed	Uninformed	Guess or Partially Informed	Well Informed

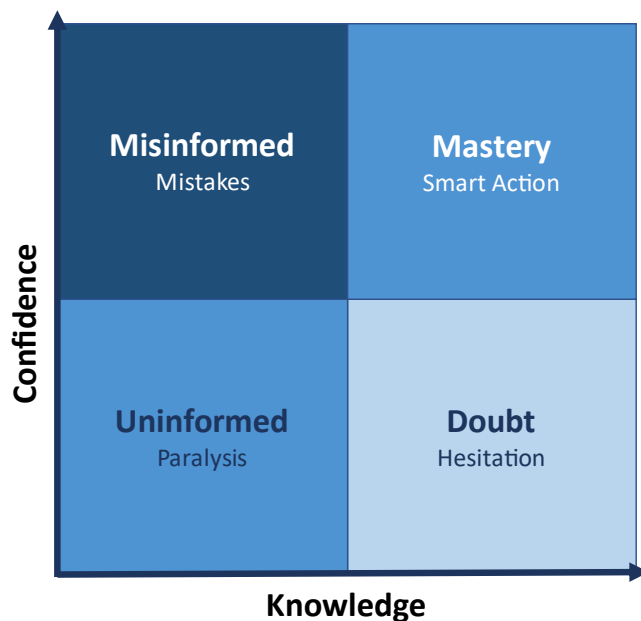
Abbildung 2.1

Zusammenhang von Richtigkeit, Zuversicht, dass eine Antwort richtig ist und Nützlichkeit des Wissens (Hunt, 2003)

Das Ziel in schulischen und hochschulischen Kontexten ist es nun, dass Lernende Wissen erwerben und behalten, das einerseits korrekt ist und andererseits auch nützlich ist. Wissen, das nützlich, aber inkorrekt ist, soll erkannt und behoben werden. Dies ist laut Hunt (1993) durch das Einbeziehen von der Angabe der eigenen Zuversicht auf gegebene Antworten bei Überprüfungen und Bewertungen möglich. Ohne einer Begründung über die gewählte Antwortmöglichkeit oder den Grad der Zuversicht, dass die Antwort auch wirklich richtig ist, ist es nicht möglich, fehlerhaftes Wissen oder Antworten, die basierend auf reinem Raten gegeben wurden, zu identifizieren. (Snow, 2019)

Bruno (1987) zog neben Wissen und Zuversicht auch das Verhalten hinzu. Wissen allein ist notwendig, aber nicht genug, um Handlungen zu tätigen. Personen werden erst durch die Verbindung von Wissen und Zuversicht zu Verhalten und Handlungen befähigt. Bruno behauptet, dass Personen, die korrektes Wissen und hohe Zuversicht haben, produktives Verhalten und ebenso produktive Handlungen setzen. Im Gegenteil können negative Handlungen von Personen entstehen, die zuversichtlich bei gleichzeitig inkorrektem Wissen sind.

Basierend auf dieser Annahme entstand das sogenannte *Information Reference Testing* (IRT). Es ist der Vorreiter des Confidence-Based Assessment. Dabei handelt es sich um eine zweidimensionale Überprüfungs- beziehungsweise Bewertungsmethode. Die zwei Dimensionen sind Zuversicht und Wissen. Daraus entwickelte Bruno die in Abbildung 2.2 ersichtlichen „knowledge-behavior quadrants“ des „Learning Behavior Model“. (Bruno, 1987)

**Abbildung 2.2**

Knowledge quadrants (modifiziert nach Bruno, 1987)

Die vier Quadranten von Bruno (1987) machen jeweils den Zusammenhang von Wissen und Zuversicht sichtbar. Dabei wird das Wissen entsprechend der Qualität der zugrundeliegenden Information in die Quadranten aufgeteilt. Die vier Quadranten lassen sich somit wie folgt beschreiben:

- *Misinformed*: Hierbei handelt es sich um Wissen, das eine lernende Person mit hoher Zuversicht für richtig hält, aber eigentlich falsch ist. Diejenigen, die sehr zuversichtlich sind, dass Fehlinformationen richtig sind, werden sehr wahrscheinlich Fehler machen. Um dies in den schulischen Kontext zu bringen, soll an dieser Stelle ein Beispiel gebracht werden. Ein*e Schüler*in ist überzeugt davon, dass man Strichrechnungen vor Punktrechnungen rechnen muss. Die Aufgabe „ $3 \cdot 2 + 5 = ?$ “ würde der*die Schüler*in also sehr zuversichtlich mit der Antwort 21 beantworten, wobei die richtige Antwort klarerweise 11 lautet. Das ist ein klassisches Beispiel für Fehlinformationen.
- *Uninformed*: In diese Kategorie fällt Wissen, das die lernende Person noch nicht erworben hat. Diese Person wird aufgrund dessen auch nicht handeln, wodurch es zu einem Lähmungszustand („Paralysis“) kommen kann.
- *Doubt* (frei übersetzt: Zweifel): Wissen zählt zu dieser Kategorie, wenn die lernende Person denkt, dass dieses richtig ist, aber auch kleine Zweifel diesbezüglich bestehen. Aus diesem Grund wird die Person zögern und eventuell nicht basierend auf diesem Wissen handeln.

- *Mastery*: Eine Person ist sehr zuversichtlich, dass ihr Wissen richtig ist. Im Vergleich zur Fehlinformation ist dieses Wissen hier auch tatsächlich richtig. Eine Person, die richtiges Wissen und auch noch ein hohes Maß an Zuversicht in dieses Wissen hat, wird sehr wahrscheinlich schlau und sehr produktiv handeln und dabei wenige Fehler machen.

Tabelle 2.1 zeigt den Zusammenhang zwischen Wissen und Zuversicht und den vier Wissens-Quadranten des Learning Behavior Model. (Hunt & Furustig, 1989)

		Knowledge	
		No	Yes
Confidence	No	Uninformed	Doubt
	Yes	Misinformed	Mastery

Tabelle 2.1

Zusammenhang zwischen Wissen und Zuversicht (in Anlehnung an Hunt & Furustig, 1989)

Das Ziel ist es, dass Lernende über viel Wissen verfügen, das dem Quadranten „Mastery“ zuzuordnen ist. Nur Personen, die wirklich zuversichtlich über korrektes Wissen sind, werden auch produktive Handlungen tätigen. (Bruno, 1993)

Folgendes Zitat soll dieses Kapitel kurz resümieren und unterstreicht dabei die Wichtigkeit dessen, dass Fehlinformationen aufgedeckt werden müssen:

It is commonly accepted that people behave in accordance with their knowledge and beliefs. The more certain the knowledge or belief then the more likely, more rapid and more reliable is the response. If a person strongly believes something to be correct which is, in fact, incorrect, then the performance of tasks that rely on this erroneous belief or misinformation may likewise be in error—even though the response may be executed with confidence. Thus, from an educational and training point of view it is important to detect and identify misinformation. (Hunt, 2003, S.107)

Bislang ging es um den Zusammenhang zwischen Zuversicht und Wissen, das folgende Kapitel befasst sich nun mit der Entwicklung von Confidence-Based Learning.

2.3 Entwicklung von Confidence-Based Learning

Vorweg sei gesagt, dass der Begriff Confidence-Based Learning in der Literatur nicht allzu viel Aufmerksamkeit bekommt. Meist werden die Begriffe Confidence-Based Learning und Confidence-

Based Assessment synonym verwendet, wodurch eine klare Unterscheidung anhand der Literatur nicht möglich ist. Im Zuge der Recherchen stellte sich heraus, dass es sehr wenig Literatur über den eigentlichen Prozess von Confidence-Based Learning gibt (siehe Kapitel 2.6). Da die beiden Begriffe ineinander verschwimmen, ist es auch in der Gliederung dieser Arbeit schwierig, eine klare Linie zu ziehen. In dem Kapitel Confidence-Based Learning liegt daher der Fokus auf der Entwicklung dieser Methode, dem Zusammenhang zu anderen Themen und der Bedeutung in der Schule. Das Kapitel Confidence-Based Assessment wird im Anschluss mehr auf die Überprüfungsmethode an sich und die damit einhergehenden unterschiedlichen Bewertungsmethoden eingehen.

Lange Zeit lag der Hauptfokus von Studien zu diesem Thema auf der Reliabilität und Validität von Bewertungen, die auf Zuversicht basieren. Erst später wurde diese Methode zum Wissenserwerb genutzt. Das erste Mal wurden die Randbedingungen von Confidence-Based Learning im Jahre 1932 in einer akademischen Publikation von Kate Hevner publiziert. Diese Publikation behandelt eine Bewertungsmethode für Richtig/Falsch-Fragen.

Ein großes Problem von Richtig/Falsch-Fragen ist, dass man sehr viele Punkte schon durch pures Raten erhalten kann. Es wurden viele Bewertungsmethoden getestet, um diesem Effekt entgegenzuwirken. Die Richtig-Minus-Falsch-Bewertungsmethode erzielte hierbei beispielsweise keine zufriedenstellende Verbesserung in Bezug auf Reliabilität und Validität von Tests. (Paterson & Langlie, 1925)

Bei der Studie von Hevner (1932) wurde nun auch der Grad der Zuversicht berücksichtigt. Es wurden zwei verschiedene Tests verwendet. Einerseits der Meier-Seashor Art Judgment-Test (Seashore, 1929) und andererseits ein Test zur ästhetischen Wertschätzung von Musik (music appreciation test) von Hevner (1930).

Die für psychologische Tests vorgesehene Reliabilität wird von diesen beiden Tests leicht unterschritten. Die Personen hören bei diesen Tests jeweils zwei sehr ähnliche Musikstücke und müssen angeben, welches sie bevorzugen und ebenso den Grad ihrer Zuversicht angeben. Es wurden vier verschiedene Bewertungsmethoden, mit dem Ziel eine Erhöhung der Reliabilität zu erreichen, empirisch untersucht. Die erste Methode nennt sich *Number Right*. Hierbei wird die Summe der richtigen Antworten gezählt. *Right-minus-Wrong* ist die zweite Methode, bei der die Anzahl der Fehler von der Anzahl der richtigen Antworten subtrahiert wird. Die dritte Methode ist *Weighted Right Answers*. Hier kommt nun die Zuversicht ins Spiel. Richtige Antworten werden entsprechend des Grades der Zuversicht gewertet. Wird A – entspricht *sehr sicher* - beim Grad der Zuversicht angekreuzt, so erhält die Person drei Punkte, bei B („ziemlich sicher“) zwei Punkte und bei C („überhaupt nicht sicher“) einen Punkt. Die vierte und letzte Bewertungsmethode ist *Weighted Right*

Minus Weighted Wrong. Es werden hier richtige Antworten mit hoher Zuversicht belohnt, aber zwei Punkte abgezogen, falls die Antwort falsch ist und A bei der Zuversicht angegeben wurde und ein Punkt, falls B angekreuzt wurde. Für eine falsche Antwort mit C wurde kein Punkt abgezogen. (Hevner, 1932)

Diese erste Studie mit dem *music appreciation test* wurde 1931 an der University of Minnesota durchgeführt. Es stellte sich heraus, dass die Methode *Weighted Right Answers* eine Begünstigung für Reliabilität und Validität mit sich bringt. (Hevner, 1932)

Die zweite Studie mit beiden oben genannten Tests wurde ebenso im Jahre 1931 durchgeführt. Auch hier lässt sich eine Begünstigung vor allem der Reliabilität bei der Methode *Weighted Right Answers* erkennen. Diese Art von Untersuchungen sind grundsätzlich für alle Tests ohne Zeitbeschränkung oder Tests mit Zeitbeschränkung, bei denen zusätzliche Zeit für die Angabe der Zuversicht gegeben wird, nützlich. Laut Hevner ist also das Messen von Zuversicht und von Wissen ein besserer Indikator für die tatsächliche Leistung der Lernenden, als wenn nur das Wissen gemessen wird. (Hevner, 1932)

Die Entwicklung von Confidence-Based Learning basiert auf jahrelanger Forschung. Maßgebend dafür war besonders die Forschung von Darwin Hunt, Dieudonne LeClerq, Emir Shuford und James E. Bruno. Dr. Darwin Hunt, ein US-amerikanischer Psychologe auf der New Mexico State University, führte bedeutende Forschungen, unterstützt von der US Navy durch. Seine Forschungen bezogen sich auf die verschiedenen Dimensionen von Wissen verbunden mit Zuversicht und Korrektheit des Wissens. Weiters erforschte er auch den Zusammenhang von Zuversicht, Wissen und Wissenserhalt. In den folgenden Kapiteln werden einige dieser Forschungen präsentiert.

2.4 Zuversicht und Gender

In einer Studie im Jahre 1994 untersuchten Hassmén und Hunt, ob die Selbsteinschätzung bezüglich der Korrektheit einer Antwort die Anzahl der richtigen Antworten beeinflusst. Weiters wollten sie ergründen, ob die Unterschiede zwischen den Ergebnissen von Frauen und Männern durch den Einbezug der Selbsteinschätzung beeinflusst werden. Bei dieser Studie nahmen je 60 Studenten und Studentinnen, die sich im ersten Semester ihres Studiums befanden, teil. Diese wurden in vier Subgruppen unterteilt. Es gab zwei Gruppen mit je 30 Studenten und zwei mit je 30 Studentinnen. Der Multiple-Choice Test enthielt 50 Items (10 mathematische Fragen und 40 verbale Fragen) und dieser wurde so genderneutral wie möglich gestaltet. Es gab jeweils 5 Antwortmöglichkeiten, wobei eine anzukreuzen war. Je eine männliche und eine weibliche

Subgruppe beantworteten die Fragen auf einem konventionellen Multiple-Choice Antwortbogen (MC), indem sie die ihrer Meinung nach richtige Antwort ankreuzten. Die beiden anderen Subgruppen hatten einen ähnlichen Antwortbogen, jedoch mussten diese auch nach jeder Frage auf einer 5-Punkte-Skala angeben, wie sicher sie sind, dass ihre Antwort richtig ist. Dieser Antwortbogen wurde in der Studie als SACAT („self-assessment answer sheet“) bezeichnet. Die Resultate der statistischen Auswertung ergaben, dass bei Frauen durch die Selbsteinschätzung eine Verbesserung gegeben war, nicht jedoch bei Männern. Bei den Subgruppen mit dem konventionellen Multiple-Choice-Antwortbogen schnitten Männer im Vergleich zu Frauen signifikant besser ab. Der Unterschied bei den richtigen Antworten zwischen Studenten und Studentinnen wurde durch den Einbezug der Selbsteinschätzung beim Antwortbogen verringert. (Hassmén & Hunt, 1994)

Es stellt sich nun die Frage, ob eine Verbesserung bei einem Test durch den Einsatz von Selbsteinschätzung nur bei Frauen erzielt werden kann. Das dem nicht so ist, zeigt eine Studie von Koivula et al. (2001), in der 494 schwedische Schüler*innen in ihrem letzten Schuljahr und 80 Psychologiestudent*innen in ihrem ersten Jahr auf der Stockholm University teilnahmen. Die Leistung der Teilnehmer*innen beim *Swedish Scholastic Aptitude Test*, ein Test, der zur Zulassung zur Hochschulbildung herangezogen wird, wurde verglichen. Dieser besteht aus mehreren Untertests. Für diese Studie wurde ein Untertest zum Wortschatz (WORD) und ein Test zu *quantitative reasoning* (DS) herangezogen. Es wurden einerseits konventionelle Multiple-Choice Tests verwendet, andererseits jene mit Einbezug der Selbsteinschätzung (SACAT).

Wie in Tabelle 2.2 ersichtlich, zeigen die Ergebnisse nur kleine Unterschiede im Subtest *WORD*. Betrachtet man die Ergebnisse des Subtests *DS*, so lässt sich hier erkennen, dass diejenigen, die die Selbsteinschätzung angaben, signifikant höhere durchschnittliche Punktzahlen erreichten als jene die einen konventionellen Multiple-Choice Test machten. Ebenso profitieren jene durch die Selbsteinschätzung, die sich selbst – gemessen an einem Geschlechtsrolleninventar - bei traditionell als männlich angesehenen Merkmalen schlecht einschätzten.

	Word	DS	Word + DS
Women (n=317)	56,0 (18,5)	41,6 (16,6)	49,8 (14,9)
MC (n=176)	54,2 (19,1)	39,8 (16,3)	48,1 (15,2)
SA (n=141)	58,2 (17,5)	43,8 (16,8)	52,1 (14,2)
Men (n=233)	57,7 (17,9)	45,7 (17,9)	52,6 (15,3)
MC (n=105)	58,0 (18,5)	41,8 (18,3)	51,1 (16,1)
SA (n=128)	57,4 (17,5)	48,9 (17,0)	53,8 (14,5)

Tabelle 2.2

durchschnittlicher Prozentsatz richtiger Antworten (und zugehörige Standardabweichungen) bei Multiple-Choice Tests (MC) und jenen mit Selbsteinschätzung (SA) in den Subtests WORD und DS für Frauen und Männer (modifiziert nach Koivula et al, 2001)

Anhand der signifikant höheren Ergebnisse der Selbsteinschätzung bei Frauen, lässt sich schließen, dass diese viel exakter bei der Selbsteinschätzung sind als Männer. (Koivula et al., 2001) Dies bestätigen auch die früheren Forschungsergebnisse von Lundeberg et al. (1994). Lundeberg et al. sammelten Daten bei Psychologiekursen von 251 Student*innen. Diese mussten nach dem Beantworten einzelner Items jeweils ihre Zuversicht über die Korrektheit der Antwort geben. Ziel dieser Studie war es, geschlechtsspezifische Unterschiede bei der Angabe der eigenen Zuversicht der Korrektheit der gegebenen Antwort zu untersuchen. Es zeigte sich, dass Geschlechtsunterschiede einerseits vom abgeprüften Gebiet und andererseits davon abhängen, ob die Items richtig oder falsch beantwortet wurden. Bei Themen wie Mathematik waren Männer deutlich zuversichtlicher als Frauen. Besonders auffallend ist, dass Teilnehmer*innen auch bei falschen Antworten eine hohe Zuversicht angaben. Besonders bei Studenten konnte nachgewiesen werden, dass sie sich bei falschen Antworten überschätzten. Allerdings ergaben die Ergebnisse auch, dass sowohl Männer als auch Frauen sich überschätzten. Aus einer späteren Studie von Hassmén et al. (2002) kann man jedoch folgern, dass Frauen weniger dazu neigen, sich zu überschätzen als Männer.

In einer weiteren Studie von Hassmén et al. im Jahr 2002 wurde ähnliches erforscht. Es konnte gezeigt werden, dass Männer bei einer richtigen Antwort eher dazu tendierten bei der Angabe der Zuversicht „Extremely sure“ zu wählen als Frauen. Jedoch gaben Männer auch bei falschen Antworten häufiger an, dass sie „Extremely sure“ wären. Zweifellos ist die Wahl des höchsten Grades an Zuversicht positiv, wenn die Antwort tatsächlich richtig ist. Ist sie das nicht, so würde diese Person nach der Terminologie von Hunt und Furustig (1989) als „misinformed“ bezeichnet werden. Eine

weitere Erkenntnis dieser Studie ist, dass Frauen besser darin sind sich richtig einzuschätzen und, wie bereits erwähnt, weniger dazu tendieren sich zu überschätzen.

Zusammenfassend lässt sich sagen, dass das Geschlecht eine wichtige Rolle bei der Selbsteinschätzung spielt. Zu diesem Thema gibt es noch wesentlich mehr Studien. Da diese Thematik allerdings nicht der zentrale Fokus dieser Arbeit ist, wird hier nicht noch weiter ins Detail gegangen und kein Anspruch auf Vollständigkeit gestellt.

2.5 Zuversicht und Wissenserhalt²

Cabigon (1993) untersuchte im Rahmen ihrer Masterarbeit unter der Betreuung von Dr. Darwin P. Hunt, wie der Grad an Zuversicht mit dem Wissenserhalt zusammenhängt. Es stellte sich heraus, dass bei korrekt beantworteten Fragen, bei denen der*die Lernende den niedrigsten Grad an Zuversicht („Not sure at all“) wählte, eine Woche später nur noch 25% dieser Fragen richtig beantwortet wurden. Bei „Very unsure“ und „Somewhat unsure“ blieben jeweils circa 75% erhalten, bei „Extremely Sure“ sogar mehr als 90% (siehe Tabelle 2.3).

Grad an Zuversicht	Prozentsatz, der eine Woche später korrekt erhalten blieb
Not sure at all	25,0
Very unsure	75,0
Somewhat unsure	75,7
Very sure	87,6
Extremely sure	91,2

Tabelle 2.3

Prozentsatz, der eine Woche später noch korrekt beantworteten Fragen abhängig vom Grad an Zuversicht (in Anlehnung an Cabigon, 1993, S.28)

Webb et al. (1994) konnte in seiner Studie ähnliche Tendenzen aus den Daten ablesen. In dieser Studie wurde der Einfluss von sofortigem und verzögertem Feedback auf das Lernen von allgemeinen Informationen bei Multiple-Choice Fragen untersucht. Die Proband*innen beantworteten 90 Fragen

² freie Übersetzung der Autorin, Originalbegriff: Retention

je einmal im Abstand von einer Woche. Dies beinhaltete auch die Angabe der Zuversicht in ihre Antworten. Auch hier konnte ein positiver Zusammenhang zwischen Zuversicht und Wissenserhalt festgestellt werden. Für diejenigen Antworten, die beim ersten Test richtig waren, nahm die Wahrscheinlichkeit mit der Zuversicht zu, auch beim zweiten Test richtig beantwortet zu sein.

Hassmén et al. (2002) erforschten ebenso, ob ein Zusammenhang zwischen der Sicherheit über erlerntes Wissen und dem Wissenserhalt besteht. In ihrer Forschung stellten sie sich die Frage, ob die oben beschriebenen Ergebnisse auch über längere Zeiträume gelten. Dabei nahmen 48 Frauen und Männer der Universität Stockholm teil. Der erste Teil der Studie bestand aus einer Lerneinheit. Die Teilnehmer*innen mussten mit möglichst wenigen Versuchen acht Bilder von unterschiedlichen Zangen den entsprechenden Namen zuordnen und jeweils angeben, wie sicher sie sich über ihre Antwort sind. Diese Begriffe wurden aus einer Studie von Hunt (1982) übernommen. Der Test endete, wenn die Person in einem Durchgang alles richtig zuordnen konnte. Im Gegensatz zu Cabigon (1993) wurden hier die Proband*innen gleichmäßig in drei Gruppen aufgeteilt, die entweder nach einer Woche, nach sechs Wochen oder nach zwölf Wochen den zweiten Teil der Studie absolvieren sollten. Dieser bestand wieder aus der Zuordnung der Bilder mit den entsprechenden Bezeichnungen. Es stellte sich heraus, dass mit der Angabe „Extremely sure“ eine Woche später noch 87,9% der Abbildungen beim ersten Versuch richtig zugeordnet werden konnten. Lediglich 62% wurden bei der Angabe „Somewhat sure“ beim ersten Versuch richtig in Erinnerung behalten. Wie erwartet sank der Prozentsatz an richtigen Antworten mit der Anzahl der Wochen, die zwischen dem ersten und dem zweiten Teil der Studie lagen. In Tabelle 2.4 sind die jeweiligen Prozentsätze ersichtlich.

Grad an Zuversicht	1 Woche	6 Wochen	12 Wochen
Somewhat sure	62,0	47,3	37,0
Very sure	59,7	39,3	27,8
Extremely sure	87,9	58,6	59,0

Tabelle 2.4

durchschnittlicher Prozentsatz an richtigen Antworten nach einer bestimmten Zeit abhängig vom Grad an Zuversicht

Vergleicht man nun die Ergebnisse der Gruppe dieser Studie, die nach einer Woche erneut getestet wurde, mit der Studie von Cabigon (1993), so zeigen sich einige Ähnlichkeiten, insbesondere in der Kategorie „Extremely sure“. Bei beiden Studien fanden sich die höchsten Prozentsätze bei der

Angabe „Extremely sure“. Bei den weiteren Werten unterscheiden sich die beiden Studien allerdings. Die Studie von Hassmén et al. (2002) deutet darauf hin, dass „Extremely sure“ definitiv einen Einfluss auf den Wissenserhalt hat. Jedoch lässt sich auch erkennen, dass eine niedrigere Zuversicht einen geringeren Einfluss auf den Wissenserhalt hat. Nicht jedoch bei Cabigon (1993), in ihrer Studie steigen die Prozentsätze an richtigen Antworten monoton an. Hassmén et al. (2002) vermutet, dass dies an der Thematik des zu erlernenden Wissens liege und auch mit dem Gender zusammenhänge. Genaueres dazu lässt sich in Hassmén et al. (2002) nachlesen.

Die Forschung von Dr. James Hunt (2003) zeigte, dass der Erhalt von neu erlerntem Wissen im Zusammenhang damit steht, wie sicher sich Lernende über die Korrektheit dieses Wissens sind. Ziel ist es, dass der*die Lernende im Anschluss an eine Ausbildung oder einen Lernprozess eine Aufgabe oder Aktivität ausführen kann. Voraussetzung dafür ist die Aneignung des dafür notwendigen Wissens beziehungsweise der Fähigkeiten, und dass dieses behalten wird, um es zu einem späteren Zeitpunkt unter anderem für Handlungen und Entscheidungen abrufen zu können.

Während bislang die Entwicklung von Confidence-Based Learning und der Zusammenhang von Zuversicht mit unterschiedlichen Variablen diskutiert wurde, beschreibt das folgende Kapitel den Prozess von Confidence-Based Learning.

2.6 Prozess

Confidence-Based Learning ist in drei Phasen unterteilt. Die erste Phase beschäftigt sich mit der Diagnose. Hier wird das vorhandene Wissen mit zugehöriger Zuversicht überprüft. Dies geschieht mit Hilfe von Confidence-Based Assessment. Eine Beschreibung dazu findet sich in Kapitel 3. In der zweiten Phase werden der lernenden Person zugeschnittene Lernunterlagen oder Informationen gegeben, um ihr Wissen zu festigen oder fehlerhaftes Wissen richtigzustellen. Die letzte Phase widmet sich dem Lernprozess und im gleichen Zuge dem Aufbau von Zuversicht in das neu erworbene Wissen. (Adams & Ewen., 2009; Chatterjee & Mandal, 2016) Mit dieser Information im Hinterkopf könnte man fast sagen, die bisherigen Studien haben sich hauptsächlich mit der ersten Phase von Confidence-Based Learning beschäftigt. Aufgrund dessen ließen sich die beiden Begriffe Confidence-Based Learning und Confidence-Based Assessment trennen. Im Grunde genommen ist Confidence-Based Assessment die erste Phase von Confidence-Based Learning.

Es steht außer Frage, dass dem damit einhergehenden Lernprozess zu wenig Beachtung geschenkt wird. An dieser Stelle soll noch einmal betont werden wie wichtig – gerade aus Sicht einer

Lehrperson – die Phasen zwei und drei sind. Allein durch die Kenntnis von Fehlinformationen, kann man noch nicht viel daran ändern. Erst wenn ein*eine Schüler*in Informationen zum eigenen Wissensstand in Verbindung mit der Zuversicht oder daran angepasste Lernunterlagen hat, ist im Anschluss ein Lernprozess und in dem Sinne ein Wachstum möglich.

Wie in einem vorigen Kapitel bereits erwähnt, ist es das Ziel, über möglichst viel korrektes Wissen zu verfügen, worüber man sehr zuversichtlich ist und das somit dem Quadranten *Mastery* zugeordnet werden kann. (Bruno, 1993) Adam & Gary (2009, S.4) halten fest: „The students continue to work through the learning until they have achieved “*mastery*” in the material – 100% correct and 100% confident.“ Was bedeutet das nun für die Schule? Diese Frage wird im folgenden Kapitel thematisiert.

2.7 Bedeutung für die Schule

Confidence-Based Learning hat es zum Ziel, die Wissenslücken von Lernenden – beziehungsweise im schulischen Kontext speziell von Schüler*innen – zu ermitteln. Wie bereits zuvor erwähnt, ist es schwierig, tatsächliche Wissenslücken mit summativen Wissensüberprüfungen zu identifizieren. Wie nun bekannt ist, ist Wissen nicht nur als das Geben richtiger Antworten definiert und ebenso nicht durch eine eindimensionale Überprüfung messbar. Es muss die Zuversicht hinzugezogen werden. (Hunt, 2003; Bruno, 1993; Bruno, 1995) Die Wichtigkeit des Confidence-Based Learning wurde in den vorangehenden Kapiteln nun schon deutlich demonstriert. Es wurde auch aufgezeigt, dass Schüler*innen möglichst viel Wissen haben sollten, das zu dem Quadranten *Mastery* zählt. Hierzu zählt jenes Wissen, bei dem die Schüler*innen sehr zuversichtlich sind und das Wissen auch tatsächlich korrekt ist.

Da Lehrpersonen nun nicht einfach schätzen können, wo das Wissen der Schüler*innen zu verorten ist, ist es wichtig, sich dabei auf empirisch messbare Daten, beziehungsweise auf evidenzbasierte Forschungen zu stützen. Dieser Gedanke basiert auf dem Prinzip von *Evidence-Based Teaching*. Groccia und Bukist (2011, S.8) definieren Evidence-Based Teaching wie folgt:

[...] the conscientious, explicit, and judicious integration of best available research on teaching technique and expertise within the context of student, teacher, department, college, university, and community characteristics.

Demnach wird empfohlen, sich auf evidenzbasierte Forschungen zu beziehen und nicht auf die Tradition, Meinung erfahrener Kolleg*innen oder gar persönliche Erfahrung basierend auf Versuch und Irrtum. (Slavin, 2008). Ebenso ist dies beim Ermitteln der Wissenslücken oder

Unsicherheiten bei Schüler*innen. Bei welcher Thematik ist sich der*die Schüler*in unsicher oder sicher? Wo liegen eventuelle Fehlinformationen vor? Fragen wie diese lassen sich nicht ausschließlich durch Beobachtungen, reines Erfragen oder basierend auf Erfahrungen beantworten. Bei der Methode Confidence-Based Learning geht man davon aus, dass eben diese Wissenslücken oder Fehlinformationen durch Confidence-Based Assessment aufgedeckt werden müssen. Erst dann kann man – etwa durch individuelle Materialien – auf die Bedürfnisse jedes*jeder Einzelnen eingehen.

3 Confidence-Based Assessment

3.1 Confidence-Based Assessment vs. Confidence-Based Marking

Es ist schwierig, eine klare Linie zwischen den Begriffen Confidence-Based Assessment und Confidence-Based Marking zu ziehen. Wie bereits in Kapitel 2 wird auch hier versucht, diese voneinander abzugrenzen. Da die Begriffe in der Literatur synonym und nicht einheitlich verwendet werden, stellt sich diese Trennung als äußerst schwierig dar. Eine Möglichkeit wäre die Definition von Confidence-Based Assessment als die Implementierung von Zuversicht in formative oder summative Leistungsbeurteilungen, ohne danach zu beurteilen. Denn allein mit der Information wo Schüler*innen in Bezug auf ihr Wissen stehen und wie später noch näher erläutert wird, wie gut sie kalibriert sind, bringt schon einen enormen Mehrwert. Confidence-Based Marking wäre entsprechend diesem Versuch der Trennung zufolge jene Beurteilung, bei der die Komponente der Zuversicht in die Beurteilung hinzugezogen wird. In *allen* bisherigen Studien beinhaltet der Einbezug von Zuversicht bei Leistungsüberprüfungen aber *immer* auch den Einbezug der Zuversicht in die Beurteilung. Daher wird bei *allen* Studien – wenn man sich auf die hier versuchte Trennung bezieht - Confidence-Based Marking erforscht. Bei keiner bisherigen Studie wurde Confidence-Based Assessment eingeführt, ohne im Anschluss dementsprechend zu beurteilen. Lediglich in einer Quelle (Novacek, 2013) konnte eine gewisse ähnliche Trennung ausgemacht werden. Novacek unterteilte Confidence-Based Assessment in *Marking System* und *Grading Scheme*. Zum einen beschreibt ersteres das zwei-dimensionale Beurteilungsschema, durch das man zwei unterschiedliche aber miteinander verbundene Ergebnisse erhält. Zum anderen wird als *Grading Scheme* die tatsächliche Bepunktung bezeichnet. Seine Studie beinhaltete allerdings auch die Erforschung der Beurteilung. Entgegen all diesen Forschungen muss aus Sicht der Autorin Confidence-Based Assessment nicht gezwungenermaßen auch eine dementsprechende Beurteilung beinhalten. Ein möglicher Grund dafür, dass Confidence-Based Assessment – ohne die dementsprechende Beurteilung – noch nicht erforscht wurde, ist, dass sich der Großteil der Studien auf post-sekundäre Bildung bezieht und kaum Studien im schulischen Bereich vorhanden sind.

Da eine sinnvolle Trennung auf Basis der bisherigen Forschungslage nicht möglich ist, wird in dieser Arbeit ebenso vorgegangen und Confidence-Based Assessment und Confidence-Based Marking annähernd synonym verwendet. Im Hinterkopf sollte aber stets behalten werden, dass Confidence-Bases Marking – wie der Name schon sagt – die Beurteilung an sich beschreibt.

3.2 Angabe der Zuversicht

3.2.1 Skala

Zunächst ist auffallend, dass sich die Art der Angabe der Zuversicht in den einzelnen Studien unterscheidet. Es variiert sowohl die Anzahl der Auswahlmöglichkeiten als auch der Zeitpunkt, zu dem die Zuversicht angegeben wird. Üblicherweise werden Multiple-Choice oder Richtig/Falsch-Fragen herangezogen und neben der Beantwortung der Frage wird ebenso noch der Grad an Zuversicht in die Korrektheit der Antwort angegeben. Bei einem einfachen System gibt es hoch, mittel und niedrig als Stufen der Zuversicht. (Barr & Burke, 2013) Gardner-Medwin (1995, 2006) verwendet eine Kardinalskala, wobei ein höherer Wert einen höheren Grad an Zuversicht bedeutet. Stellvertretend werden auch Deskriptoren verwendet. Unter anderem werden hierfür Phrasen wie „Very Confident, Fairly Confident, Not Confident“ (Davies, 2002) oder „Almost guess, Probable guess, Neutral, Fairly Neutral, Almost Certain“ (Hassmén & Hunt, 1994) verwendet. Gvozdenko & Chambers (2007) ziehen eine 5-Punkte-Skala heran: „most likely wrong, probably wrong, maybe right maybe wrong, probably right, most likely right“. Abbildung 3.1 zeigt eine der Fragen, bei denen diese Skala von Gvozdenko & Chambers bei einer Studie verwendet wurde.

The screenshot displays a test interface. On the left, a sidebar shows 'Test length: 55 min', 'Test timer: 19:15', and a 'Test progress' section with a grid of question status icons (Q1-Q18). The main area contains question 15: '15. How many minutes are there between 11:15am and 4:35pm?'. Below the question is an input field with the value '320'. Navigation buttons 'Previous' and 'NEXT' are present. A note says '<. Use buttons to the left and right. -> Do NOT use browser Forward or Back Buttons!'. At the bottom, a 5-point confidence scale is shown with colored buttons and labels: 'most likely wrong' (red), 'probably wrong' (orange), 'This answer is ... maybe right maybe wrong' (yellow), 'probably right' (teal), and 'most likely right' (green). Above the scale, there is a 'Jump to' dropdown, 'Go' and 'mark' buttons, and a 'Test progress' section.

Abbildung 3.1

Interface einer Frage mit 5-Punkte-Skala (Gvozdenko & Chamber, 2007, S.207)

Da Deskriptoren von verschiedenen Personen jeweils unterschiedlich aufgefasst werden können, sprechen sich Gardner-Medwin und Gahan (2003) gegen Deskriptoren bei der Angabe der Zuversicht aus. Bei ebendiesen werden die einzelnen Stufen an Zuversicht mit den dafür vorgesehenen Punkten benannt.

Leclercq und Poumay (2005) verwenden bei ihrer Studie eine symmetrische Skala, die von 0% bis 100% geht. Diese wird in 6 Degrees of Certainty – diese Bezeichnung verwenden sie anstelle von

Confidence-Based Marking – unterteilt: „0%, 20%, 40%, 60%, 80%, 100%“ (S.2) Bezieht man sich hier auf die Unterteilung von Wissen nach Hunt (1993) (*misinformed, uninformed, informed*), so könnte man inkorrekte Antworten mit der Zuversicht von 60%, 80% oder 100% als *misinformed knowledge* bezeichnen. Leclercq und Poumay (2005) bezeichnen dieses Wissen als „dangerous knowledge“ (S.3). Richtige Fragen mit einem Grad an Zuversicht von 60%, 80% und 100% entspricht *informed knowledge*. Dieses Wissen lässt sich nach Hunt (1993) auch als *usable knowledge* und alles zwischen den beiden Extrema als *unusable knowledge* bezeichnen.

Eine weitere Skala tauchte in einem Artikel von Foster et al. (2022) auf. Diese führten eine Studie mit 6-16-jährigen Schüler*innen aus Großbritannien durch. Die Schüler*innen mussten dabei in Mathematik ihre Zuversicht auf einer 5-stufigen Likert-Skala, einer graduellen Antwortskala angeben. In Kontrast zu den bisherigen Skalen wurden hier für die Angabe der Zuversicht *confidence emojis* verwendet (siehe Abbildung 3.2).

How confident are you in your answer?



Abbildung 3.2

Angabe der Zuversicht mittels „Confidence emojis“ (modifiziert nach Foster et al., 2022, S. 493)

Es fällt auf, dass bei vielen Studien zu dieser Thematik eine ungerade Anzahl an Skalenpunkten gewählt wurde. Dadurch wird eine Mittelkategorie angeboten, die abhängig von der Polarität der Ratingskala unterschiedliches bedeuten kann. Die Forschung ist sich bezüglich der Wahl von geraden oder ungeraden Skalenpunkten nicht einig. Beides hat Vor- und Nachteile, die in Kapitel 5.2 genauer beleuchtet werden.

3.2.2 Zeitpunkt

Hunt untersuchte in einer Studie im Jahr 1982, ob die Angabe der Selbsteinschätzung das Lernen beeinflusst. Bei dieser Studie waren 90 Studenten und 90 Studentinnen beteiligt. Die Proband*innen mussten jeweils entweder unmittelbar vor oder nach jeder Antwort ihre Selbsteinschätzung zum Grad ihrer Zuversicht in Bezug auf die Richtigkeit ihrer Antwort angeben, indem sie auf unterschiedliche der Zuversicht entsprechenden Schaltflächen drückten. Bei den einzelnen Gruppen wurde sowohl in der Anzahl der möglichen Schaltflächen für die Angabe der Zuversicht (zwei, vier oder acht) unterschieden, als auch im Zeitpunkt der Angabe (vor oder nach der

Antwort). Eine einzige Gruppe erledigte ausschließlich die Lernaufgabe ohne der Angabe der Zuversicht. Es wurde jeweils die Anzahl der notwendigen Versuche gezählt, bis die Proband*innen die Namen von acht Zangen beherrschten und die Ergebnisse wurden im Anschluss verglichen. Hunt (1982) kommt zu dem Ergebnis, dass der Wissenserwerb durch die Angabe der Zuversicht um 25% schneller war. Darüber hinaus konnte ein schnellerer Wissenserwerb erzielt werden, wenn die Angabe der Zuversicht nach dem Beantworten stattfand. Die Daten deuten auch darauf hin, dass eine 8-Punkte-Skala das Lernen positiver beeinflusst als eine 2-Punkte-Skala.

3.3 Kalibrierung

Wie bereits in einem vorigen Kapitel beschrieben, wird bei Confidence-Based Assessment der Fokus auf die Zuversicht einer Person auf eine Antwort zu einem bestimmten Thema gelegt. Es liegt demnach nahe, dass Zuversicht und Kompetenz zusammenhängen. Foster (2015) formuliert dies wie folgt: „It is natural to see the variable confidence and competence as defining a two-dimensional space [...]“ (S. 274) Abbildung 3.3 zeigt den von ihm genannten zwei-dimensionalen Raum.

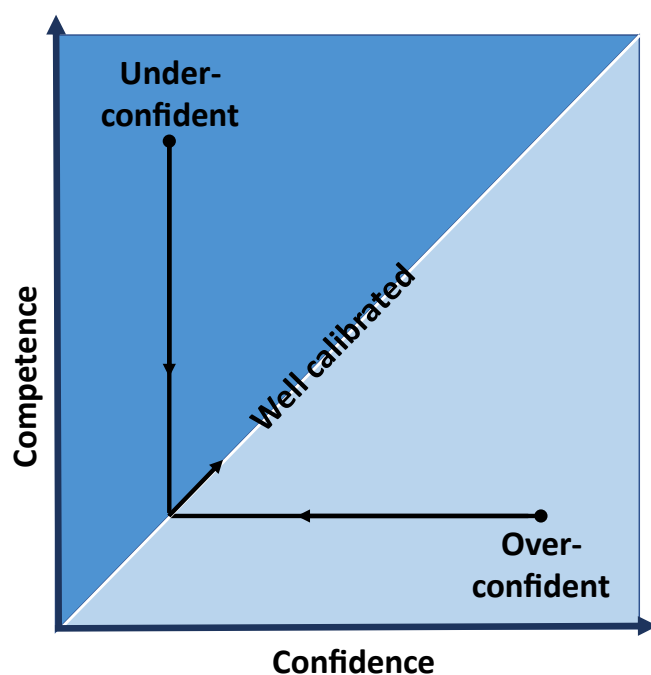


Abbildung 3.3

Zusammenhang zwischen Zuversicht und Kompetenz (modifiziert nach Foster, 2016, S.274)

Unter anderem ist in dieser Abbildung die 1. Mediane eingezeichnet, die mit *Well calibrated* beschriftet wurde. Gut-kalibrierte Schüler*innen sind in dieser Abbildung auf ebendieser 1. Mediane

zu verorten. Es stellt dar, dass ihre Zuversicht und ihre Kompetenz genau zusammenpassen. Ein*eine Schüler*in ist demnach *well calibrated*, „[...]if, over the long run, for all propositions assigned a given probability, the proportion that is true is equal to the probability assigned[...]“ (Fischhoff et al., 1977, S. 552) Zum einen befinden sich diejenigen Schüler*innen, die dazu tendieren sich zu überschätzen unterhalb der ersten Mediane – diese Schüler*innen werden als *over-confident* bezeichnet. Zum anderen sind Schüler*innen, die sich unterschätzen, und als *under-confident* bezeichnet werden, oberhalb der ersten Mediane zu verorten. Sowohl für summative als auch für formative Zwecke ist es wichtig zu wissen, wo sich die Schüler*innen in Abbildung 3.3 horizontal beziehungsweise vertikal befinden. (Foster, 2016) Je nachdem, wo sich Schüler*innen befinden, wählen Lehrpersonen unterschiedliche Möglichkeiten, um zu intervenieren. Bei zwei Schüler*innen mit ähnlicher Kompetenz, aber unterschiedlicher Zuversicht würden Lehrpersonen demnach unterschiedlich agieren. Ein*eine Schüler*in mit hoher, aber unangebrachter Zuversicht, die in Abbildung 3.3 rechts unten zu verorten wäre, wird von einem kognitiven Konflikt lernen, durch den Fehlinformationen aufgeklärt werden können. Der*die andere Schüler*in ist *under-confident*, hat eine hohe Zuversicht, aber niedrige Kompetenz (links oben in Abbildung 3.3). Diese Person könnte man dazu ermutigen, wieder daran zu arbeiten, worin er*sie sehr zuversichtlich ist, um von da an Zuversicht aufzubauen. (Foster, 2015) Der Autor hält dazu schließlich fest: „In both cases, movement towards better calibration is desired, but this might be effected in quite different ways.“ (S. 275)

3.4 Confidence-Based Marking

Confidence-Based Marking (CBM) ist eine Bewertungsmethode, die die Angabe der Zuversicht in der Beurteilung berücksichtigt. Oftmals findet sich auch der Begriff Certainty-Based Marking oder Confidence-Based Scoring in der Literatur, wobei diese synonym verwendet werden. Wie bereits in Kapitel 2 thematisiert, wird in dieser Arbeit der Begriff Confidence verwendet. Es gibt diesbezüglich jedoch kontroverse Meinungen. Gardner-Medwin (o.D., 2019) verwendet seit 2005 den Begriff Certainty anstelle von Confidence, da er meint, dieser wäre weniger missverständlich. Laut Gardner-Medwin (2019a) nahmen viele Personen an, dass es bei Confidence-Based Marking darum geht Selbstbewusstsein zu stärken oder zu belohnen. Dem stellt der Forscher folgende Aussage gegenüber:

Its aim is much more nuanced - to reward the identification of *uncertainty* as well as genuinely reliable conclusions. It may even diminish unwarranted self-confidence, though knowledge of what one does and doesn't understand is ultimately the basis for confident decision making. (Gardner-Medwin, 2019a, S. 1)

Confidence-Based Marking wurde vor allem zu dem Zeitpunkt erforscht, in dem computergestützte Beurteilung noch nicht derart praktikabel war. (Gardner-Medwin, 2006; Ahlgren, 1969) Im Bereich der Sozialpsychologie und der Didaktik taucht diese Thematik seit den 1930ern auf. Der Großteil der Studien bezieht sich allerdings auf Hochschulbildung in den Bereichen Medizin und Psychologie. Seit 2001 wird Confidence-Based Marking im ersten und zweiten Jahr des Medizinstudiums am University College London (UCL) verwendet. Ursprünglich wurde laut Gardner-Medwin und Gahan (2003) Confidence-Based Assessment am UCL eingeführt, um den Studierenden dabei zu helfen zu erkennen, wo sie sich auf einer Skala („knowledge, uncertainty, ignorance, misconception, delusion“, S.1) befinden, die unterschiedliche Level an Zuversicht bezüglich einer richtigen Aussage angibt. In Abbildung 3.4 wird diese Skala dargestellt. Ebenso galt das Ziel, dadurch formatives self-assessment zu verbessern und die Studierenden dazu zu bringen, genauer über ihre Antworten nachzudenken. (Gardner-Medwin, 2006)

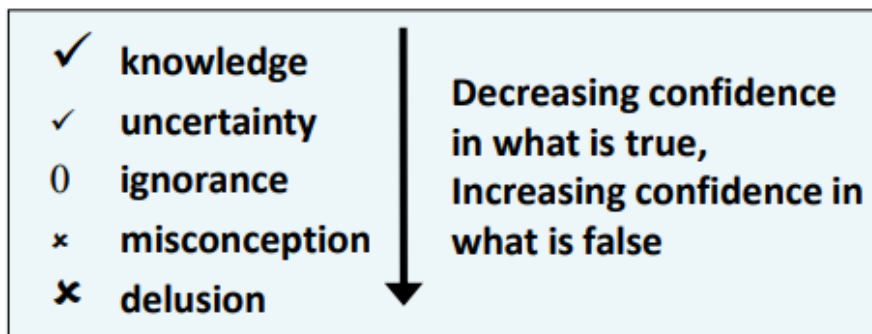


Abbildung 3.4

Unterschiedliche Level an Zuversicht über eine korrekte Aussage (Gardner-Medwin, 2019b, S.12)

In der Forschung zum Thema Confidence-Based Assessment/Marking tauchen unterschiedliche Bewertungsschemata auf. Diese unterscheiden sich hauptsächlich in der Anzahl an unterschiedlichen Skalenpunkten für die Angabe der Zuversicht und an der Benotung von richtigen und falschen Antworten (Salehi et al., 2015). Für gewöhnlich wird Confidence mit C abgekürzt und durch dieses wird der Grad an Zuversicht angegeben. Wie bereits weiter oben Kapitel erwähnt, überwiegt die Anzahl an ungeraden Skalenpunkten bei Studien zu diesem Thema. Bei Confidence-Based Marking wurde meist eine 3-Punkte-Skala herangezogen: niedrig (C=1), mittel (C=2) und hoch (C=3). Nach dem Beurteilungsschema, das Gardner-Medwin (1995) am University College London (UCL) im Rahmen des *London Agreed Protocol for Teaching* (LAPT) ursprünglich für Richtig/Falsch-Fragen entwickelte, wurde entsprechend einer 3-Punkte-Skala, mit der die Zuversicht in die Korrektheit der Entscheidung angegeben wird, beurteilt (siehe Tabelle 3.1). Wenn die Antwort richtig

ist, entspricht der Grad an Zuversicht der vergebenen Punktezahl. Ist die Antwort jedoch falsch, erhält man abhängig vom gewählten Grad an Zuversicht eine unterschiedliche Punktzahl. Wählt man die erste Stufe an Zuversicht (C=1), bekommt man keine Punkte, steigend mit dem Grad an Zuversicht, werden 2 oder sogar 6 Punkte abgezogen. Beantwortet man eine Frage nicht, erhält man 0 Punkte (Gardner-Medwin, 1995, 2006; Gardner-Medwin & Curtin, 2007). Dieses Beurteilungsschema wurde nicht nur bei Studien von Gardner-Medwin herangezogen, sondern unter anderem auch bei einer Studie von Hendriks et al. (2019) in einem Zellbiologiekurs und von Novacek (2015), der im Rahmen seiner Dissertation die Implementierung von Confidence-Based Assessment in einem Flugausbildungsprogramm erforschte.

Confidence level:	C=1 (low)	C=2 (mid)	C=3 (high)	No Reply
Mark if correct:	1	2	3	(0)
Penalty if wrong:	0	-2	-6	(0)

Tabelle 3.1

Beurteilungsschema (UCL scoring scheme) für Richtig/Falsch-Fragen nach Gardner-Medwin (modifiziert nach Gardner-Medwin, 2006, S.2)

Das zuvor beschriebene UCL-Schema sollte primär vier Anforderungen erfüllen (Gardner-Medwin & Gahan, 2003, S.2):

- Simplicity:* Das Schema soll mit wenig bis gar keiner Übung leicht zu verstehen sein.
- Motivation:* Studierende sollen immer von der ehrlichen Angabe ihrer Zuversicht profitieren, sei es hoch oder niedrig.
- Flexibility:* Das Beurteilungsschema ist anwendbar für alle Fragen, bei denen es eindeutig richtige beziehungsweise eindeutig falsche Antworten gibt, ohne die Fragen speziell anpassen zu müssen.
- Validity:* Bei diesem Schema sollte eine angemessene Übereinstimmung mit Wissensüberprüfungen, die durch die mathematische Theorie gestützt wird, aufrechterhalten bleiben.

Unbeachtet darf nicht bleiben, warum die Entscheidung für dieses Beurteilungsschema fiel. Ein Grund war die Tatsache, dass dieses Studierende motiviert, den Grad an Zuversicht ehrlich anzugeben. Da Punkte für falsche Antworten mit mittlerer oder hoher Zuversicht abgezogen werden, werden Studierende eine hohe Zuversicht nicht wählen, wenn sie nicht ganz sicher sind, dass die Antwort auch richtig ist. (Gardner-Medwin, 2006; Gardner-Medwin & Gahan, 2003) Dass die richtige

Entscheidung getroffen wurde, bestätigt eine Befragung von Studierenden. „CBM is seen by our students as simple, fair, readily understood and beneficial” (Gardner-Medwin, 2006, S. 1) Entgegen der Meinung die Minuspunkte seien zu hoch, beschreibt Gardner-Medwin (2006), dass die Wahrscheinlichkeit 50% und damit viel zu hoch ist, die Frage durch pures Raten richtig zu beantworten. Für jeden möglichen Grad an Zuversicht hängt die zu erwartende Punktzahl bei einer Frage von der Wahrscheinlichkeit ab, diese Frage richtig zu beantworten. (siehe Abbildung 3.5) Ist die Zuversicht hoch ($<80\%$), dann ist $C=3$ die beste Wahl, bei einer geringen Zuversicht ($<67\%$) ist es $C=1$. $C=2$ empfiehlt sich bei einer mittleren Zuversicht mit einer Wahrscheinlichkeit größer als 67% aber kleiner als 80%. Abbildung 3.5 zeigt, dass es niemals am besten ist, nicht zu antworten, da man mit der Angabe $C=1$ Punkte erhalten kann, ohne das Risiko einzugehen Punkte zu verlieren. Studierende denken bei ihrer Wahl aber nicht explizit über diese Wahrscheinlichkeiten nach, viel mehr entwickeln sie ein intuitives Verständnis für die Risiken und Nutzen. Die Angabe der Zuversicht erfolgt hier nicht mittels Deskriptoren wie etwa *very sure*, da dies verschiedenes für unterschiedliche Personen bedeuten könnte. Daher werden die einzelnen Stufen an Zuversicht mit den zugehörigen Punkten für richtige Antworten beschrieben. (Gardner-Medwin & Gahan, 2003)

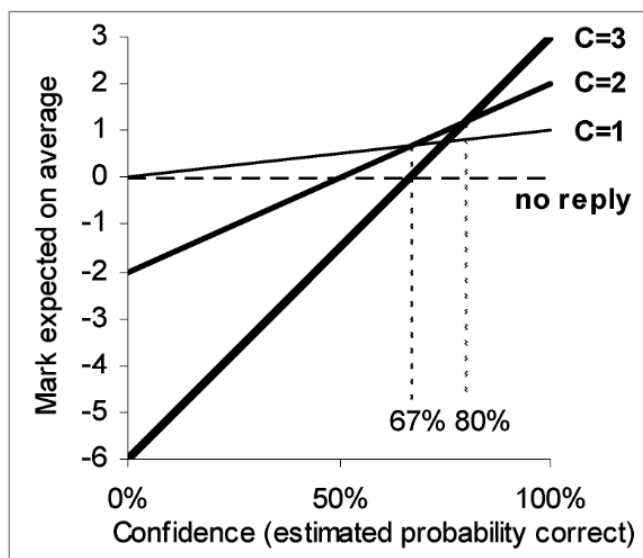


Abbildung 3.5

UCL Beurteilungsschema für Richtig/Falsch-Fragen und die Bereiche, wann welcher Grad an Zuversicht optimal ist (Gardner-Medwin & Gahan, 2003, S. 3)

Die zuvor beschriebene Beurteilungsmethode entwickelte Gardner-Medwin (2006) für Richtig/Falsch-Fragen, für Multiple-Choice-Fragen empfiehlt er eine abgewandelte Form davon, da bei diesem Fragenformat die Wahrscheinlichkeit das Ergebnis zu erraten geringer ist als 50%. Das Beurteilungsschema ist für richtige Antworten ident zu dem für Richtig/Falsch-Fragen. Bei einer

richtigen Antwort erhält man demnach dem Grad an Zuversicht C=1, C=2 und C=3 entsprechend 1, 2 oder 3 Punkte. Bei falschen Antworten hingegen erhält man mit aufsteigendem Grad an Zuversicht keine Punkte oder 1 oder 4 Punkte werden abgezogen (siehe Tabelle 3.2). Gardner-Medwin empfiehlt die Angabe von C=1, wenn die Zuversicht niedriger als 50% ist, die Frage richtig beantwortet zu haben. Zwischen 50% und 75%, wird C=2 und bei über 75% C=3 empfohlen.

Confidence level:	C=1 (low)	C=2 (mid)	C=3 (high)	No Reply
Mark if correct:	1	2	3	0
Penalty if wrong:	0	-1	-4	0

Tabelle 3.2

Beurteilungsschema für Multiple-Choice-Fragen nach Gardner-Medwin (2006)

Mit dieser und der vorhin beschriebenen Methode ist es nicht möglich Punkte zu erhalten, indem man die Zuversicht systematisch überschätzt oder unterschätzt. Die negativen Punkte, die man mit falschen Antworten und mittlerer oder hoher Zuversicht erhält, verhindern, dass eine hohe Zuversicht angegeben wird, wenn man sich nicht wirklich sicher über die Antwort ist. Aus diesem Grund bringt dieses Beurteilungsschema Schüler*innen oder Studierende dazu ihren wirklichen Grad an Zuversicht anzugeben. „This is the motivating characteristic of the mark scheme, rewarding the student’s ability to judge the reliability of an answer, not their self-confidence or diffidence. CBM is quite unlike games such as poker, in which misrepresentation of confidence can gain advantage.” (Gardner-Medwin, 2006, S.3f)

Ein weiteres Beurteilungsschema, auf das eingegangen wird, stammt von Hassmén und Hunt (1994). Diese verwendeten in ihrem Beurteilungsschema für Multiple-Choice-Fragen fünf statt drei Stufen an Zuversicht. In Tabelle 3.3 lassen sich die Punkte ablesen, die bei einer richtigen Antwort hinzugefügt und bei einer falschen Antwort abgezogen werden. Im Gegensatz zu den bisher vorgestellten Beurteilungsschemata, wird festgestellt, dass hier viel mehr Punkte vergeben, aber auch abgezogen werden. Gardner-Medwin & Gahan (2003) stufen dieses Beurteilungsschema als motivierend ein, da es für jeden Grad an Zuversicht eine optimale Spanne an Wahrscheinlichkeiten gibt. Für Richtig/Falsch-Fragen müssen die untersten beiden Stufen weggelassen werden. Unberücksichtigt sollte auch nicht bleiben, dass dieses Beurteilungsschema gerade im Vergleich zu den beiden vorhin genannten sehr komplex, schwer zu merken und zu verstehen ist. Aus diesem Grund wurde dieses Beurteilungsschema oft verwendet, ohne dass Studierende wussten, wie die Punkte vergeben werden. (Gardner-Medwin, 2006)

Confidence level:	Almost guess	Probable guess	Neutral	Fairly certain	Almost certain
Mark if correct:	20	54	74	94	100
Penalty if wrong:	10	-8	-32	-64	-120

Tabelle 3.3

Beurteilungsschema für Multiple-Choice-Fragen von Hassmén und Hunt (in Anlehnung an Hassmén und Hunt, 1994)

Bis jetzt wurden Beurteilungsschemata vorgestellt, bei denen die Punkte, die hinzugefügt oder abgezogen werden, unterschiedlich sind. Anders ist dies in einem Beurteilungsschema für Multiple-Choice-Fragen von Davies (2002). Bei einer richtigen Antwort erhält man abhängig vom Grad an Zuversicht 1, 2 oder 3 Punkte, bei einer falschen Antwort werden 1, 2 oder 3 Punkte abgezogen (siehe Tabelle 3.4). Festzuhalten ist, dass nicht alle dieser Schemata gleich motivierend sind. Jenes von Davies (2002) ist laut Gardner-Medwin & Gahan (2003) und Gardner-Medwin (2006) ein Beispiel für ein nicht motivierendes Beurteilungsschema. Es wird damit begründet, dass es die beste Strategie ist C=3 zu wählen oder gar nicht zu antworten. Aufgrund dessen wird dieses Schema als nicht fair eingestuft, da es diejenigen bevorzugt, die immer C=3 wählen unabhängig davon, ob sie ehrlich oder unehrlich sind.

	C=1	C=2	C=3
Mark if correct:	1	2	3
Penalty if wrong:	-1	-2	-3

Tabelle 3.4

Beurteilungsschema für Multiple-Choice-Fragen nach Davies (2002)

Ein weiteres Beurteilungssystem, ist jenes mit einer 6-stufigen Skala von Leclercq und Poumay (2005), das in Kapitel 3.2.1 bereits vorgestellt wurde. Es wird in unterschiedlichen Universitäten wie in Liège, Paris und Aosta verwendet. Leclercq und Poumay beschreiben in ihrem Artikel die von ihnen im Jahr 2003 eingeführten Indizes: „The Confidence index is the average confidence degree given with correct answers. The Imprudence index is the average confidence degree given with incorrect answers [sic] The Nuance index is the difference between those two indices.” (S. 5) Folgt man diesem Beurteilungssystem, erhält man 1 Punkt für jede richtige Antwort, für jede falsche Antwort werden 0,25 Punkte abgezogen unabhängig vom Format der Frage. Dies wird als *classical score* bezeichnet. Anschließend wird die Gesamtpunktzahl auf eine 20-Punkte-Skala transformiert. Im Anschluss werden zum *classical score* Punkte hinzugefügt, jedoch nur, falls der Grad

an Zuversicht bei allen Fragen angegeben wurde, falls nicht überall das gleiche gewählt und falls die Punkteanzahl zum Bestehen schon erreicht wurde. Treffen diese Kriterien zu, wird 1 Punkt für Confidence, 1 Punkt für Prudence und 1 Punkt für Nuance hinzugefügt, wobei letzterer nur vergeben wird, falls die ersten beiden Punkte vergeben wurden. Leclercq und Poumay (2005) stellen in ihrem Artikel neben dieser Beurteilungsweise noch weitere vor. Diese finden sich in ihrem Artikel auf Seite 7ff.

Nicht unbesprochen bleiben sollte das Beurteilungsschema von Foster (2016, 2021) bleiben, da es den bisher vorgestellten kaum ähnelt. Foster führte seine Studien in Schulen durch und verwendete dabei eine Skala von 0 bis 10, wobei 0 niedrig und 10 sehr hoch ist. Das Ergebnis der Schüler*innen wird berechnet, indem die Angaben der Zuversicht bei korrekten Antworten addiert und die Angaben der Zuversicht bei falschen Antworten subtrahiert werden. Dieses Bewertungssystem soll Schüler*innen dazu anregen, wahrheitsgetreue Angaben der Zuversicht zu tätigen.

Abschließend soll noch das Ergebnis einer Studie von Gvozdenko und Chambers (2007) kurz erwähnt werden. Diese Studie ergab, dass 85% der Teilnehmer*innen eine 3- oder 4-stufige-Skala gegenüber einer 5-stufigen-Skala bevorzugen, da diese weniger komplex und dadurch einfacher zu verstehen seien. Nachdem in diesem Kapitel die unterschiedlichen Bewertungsschemata für Confidence-Based Marking erläutert wurden, konzentriert sich das folgende Kapitel auf die Vorteile von Confidence-Based Assessment/Marking.

3.5 Vorteile von Confidence-Based Assessment und Marking

Mit dem Einbeziehen der Zuversicht als Komponente des Wissens gehen einige Vorteile einher. Hassmén et al. (2002) führt in seinem Artikel einige davon an. Durch den Einbezug von Zuversicht bei Wissensüberprüfungen kann beispielsweise unterschieden werden, ob Personen *uninformed* oder *misinformed* sind. Ebenso ist eine Unterscheidung zwischen *usably informed* und *unusably informed* nach Hunt und Furustig (1989) möglich. Betrachtet man die Definition von Hunt (1993), lässt sich durch Confidence-Based Assessment auch zwischen *usable* und *unusable knowledge* (siehe Kapitel 2.2) differenzieren. Ein weiterer Grund, der für den Einbezug von Zuversicht im Sinne von Confidence-Based Assessment spricht, ist die Verringerung geschlechtsbezogener Bevorzugung (siehe Kapitel 2.4). Weiters ist es möglich, dadurch zu erkennen wie viel Wissen abhängig vom Grad an Zuversicht über bestimmte Zeit erhalten bleibt. Mit dieser Thematik haben sich einige Studien

befasst. Dies ist im Kapitel 2.5 nachzulesen. Hier ist außerdem der Gedanke von Hunt (1982) und Hassmén et al. (1996) aufzugreifen, dass die Verwendung von Confidence-Based Assessment sogar das Lernen an sich verbessert. Hunt (1982) behauptet dazu in seinem Artikel, dass „[...]making an accurate self-assessment along with making a correct response to be learned may be more rewarding than simply making a correct response alone.“(S.76) Ein weiterer Punkt auf den eingegangen werden soll, ist, dass man durch Confidence-Based Assessment die Fähigkeit der Selbstkontrolle beim Lernen und die metakognitive Fähigkeit die eigene Antwort zu begründen entwickelt. Zudem lernt man, dass es mehrere Möglichkeiten gibt, Probleme zu lösen. Effizientes Lernen und großes Wissen beinhalten die Fähigkeit unterschiedliche Ideen miteinander zu vergleichen. Wie bereits erwähnt, benötigen falsche Antworten mit hoher Zuversicht besondere Aufmerksamkeit – in diesem Zusammenhang erkennt man, dass man hier fehlinformiert ist. Hinzu kommt, dass Lernende dazu animiert werden, Antworten mit anderem Wissen in Beziehung zu setzen. Unerwähnt darf nicht bleiben, dass es Lernende anspricht sich nicht auf auswendiggelernte Fakten zu verlassen, sondern Zusammenhänge zu verstehen und willkürliches Raten in der Hoffnung auf Zufallstreffer nicht als Option zu sehen. (Davies, 2000; Gardner-Medwin & Gahan, 2003)

Bei gewöhnlichen Prüfungen erhält man einerseits selten Informationen darüber, wie zuversichtlich Studierende in Bezug auf ihre Antworten sind, andererseits ist es auch oft mit der Erstellung von sehr komplexen Fragen verbunden, da Lehrende beziehungsweise Prüfer*innen versuchen diese so zu gestalten, dass Lernende keine gute Note erhalten können, wenn man nur Teilwissen hat oder gar bei der Beantwortung der Frage rät. (Gardner-Medwin, 1995) Mit Confidence-Based Marking lässt sich so nun viel Zeit bei der Erstellung von Fragen sparen. Man ist nicht darauf angewiesen Fragen mit einem angemessenen Schwierigkeitsgrad zu stellen, da die Zuversicht in Antworten hinzugezogen wird. Damit erübrigt sich das oben beschriebene Problem. Sehr einfache Fragen, die unter anderen Umständen erniedrigend wirken könnten, haben hier das Potential die Denkweise „OK, definitely a C=3 here“ (Gardner-Medwin, 2019a, S.4) anzuregen. Ein weiterer Vorteil wird an dieser Stelle besonders betont. Für Lehrende können die Ergebnisse von Prüfungen mit Confidence-Based Marking äußerst aufschlussreich sein. Auf der einen Seite erhält man eine Rückmeldung über die Qualität der Fragen, auf der anderen Seite lassen sich auch Unsicherheiten bei einfachen Fragen aufdecken. Letzteres wäre durch richtige Antworten und einen niedrigen Grad an Zuversicht zu identifizieren. Schwache Schüler*innen oder Studierende werden ermutigt sowohl ihre Schwächen als auch ihre Stärken zu identifizieren. (Gardner-Medwin, 2019) Durch Confidence-Based Marking erhalten Schüler*innen oder Studierende auch ein besonderes Feedback, das unter anderem als „wake-up call“ (Gardner-Medwin, 2018, Certainty-Based Marking) dienen kann. Gemeint ist damit, dass diejenigen Probleme angesprochen werden können, deren Existenz zuvor nicht bekannt

war. An dieser Stelle soll auch betont werden, dass geschlechtsspezifische Vorurteile durch Confidence-Based Marking sehr stark reduziert werden oder gar ausbleiben. (Hassmén & Hunt, 1994; Gardner-Medwin, 1995; Salehi et al., 2015) Siehe dazu auch Kapitel 2.4.

Im Artikel von Gardner-Medwin (2006) werden Gründe aus Studierendensicht genannt, die für Confidence-Based Marking sprechen. Zum einen muss man fähig sein seine Antwort so weit zu vertreten, dass man im Falle bereit ist negative Punkte zu erhalten. Nur mit dieser Einstellung kann man volle Punktezahlen erhalten. Zum anderen ist es auch möglich bei Unsicherheit gegenüber einer Antwort Punkte zu erhalten, indem man einen niedrigeren Grad an Zuversicht wählt. Das ist jener Aspekt, den Gardner-Medwin (2003) als motivierenden Faktor von Confidence-Based Marking bezeichnet. Da die fatalste Form von Wissen ist, wenn man sehr zuversichtlich eine falsche Antwort vertritt (siehe Kapitel 2.2), werden dabei auch mehr Punkte abgezogen, als wenn man eine falsche Antwort angibt mit einer niedrigeren Zuversicht. Auch diesen Aspekt sehen Studierende als Grund, der für die Verwendung von Confidence-Based Marking spricht. (Gardner-Medwin, 2006)

In einer Evaluationsstudie von Issroff und Gardner-Medwin (1998) werden einige Kommentare der Studierenden in Bezug auf Confidence-Based-Marking angeführt, von denen hier ein paar zitiert werden: „Useful guide to level of knowledge and real ability“(S. 175), „If a low score is given, I know I need to work much harder and go back to the books“(S. 176), “Highlighted difficulties so I knew what topics to revise” (S. 177) An letzter Stelle wird ein weiteres Zitat angebracht, das im Artikel von Issroff und Gardner-Medwin aus dem Jahr 1998 nicht angeführt wird, allerdings in der Publikation von Gardner-Medwin (2006) aufscheint: „It .. stops you making rush answers.“ (S. 2)

Die Beurteilung mit Einbezug der Zuversicht verhindert, dass es Studierende gibt, die jedes Mal den höchsten Grad an Zuversicht wählen, da sie davon nicht profitieren, sondern bei falschen Antworten, sogar je nach Beurteilungsschema sehr viele Punkte verlieren können. (Novacek, 2013) Wie nun klar ersichtlich ist, gibt es für Confidence-Based Marking einige Vorteile. Jedoch ist es auch nicht unumstritten. Oftmals kommt auf, dass Confidence-Based Marking unfair sei. Ahlgren (1996) stellt dem entgegen, dass Confidence-Based Assessment „less in a purely psychometric or psychological context, and more in a context of education“ (S. 9) gesehen werden sollte, und meint damit etwa die Verbesserung der Reliabilität. Die angebliche Unfairness bedarf es auch zu erforschen, aber diese sei nur sekundär zu betrachten. Obwohl es viele Gründe gibt, die für Confidence-Based Marking sprechen, ist es noch sehr wenig verbreitet. Oftmals wird angeführt, dass Confidence-Based Marking bestimmte Persönlichkeitsmerkmale bevorzugt, unter anderem, dass es zu geschlechtsspezifischen Vorurteilen führt. Dies wird vor allem darin begründet, dass es schüchterne und risikoscheue Persönlichkeiten, die mutmaßlich häufiger unter Frauen vorkommen, benachteiligt.

(Gardner-Medwin, 2006) In dem Artikel von Gardner-Medwin und Gahan (2003) werden Daten aus ihrer Erfahrung am University College London präsentiert. In Abbildung 3.6 lässt sich erkennen, dass es keine signifikanten geschlechtsspezifischen Unterschiede gibt. Ausführlichere Informationen bezüglich dieser Studie finden sich im Artikel von Gardner-Medwin und Gahan (2003).

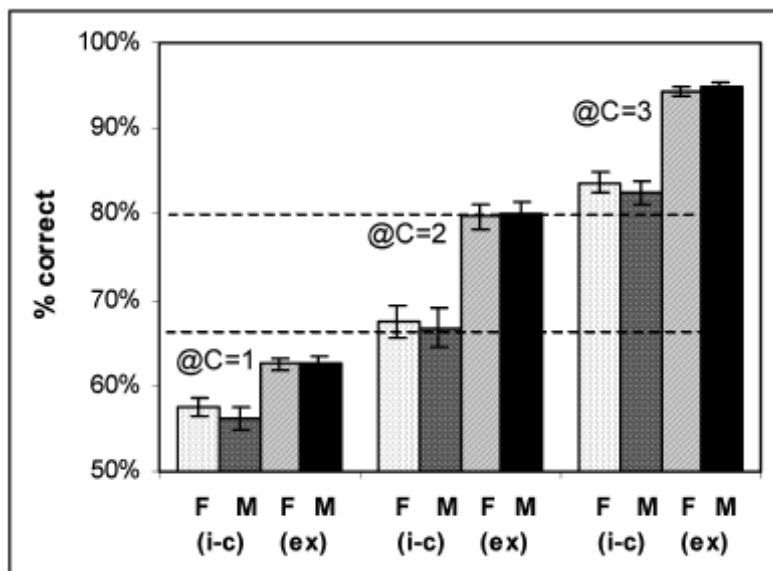


Abbildung 3.6

Durchschnittlicher Prozentsatz an richtig beantworteten Fragen für Richtig/Falsch-Fragen für die unterschiedlichen Level an Zuversicht ($C=1$, 2 oder 3) und für die unterschiedlichen Geschlechter (Gardner-Medwin & Gahan, 2003, S. 5)

Die in diesem Kapitel beschriebenen Vorteile beziehen sich zum Teil nur auf den Einbezug der Zuversicht als Komponente des Wissens. Die Autorin schlägt vor diese Vorteile auch Confidence-Based Assessment – nach ihrem Versuch der Trennung – zuzuschreiben. Denn ohne die Zuversicht in die Beurteilung einfließen zu lassen, lässt sich erkennen, ob Schüler*innen zum Beispiel *usuably* oder *unusuably informed* sind. Im Anschluss an die Darlegung der Vorteile von Confidence-Based Assessment/Marking beschäftigt sich das folgende Kapitel mit Studien, die im Mathematikunterricht an Schulen durchgeführt wurden.

3.6 Confidence-Based Assessment im Mathematikunterricht

Confidence-Based Assessment wurde in der Hochschuldidaktik schon mehrfach vor allem in Bereichen der Medizin und verwandten Gebieten erforscht. (Gardner-Medwin, 1995, 2006, 2019a, 2019b; Issroff & Gardner-Medwin, 1998) Es ist klar, dass es vor allem in der Medizin unerlässlich ist, dass das Wissen jeder einzelnen Person nicht nur solide ist, sondern auch, dass er*sie beurteilen kann, was er*sie weiß und was er*sie nicht weiß. Außerhalb der Hochschulbildung oder generell außerhalb post-sekundärer Bildung, wurde Confidence-Based Assessment/Marking noch relativ wenig erforscht.

Dennoch gibt es ein paar wenige Studien, die Confidence-Based Marking in der Schule zum einen in Physik und zum anderen in Mathematik behandeln. Da sich dieses Kapitel ausschließlich auf Studien im Mathematikunterricht bezieht, wird hier auf keine Studien anderer Unterrichtsfächer eingegangen. Trotz dessen wird an dieser Stelle als Beispiel auf eine Studie im Physikunterricht von Clark (2020) verwiesen. Wie bereits erwähnt, stellt sich die Studienlage äußerst schwach dar. Es gibt jedoch Studien aus den letzten Jahren zu Confidence-Based Assessment und Marking im Mathematikunterricht. Um es vorwegzunehmen, auch diese Studien befassten sich alle mit der Erforschung des Einbezugs der Zuversicht in die Beurteilung.

Florian (2010) beschäftigte sich im Rahmen seiner Doktorarbeit mit dem Einfluss von Confidence-Based Assessment auf Schüler*innen der Sekundarstufe. Seine Forschungsfrage befasste sich mit der Verwendung von Moodle, einer Lernplattform, um Confidence-Based Assessment zu implementieren, mit dem Ziel die Fähigkeit kritisches Denken zu messen, zu verbessern und zu erweitern. Er zog dafür Fragebögen und Interviews mit Lehrpersonen, Administrator*innen und Entwickler*innen hinzu und sammelte daraus qualitative Daten. Es stellte sich heraus, dass Confidence-Based Assessment in Moodle nicht nur das Hintergrundwissen einer Person evaluieren, sondern auch Schüler*innen darin ausbilden kann kritisch innerhalb dieser Disziplin zu denken.

Eine weitere Studie führte Foster (2016) mit 14 Klassen in fünf Schulen in drei unterschiedlichen Städten im Vereinten Königreich durch. Die Forschungsfragen lauteten:

1. How do school mathematics pupils respond to the use of a CA instrument designed to incentivise [sic] the eliciting of truthful confidence ratings? How easy or difficult do they find it to understand? How readily or not do they accept it as a formative assessment tool?
2. What is the relationship between school mathematics pupils' confidence and competence within the topic of directed (positive and negative) numbers? In particular, how well calibrated are pupils within this topic? (S. 276)

Der Auftrag der Lehrpersonen war es Klassen zu wählen, die schon Grundlagen im Bereich Ganze Zahlen haben. Welche Grundlagen im genauen gemeint sind, wurde nicht spezifiziert. Die Studie beinhaltete 10 Fragen zu Ganzen Zahlen, bei denen die Schüler*innen jeweils ihre Zuversicht auf einer Skala von 0 bis 10 angeben mussten. Die Gesamtanzahl berechnete sich anschließend aus der Summe der Angaben der Zuversicht für richtige Antworten minus der Summe der Angaben der Zuversicht der falschen Antworten. Obwohl üblicherweise eine 3-Punkte-Skala verwendet wird, wurde hier eine Skala von 0 bis 10 herangezogen, da sich daraus eine maximale Punkteanzahl von 100 Punkten für insgesamt 10 Fragen ergibt. Foster (2016) gab den Lehrpersonen genaue Anweisungen wie die Studie in den Klassen durchzuführen ist (siehe Foster, 2016, Kapitel 4.3). Unerwähnt darf nicht bleiben, dass die Schüler*innen ihr Aufgabenblatt, das in Abbildung 3.7 ersichtlich ist, selbst korrigierten, indem sie in der Spalte *Leave Blank* nach Beantwortung aller Fragen ein Plus für eine richtige Antwort und ein Minus für eine falsche Antwort eintrugen. Insgesamt nahmen 345 Schüler*innen an dieser Studie teil. Die Ergebnisse zeigen zum einen, dass die Schüler*innen das Instrument sehr gut verstanden und versuchten ihren wahren Grad an Zuversicht anzugeben. Zum anderen konnte auch festgestellt werden, dass die Schüler*innen Confidence-Based Assessment akzeptierten. Neben diesen Ergebnissen konnte ebenso nachgewiesen werden, dass in Bezug auf das Thema Ganze Zahlen die Zuversicht und Kompetenz der Schüler*innen gut zusammenpasste (siehe Kapitel 3.3).

Directed Numbers

For each of the 10 questions below, write the answer in the "Answer" box.

Each time, also write in the "How Sure?" box a whole number from 0 to 10 to indicate how sure you are that your answer is correct. On this scale 0 is "completely unsure" and 10 is "completely certain".

	Question	Answer	Leave Blank	How Sure?
1	$5 - 8$			
2	$-2 + 7$			
3	$-3 - 5$			
4	$6 - (-1)$			
5	$4 + (-9)$			
6	$(-3) \times 4$			
7	$5 \times (-2)$			
8	$(-2) \times (-3)$			
9	$\frac{15}{-3}$			
10	$\frac{-10}{-2}$			

Comments

--

Abbildung 3.7

*Aufgabenblatt der Schüler*innen (Foster, 2016, S. 278)*

Mit dieser Studie sollte auch gezeigt werden, dass es sehr einfach ist, Confidence-Based Assessment in der Schule ohne viel Aufwand zu implementieren, da Lehrpersonen die ursprünglichen Überprüfungsformate beibehalten und die Schüler*innen bitten können, am Rand bei jeder Aufgabe eine Zahl von 0 bis 10 für ihre Zuversicht dazuschreiben. (Foster, 2016; Foster et. al., 2022) Es wurde bereits deutlich gemacht, dass Confidence-Based Assessment und Marking viele Vorteile im Hochschulbereich mit sich bringt. Die Behauptung wird aufgestellt, dass diese Vorteile auch für formative Beurteilung in Mathematik in der Schule gelten könnten. (Foster, 2016, 2017, 2021; Foster et al., 2022) Ein möglicher Vorteil ist, dass es entmutigt bei Aufgaben einfach zu raten. Weiters nennt Foster (2021), dass es sowohl Selbstüberschätzung als auch Selbstunterschätzung reduziert. Es wird

besonders betont, dass es nicht Sinn und Zweck ist, nur die Zuversicht der Schüler*innen zu steigern. Teilweise ist es sogar notwendig, diese für eine gewisse Zeit bei einem bestimmten Aspekt zu reduzieren, wie etwa bei falschen Antworten mit hoher Zuversicht.

An diesem Punkt soll eine weitere Forschung präsentiert werden. Die Forschungsfrage von Foster (2021) lautete: „Does repeated use of CA in formative assessments in school mathematic lessons over an extended time period result in improvements in students’ mathematics attainment in summative assessments?“ (S. 5) An dieser Studie nahmen vier Schulen mit je zwei Klassen teil, je eine Kontrollgruppe und eine Interventionsgruppe. Die Intervention wurde sehr geringgehalten, indem die Schüler*innen bei formativen Beurteilungen neben die Aufgaben ihre Zuversicht von 0 bis 10 notieren und im Anschluss nach dem Beurteilungsschema von Foster (2016) ihre Punkte berechnen sollten. Es wurden dann die Noten der Schüler*innen bei summativen Beurteilungen vor und nach dem Zeitraum, in dem Confidence-Based Assessment genutzt wurde, verglichen und überprüft, ob die Verwendung von Confidence-Based Assessment eine Verbesserung darstellt. Die Ergebnisse dieser Studie ergaben, dass die Intervention für die Leistungen der Schüler*innen nicht nachteilig war, aber diese dadurch auch nicht verbessert wurden. (Foster, 2021) Im Artikel von Foster (2021) werden verschiedene Begründungen für dieses Ergebnis angeführt. Das Angeben der Zuversicht könnte den Flow der Schüler*innen während der Beantwortung von Fragen unterbrochen haben und somit das Lernen negativ beeinflussen. Weiters könnte Confidence-Based Assessment viel Zeit im Unterricht in Anspruch genommen haben, die für „more effective aspects of the lesson“ (Abschnitt Discussion) notwendig gewesen wäre. Dies konnte jedoch nicht belegt werden. Ebenso werden Gründe genannt, die für eine Annahme der Verbesserung der mathematischen Leistung von Schüler*innen sprechen. Warum dies jedoch nicht eingetreten ist, begründet der Autor wie folgt. Ein möglicher Grund ist, dass zwischen der Messung und der Intervention zu viel Zeit lag und, dass die Annahme einer Verbesserung durch die Verwendung von Confidence-Based Assessment von Zeit zu Zeit zu optimistisch war. Foster (2021) vermutet demnach, dass die tägliche Verwendung von Confidence-Based Assessment notwendig wäre, um eine messbare Verbesserung zu bemerken. Da die Schulen unter keiner Beobachtung standen, kann nicht gesagt werden, wie regelmäßig und sorgfältig Confidence-Based Assessment von den Lehrpersonen eingesetzt wurde. Dem ist hinzuzufügen, dass die Schulen lediglich einen Zettel erhielten, die diese Methode schilderten. Demnach ist es möglich, dass wichtige Informationen untergegangen sind und den Lehrpersonen zu wenig Zeit zur Verfügung stand, um über Confidence-Based Assessment zu reflektieren und gegebenenfalls darüber zu diskutieren.

Die Studie von Foster et al. (2022) darf ebenso nicht unerwähnt bleiben. Sie verwendeten bei ihrer Studie eine 5-stufige Likert-Skala in Form von *confidence emojis*, die in Kapitel 3.2.1 bereits vorgestellt wurden. Ihre Studie beinhaltete 7302 Mathematikschüler*innen im Alter zwischen 6 und 16 Jahren des Vereinigten Königreichs. Bei ihrer Forschung zogen sie ein Datenset von einer online Lernplattform namens *Eedi* (eedi.co.uk) heran, die von Woodhead und Barton, Co-Autoren dieses Artikels, entwickelt wurde. Diese Lernplattform umfasst über 50.000 Multiple-Choice Fragen, die die Angabe der Zuversicht beinhalten (siehe Abbildung 3.8). Mit ihrer Studie wollten Foster et al. (2022) folgende Fragen beantworten:

1. How are students' confidence ratings related to their mean facility, age, gender and socioeconomic disadvantage, and how do these variables interact?
2. Is there evidence for the hypercorrection effect in students' responses to a second set of questions (quiz B) administered 3 weeks after the first (quiz A)? (S. 495)

Mit ihrer Studie konnten sie einen positiven Zusammenhang zwischen der Zuversicht eines*einer Schüler*in und deren mittleren Anzahl an korrekten Antworten nachweisen. Schüler gaben eine höhere Zuversicht im Vergleich zu Schülerinnen an. Weiters konnte gezeigt werden, dass die Zuversicht niedriger bei sozioökonomisch benachteiligten Schüler*innen war, gleiches gilt für ältere Schüler*innen. Ebenso konnte der erste Nachweis für den *hypercorrection effect* (Butterfield & Metcalfe, 2001) in einem schulmathematischen Kontext erbracht werden. Genauere Details zur Studie finden sich in Foster et al. (2022).



Which list below has the numbers in **descending** order?

A	B	C	D
7, -5, -3, -1	9, 1, -2, -6	-4, -2, 1, 3	-3, 4, 6, -8

Copyright © AQA and its licensors. All rights reserved.

A

B

C

D

Explain why you answered B

How confident are you in your answer?



Next

Abbildung 3.8

Typische Frage von Eedi (Foster et al., 2022, S.493)

3.7 Einsatzmöglichkeiten im schulischen Kontext

Vorweg werden zwei Punkte erwähnt, einerseits, dass sich dieses Kapitel bezüglich der Abgrenzung der Begriffe Confidence-Based Assessment und Confidence-Based Marking an den Versuch aus Kapitel 3.1 hält. Andererseits wird an dieser Stelle betont, dass der Einbezug von Confidence-Based Assessment nicht auf den Mathematikunterricht beschränkt ist, sondern in jedem Schulfach auf seine eigene Art und Weise Einzug finden kann. Wie in dieser Arbeit gezeigt werden kann, ist die Implementierung von Confidence-Based Assessment sehr sinnvoll. Hingegen erscheint es als nicht notwendig Confidence-Based Marking einzuführen, denn allein die Information, die man aus Confidence-Based Assessment erhält, ohne danach zu beurteilen, ist sehr aufschlussreich. Auch ohne Confidence-Based Marking lassen sich nützliche Informationen sowohl für Schüler*innen und Lehrpersonen erhalten. Äußerst wichtig erscheint, dass den Schüler*innen der Nutzen des Einbezugs der Zuversicht bewusst ist. Eine Möglichkeit, dies den Schüler*innen bewusst zu machen, ist ein Peer-Nutzengespräch, in dem sich Peers über den Nutzen von Confidence-Based Assessment unterhalten und austauschen. Im Anschluss werden die Ergebnisse an der Tafel gesammelt oder zum Beispiel auf Padlet (www.padlet.com) grafisch dargestellt werden, um abgespeichert werden zu können.

Die reine Implementierung der Angabe der Zuversicht ohne weitere Schritte ist nicht zielführend. Wenn die Schüler*innen bloß angeben, wie zuversichtlich sie sind, ohne dann eine Rückmeldung zu erhalten, wird bald die Motivation nicht mehr da sein, die Zuversicht überhaupt anzugeben, geschweige denn die Wahl der Zuversicht ehrlich zu gestalten. Daher sieht die Autorin es als sehr wichtig an, dass die Schüler*innen vorweg über Vorteile und den Nutzen informiert werden, die damit einhergehen können. Florian (2010) demonstriert in seinem Artikel auch, dass es von hoher Bedeutung ist, dass Schüler*innen aktiv dabei sind, da es ansonsten sehr schwierig ist positive Resultate durch Confidence-Based Assessment zu erzielen. Er hält fest, dass gerade die Einführung in Confidence-Based Assessment und die anfängliche Übung unbedingt positiv sein müssen. „During the beginning stages, students need to understand that this is a new method of assessment. It will take time for them to understand how to take a CBA and to use the information that it gives each student.“ (Florian, 2010, S. 109)

Eine Möglichkeit Confidence-Based Assessment in den Unterricht zu implementieren, ist, dass Lehrpersonen Aufgaben, die falsch gelöst wurden und bei denen eine hohe Zuversicht angegeben wurde, speziell markieren, beispielsweise ein rotes Rufzeichen neben jene Aufgabe zu setzen. Für die weitere Vorgehensweise gibt es unterschiedliche Optionen, wobei an dieser Stelle zwei genannt werden. Es wird entweder in einem Lehrer*in-Schüler*in-Gespräch (z.B. Zuversichtsgespräch)

besprochen oder eine bestimmte Vorgangsweise in dieser Klasse ausgemacht, die dementsprechend sehr individuell gestaltet werden kann. Die beiden eben genannten Vorschläge schließen einander nicht aus. Ein Beispiel für eine Vorgangsweise ist, dass Schüler*innen die Thematik des Beispiels im Schulübungsheft nachschlagen und dieses und ein weiteres ähnliches selbst wählbares Beispiel dazu lösen. Im Falle einer Hausübung wird dies mit der kommenden Hausübung, im Falle einer Lernzielkontrolle gemeinsam mit der Verbesserung abgegeben. Eine derartige Vorgangsweise ist natürlich nicht nur sinnvoll bei falsch gelösten Aufgaben mit hoher Zuversicht, um hier Fehlinformationen auszubessern, sondern auch bei richtig gelösten Aufgaben mit niedriger Zuversicht, um dabei das Wissen zu festigen und dadurch auch die Zuversicht in dieses Wissen zu stärken. Hierfür kann entweder ein anderes Symbol verwendet werden oder das gleiche und die Schüler*innen wissen anhand der Angabe der Zuversicht und Richtigkeit der Aufgabe Bescheid. An dieser Stelle wird erwähnt, dass die Angabe der Zuversicht individuell sein kann. Foster (2016, 2021) schlug vor die Zuversicht in Form von einer Zahl von 0 bis 10 neben die Aufgabe schreiben zu lassen. Wie später ersichtlich (Kapitel 5), wählte die Autorin in ihrer Studie eine andere Methode, die der von Foster et al. (2022) ähnelt.

An dieser Stelle wird ein Zitat aus einem Artikel von Foster (2017) genannt: „Sometimes in maths, ‘I don’t know’ is exactly the right answer[...]“ (S. 1). Aus persönlicher Erfahrung der Autorin als Mathematiklehrerin wird dieser Gedanke nun weitergeführt und es kann festgehalten werden, dass es nicht zu selten vorkommt, dass Schüler*innen Hausaufgaben abschreiben. Versetzt man sich in die Lage der Schüler*innen, erscheint dies – natürlich nicht immer – als Möglichkeit einer schlechten Beurteilung zu entfliehen. Falls er*sie einfach nicht weiß, wie eine Aufgabe zu lösen ist und diese deshalb nicht gemacht wurde, würde sich dies negativ auf die Mitarbeit auswirken. In Bezug auf die Zuversicht und das obige Zitat, wäre es sehr sinnvoll eine Möglichkeit anzubieten, mit der Schüler*innen angeben können, dass sie einfach keine Ahnung haben – also „I don’t know“. Dies würde den Lehrpersonen eine sehr sinnvolle Information liefern. Wie Lehrpersonen dann verhindern, dass Schüler*innen wahllos angeben, dass sie keine Ahnung haben, und, ob es vielleicht doch eine sinnvolle Form von Confidence-Based Marking für den schulischen Alltag gibt, sind weitere Fragestellungen.

Natürlich lässt sich so wie in den bisherigen Erläuterungen in diesem Kapitel Confidence-Based Assessment nicht nur analog verwenden. Confidence-Based Assessment ist bereits in der LMS-Plattform Moodle (www.moodle.org) mit vielen Entwicklungsmöglichkeiten implementiert. Auf folgender Seite, die von Gardner-Medwin verwaltet wird, erhält man Zugriff auf diese Plugins für Moodle: https://moodle.org/plugins/view.php?plugin=quiz_cbmgrades. Florian (2010), Gardner-

Medwin (2019a) und auch Smrkolj et al. (2022) nutzten Moodle bereits für ihre Studien, ersterer forschte sogar im schulischen Rahmen.

„Because CBA has had limited exposure in secondary education institutions, there may be a misconception that it is another new program that will be gone in short fashion.“ (Florian, 2010, S. 109) Um genau diesem Phänomen entgegenzuwirken, wurde hier versucht Möglichkeiten zu sammeln, wie Confidence-Based Assessment in den schulischen Alltag ohne viel Aufwand implementiert werden kann. Ein großer Punkt beim Einbezug der Zuversicht als Komponente des Wissens ist die Selbsteinschätzung der Studierenden oder Schüler*innen. Da man als Schüler*in oder Student*in niemals davon ausgehen kann, durch dauerhaftes Über- oder Unterschätzen zu profitieren, ist es unerlässlich zu lernen sich richtig einzuschätzen. Demnach befasst sich das folgende Kapitel mit dem Thema der Selbsteinschätzung.

4 Selbsteinschätzung

Die beiden vorangegangenen Kapitel behandeln eine Methode, bei der Schüler*innen angeben müssen, wie zuversichtlich sie in Bezug auf ihre Antwort sind. Im Prinzip steckt hinter dieser Angabe eine Selbsteinschätzung, die die „Einschätzung der eigenen Person in Hinblick auf bestimmte Fähigkeiten, Fehler u.Ä.“ (Dudenredaktion, o.D.) ist. Diese wird von vielen verschiedenen Faktoren beeinflusst, wie etwa dem Selbstkonzept und verschiedenen kognitiven Verzerrungen, die im Folgenden näher beleuchtet werden.

4.1 Selbstkonzept

Einer der Faktoren, von denen die Selbsteinschätzung abhängt, ist das Selbstkonzept. Die Begriffe Selbstkonzept und damit einhergehend auch der Selbstwert sind unter den Forscher*innen weitestgehend geklärt und nahezu einheitlich definiert, wodurch diese sich untereinander kaum unterscheiden. Lohaus et al. (2010) definieren in ihrem Buch über Entwicklungspsychologie im Kindes- und Jugendalter das Selbstkonzept wie folgt: „Das Selbstkonzept besteht als kognitive Komponente [sic] des Selbst aus der Selbstwahrnehmung und dem Wissen um das, was die eigene Person ausmacht. Neben persönlichen Eigenschaften und Fähigkeiten, die man besitzt, gehören zu diesem Wissen auch Neigungen, Interessen und typische Verhaltensweisen.“ (S. 165) Ebenso wird der Selbstwert definiert: „Der Selbstwert resultiert als affektive Komponente des Selbst aus den Bewertungen der eigenen Person oder von Aspekten, die die eigene Person ausmachen. Somit können sich die Bewertungen auf Persönlichkeitseigenschaften, Fähigkeiten oder aber auch auf das eigene emotionale Erleben beziehen.“ (S. 165)

In der modernen Forschung zum Selbstkonzept ist man sich einig, dass sich das Selbstkonzept aus vielen unterschiedlichen Bereichen zusammensetzt. Bereits 1983 konnten Marsh et al. mit ihrer Studie die Multidimensionalität des Selbstkonzepts untermauern. Demnach wird nicht mehr von einem globalen Selbstkonzept ausgegangen. (Lohaus et al., 2010) Bevor im Anschluss das akademische Selbstkonzept näher erläutert wird, wird im Folgenden ein Überblick über die Entwicklung der Selbstkonzeptforschung gegeben.

4.1.1 Entwicklung der Selbstkonzeptforschung

Bereits William James beschreibt das Selbst im Jahr 1890 als duales Phänomen. Er sieht es als „[...] das erkennende Subjekt (Selbst als Subjekt, »self as knower«) und das zu erkennende Objekt

(Selbst als Objekt, »self as known«).“ (Lohaus et al., 2010, S. 165) Das erkennende Subjekt meint das Selbst-Erleben, das erkennende Objekt hingegen das Wissen um die eigene Person und im weiteren Sinne das Selbstkonzept an sich. Dieses Wissen wird nach James in drei Bereiche gegliedert:

- „Das materielle Selbst (Wissen um den eigenen Körper),
- das spirituelle Selbst (Wissen um die eigenen geistigen Eigenschaften)
- und das soziale Selbst (Wissen um die Sicht anderer auf sich selbst).“ (Lohaus et al., 2010, S. 165)

Ersteres bezieht sich auf das unmittelbare Selbst-Erleben, während Letzteres dem entspricht, was eingangs als Selbstkonzept definiert wurde, nämlich dem Wissen um die eigene Person, die dabei zum Objekt des Erkenntnisgewinns wird. Dieses Wissen setzt sich nach James aus dem materiellen, dem spirituellen sowie dem sozialen Selbst zusammen. Damit meint James nichts anderes als das Wissen über den eigenen Körper (Körper-Selbstkonzept), die eigenen Persönlichkeitseigenschaften und Einstellungen sowie das Wissen um die Sicht anderer auf die eigene Person. Cooley (1902) greift letzteren Aspekt auf und differenziert dies zu dem sogenannten *Spiegel-Selbst* beziehungsweise *looking-glass self* aus. Das beschreibt, dass das Selbst eine Summe aus wahrgenommenen Zuschreibungen anderer Personen ist, die nicht damit übereinstimmen müssen, was diese Personen tatsächlich denken: „Für den Aufbau des Selbstkonzepts kommt es also Cooley zufolge nicht darauf an, was Personen über mich denken, sondern darauf, wovon ich überzeugt bin, dass sie es denken.“ (Lohaus et al., 2010, S. 165f)

Auch Mead (1934) greift die Überlegungen von James auf und unterteilt das *Selbst* in *Me* und *I*. Ersteres beschreibt, wie das Selbst durch äußere beziehungsweise soziale Aspekte beeinflusst wird und zweiteres stellt die tatsächliche Persönlichkeit dar. Die Psychoanalyse (vor allem Freud) bietet eine andere Sichtweise auf das Selbst, das sich durch innerpsychische Konflikte und der Qualität deren Lösung entwickelt. (Lohaus et al., 2010) In diesem Gebiet wird oftmals statt vom Selbst von der sogenannten Identität gesprochen. Dies zu definieren ist jedoch schwierig, da die Identität sehr vielschichtig ist und in den verschiedenen wissenschaftlichen Bereichen von unterschiedlichen Seiten beleuchtet wird. Rodgers und Scott (2008) beschreiben in ihrem Artikel vier Grundannahmen derzeitiger Vorstellungen der Identität:

- Identität hängt von sozialen, kulturellen, politischen und historischen Kontexten ab und wird von diesen geformt.
- Identität wird in Verbindung mit anderen gebildet und beinhaltet Emotionen.
- Identität „[...]is shifting, unstable and multiple.“(S. 733)

- Identität beinhaltet die Konstruktion und Rekonstruktion von Sinn durch Geschichten im Laufe der Zeit.

Die Bildung der Identität ist nach Erikson (1974) eine der essenziellen Aufgaben in der Entwicklung des Jugendalters, jedoch streckt sich dieser Prozess über das gesamte Leben. (Lohaus et al., 2010) Marcia (1980) baut darauf auf und nennt verschiedene Identitätsstadien, die unterschieden werden, abhängig davon, ob gerade ein Entscheidungszeitraum (Krise) besteht oder nicht (siehe Tabelle 4.1). Die diffuse Identität beschreibt ein Stadium, indem die Person keine klare Richtung hat. In der übernommenen Identität³ sind die Ziele und Werte, denen nachgegangen wird, nicht selbst ausgewählt, sondern von den Eltern. Wichtig in diesem Modell ist die sogenannte Krise. Im Moratorium werden verschiedene Ziele und Werte miteinander verglichen. Im letzten Stadium – die erarbeitete Identität – hat die Person nun die eigene Identität gefunden und fühlt sich gegenüber bestimmten Zielen und Werten verpflichtet. (Marcia, 1980; Lohaus et al., 2010)

	Krise	Keine Krise
Innere Verpflichtung	Erarbeitete Identität	Übernommene Identität
Keine innere Verpflichtung	Moratorium	Diffuse Identität

Tabelle 4.1

Vier Identitätsstadien nach Marcia (modifiziert nach Lohaus et al., 2010, S. 166)

Hier kann nur ein kurzer Überblick über den Begriff Selbstkonzept und dessen Entwicklung gegeben werden, da eine genauere Betrachtung den Rahmen dieser Arbeit sprengen und für eine fachdidaktische Arbeit zu sehr in die psychologische Ebene gehen würde. Im folgenden Kapitel wird nun der Bogen zur Schule gespannt und das schulische beziehungsweise akademische Selbstkonzept beschrieben.

4.1.2 Akademisches Selbstkonzept

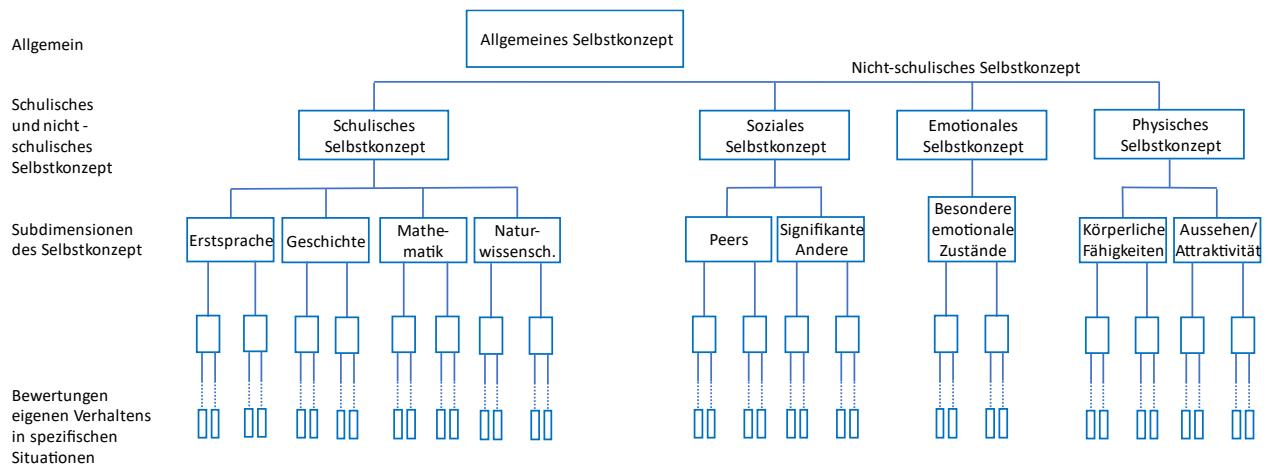
In Bezug auf die pädagogische Psychologie ist das Selbstkonzept ein weit erforschtes Thema. Dies begründet sich damit, dass ein positives Selbstkonzept nicht nur für das psychische Wohlempfinden, sondern auch für die eigene Leistungsfähigkeit von Kindern beziehungsweise Jugendlichen dienlich ist. Da es im schulischen Kontext um die Selbstbeschreibung in einem spezifischen Gebiet geht, handelt es sich beim schulischen Selbstkonzept um ein bereichsspezifisches

³ Originalbegriff: Foreclosure

Selbstkonzept. Das schulische Selbstkonzept hat in der Forschung verschiedene Bezeichnungen, die von manchen Forscher*innen aufgrund der Ähnlichkeiten synonym verwendet werden. Bei genauerer Betrachtung unterscheiden sich diese allerdings in kleinen Nuancen. Einerseits ist der Begriff Fähigkeitsselbstkonzept ein weit verbreiteter, der „den Aspekt des wahrgenommenen Leistungsstands (die Performanz)“ (Möller & Trautwein, 2009, S. 181) stärker hervorhebt. Andererseits wird oft der Begriff Selbstkonzept der Begabung verwendet, der im Gegensatz zum vorigen Begriff die „potenzielle Leistung (bzw. die Anlagen, die eine Person besitzt)“ (Möller & Trautwein, 2009, S.181) beachtet. In dieser Arbeit wird der Begriff Selbstkonzept verwendet. Im schulischen Alltag sind Schüler*innen tagtäglich mit der Tatsache konfrontiert, dass ihre Handlungen beziehungsweise Fähigkeiten in der Schule Konsequenzen – wie etwa Noten – mit sich bringen. Aufgrund dessen lässt sich schließen, dass das schulische Selbstkonzept auch evaluative Komponenten enthält. (Möller & Trautwein, 2009) Im Folgenden werden nun zwei verschiedene Selbstkonzeptmodelle vorgestellt.

4.1.2.1 *Hierarchisches Selbstkonzeptmodell von Shavelson et al.*

Shavelson et al. (1976) beschrieben das Selbstkonzept mit sieben Aspekten: „Seven features can be identified as critical to the construct definition. Self-concept may be described as: organized, multifaceted, hierarchical, stable, developmental, evaluative, differentiable.“ (S. 411) Essentiell an diesem Selbstkonzeptmodell ist die hierarchische und multidimensionale Struktur. Abbildung 4.1 zeigt das hierarchische Selbstkonzeptmodell von Shavelson et al. (1976).

**Abbildung 4.1**

Das hierarchische Selbstkonzeptmodell (modifiziert nach Möller & Trautwein, 2009, S. 182)

In dieser Abbildung lässt sich die Multidimensionalität und Hierarchie klar erkennen. Das allgemeine Selbstkonzept, das bereichsunspezifisch ist, ist ganz an der Spitze der Hierarchie zu verorten. Eine Ebene darunter wird in das akademische und nicht-akademische Selbstkonzept unterteilt. Das akademische Selbstkonzept gliedert sich wiederum in unterschiedliche Schulfächer wie die Erstsprache als gesetzlich verankerte Sprache (in Österreich Deutsch), Geschichte, Mathematik und Naturwissenschaften. Es ist demzufolge ein bestimmter Teil des Selbstkonzepts einer Person. (Shavelson et al., 1976) Nach Goetz et al. (2010) ist das Selbstkonzept in Bezug auf die Schule auch in noch spezifischeren Bereichen leicht beurteilbar. Als Beispiel wird der Bereich Mathematik genannt, mit den Subbereichen Geometrie, Algebra und Statistik. Beim nicht-akademischen Selbstkonzept wird zuerst zwischen einem sozialen, emotionalen und physischen Selbstkonzept unterschieden. Nach unten hin erfolgt eine zunehmende inhaltliche Ausdifferenzierung. Shavelson et al. bezeichnen das allgemeine Selbstkonzept als stabil. Steigt man in der Hierarchie jedoch ab, erkennt man, dass das Selbstkonzept von speziellen Situationen abhängt und dadurch instabiler wird. (Shavelson et al., 1976)

In einer Studie von Marsh und Shavelson (1985) wurde das hierarchische und multidimensionale Selbstkonzeptmodell von Shavelson et al. (1976) überprüft. Dazu füllten 662 Studierende unterschiedlicher Semester das Self-Description Questionnaire (SDQ) aus. Die Studie (Marsh & Shavelson, 1985) ergab, dass eine hierarchische Struktur nachweisbar ist. Diese ist jedoch komplexer als ursprünglich von Shavelson et al. im Jahr 1976 angenommen. Weiters gingen die Forscher aufgrund der Studie davon aus, dass es eine klare Trennung zwischen dem verbalen und mathematischen Selbstkonzept gibt und diese sich nicht in einem gemeinsamen akademischen

Selbstkonzept befinden können. Dies bestätigte eine spätere Studie (Marsh et al., 1988), die nochmals untermalte, wie wenig bis gar nicht das verbale und mathematische Selbstkonzept miteinander korrelieren. Dies führte zur Überarbeitung des Modells. Das ursprüngliche akademische Selbstkonzept wurde somit in das verbale-akademische und das mathematische-akademische Selbstkonzept aufgeteilt. Dieses revidierte Modell wird in Abbildung 4.2 gezeigt. Während das mathematische-akademische Selbstkonzept naturwissenschaftliche Fächer beinhaltet, umfasst das verbale-akademische Selbstkonzept vor allem Fremdsprachen und die Erstsprache. Dieses Modell stellte die Basis für viele weitere Forschungen, die sich vor allem mit der Korrelation zwischen dem mathematischen und verbalen Selbstkonzept beschäftigten. (Möller & Trautwein, 2009) Die Gründe, warum das mathematische-akademische und das verbale-akademische Selbstkonzept im revidierten hierarchischen Selbstkonzeptmodell getrennt wurden, werden im Internal/External-Frame-of-Reference Modell von Marsh (1986) beschrieben.

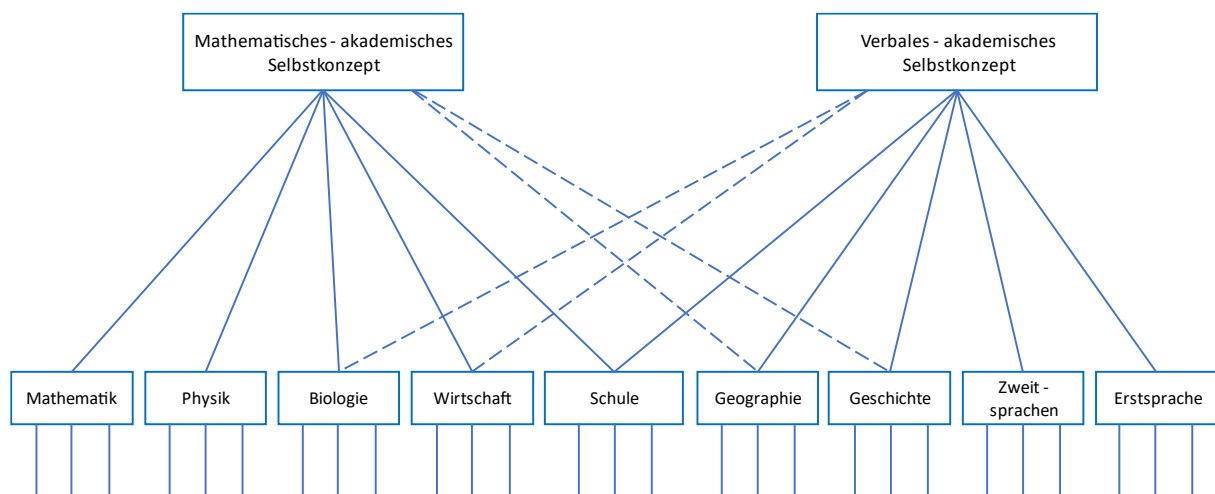


Abbildung 4.2

Struktur des revidierten akademischen Selbstkonzeptmodells (modifiziert nach Marsh et al., 1988)

4.1.2.2 Integratives Selbstkonzeptmodell

Marsh et al. (2018) schlagen in ihrem Artikel ein neues integratives akademisches Selbstkonzeptmodell vor. Dieses vereint drei essenzielle Theorien der Forschung des akademischen Selbstkonzepts: Internal/External Frame of Reference Modell (I/E), Reciprocal Effect Modell (REM) und den Big-Fish-Little-Pond Effekt (BFLPE). Ebenso werden Perspektiven zur Entwicklung implementiert. Obwohl alle drei Theorien für die Forschung relevant sind, wurden diese bis zu dieser Studie nur separat betrachtet. Dadurch wurde zuvor nicht erforscht, wie diese Modelle untereinander interagieren und ob und inwiefern diese Modelle einander beeinflussen. Das in diesem Artikel

vorgestellte integrative akademische Selbstkonzeptmodell erforscht diese drei Effekte, die auf früherer Forschung basieren, erstmals gemeinsam. Ein weiteres Ziel dieser Studie von Marsh et al. (2018) war es, zu erforschen wie sich die Ansicht über die eigene Kompetenz in der Zeit der Jugend entwickelt. Hierzu werden die drei oben genannten Modelle berücksichtigt, die im Weiteren genauer erläutert werden, bevor im Anschluss das integrative akademische Selbstkonzeptmodell nähergebracht wird.

Das erste Modell ist das sogenannte *Internal/External Frame of Reference Modell (I/E-Modell)*, das beschreibt, wie verbale und mathematische Selbstkonzepte gebildet werden. In der deutschsprachigen Literatur findet es sich auch unter dem Begriff Bezugsrahmenmodell. Bereits im Jahr 1983 konnten Marsh et al. zeigen, dass die mathematische Leistung mit dem mathematischen Selbstkonzept korreliert. Weiters konnten Hansford und Hattie (1982) in ihrer Metaanalyse herausfinden, dass ein höherer Zusammenhang zwischen Leistung und akademischem Selbstkonzept als zwischen Leistung und nicht-akademischem Selbstkonzept besteht. Ähnliches stellten Marsh et al. (1983) fest. Das Internal/External Frame of Reference Modell liefert nun eine theoretische Erklärung für die Tatsache, dass mathematische und verbale Leistungen stark zusammenhängen, die zugehörigen akademischen Selbstkonzepte jedoch kaum. Marsh (1986) möchte dieses Modell mit seiner Studie sowohl theoretisch als auch empirisch stützen. Auf Basis dieses Modells zeigt sich, dass das akademische Selbstkonzept in Schulfächern auf zwei Arten gebildet wird. Einerseits ist ein externaler Bezug Teil der Bildung des akademischen Selbstkonzepts. Dies entspricht einem sozialen Vergleich. Eine Person vergleicht die eigenen mathematischen beziehungsweise verbalen Fähigkeiten mit jenen anderer Personen in ihrem Bezugsrahmen. (Marsh et al., 2018) Andererseits spielt der interne Vergleich ebenfalls eine tragende Rolle. Personen vergleichen ihre von sich wahrgenommen Fähigkeiten im mathematischen Bereich mit den Fähigkeiten im verbalen Bereich und umgekehrt. Auf dieser Grundlage wird das eigene akademische Selbstkonzept in einem bestimmten Schulfach gebildet. Möller und Trautwein (2009) beschreiben dazu vier Prozesse. Schüler*innen vergleichen ihre Leistungen in verschiedenen Schulfächern mit denen ihrer Mitschüler*innen (externaler Bezugsrahmen). Im nächsten Schritt geschieht der soziale/interindividuelle Vergleich. Schüler*innen mit niedriger Fähigkeit in diesem Schulfach bilden ein niedriges Selbstkonzept aus und jene mit hohen Fähigkeiten ein hohes Selbstkonzept. Nun wird ein internaler Bezugsrahmen hinzugezogen. Schüler*innen ziehen nun ihre eigenen Fähigkeiten in anderen (verbalen) Fächern zum Vergleich mit ihren mathematischen Fähigkeiten heran. Zuletzt führt nun der dimensionale/intraindividuelle Vergleich dazu, dass Schüler*innen ihr Selbstkonzept in einem Schulfach bilden. Es kann dazu kommen, dass Schüler*innen mit hohen mathematischen Fähigkeiten dadurch ihr verbales-akademisches Selbstkonzept abwerten oder, dass Schüler*innen mit niedriger mathematischer

Fähigkeit im sozialen Vergleich ihr verbales-akademisches Selbstkonzept aufwerten. Erkennbar ist hier ein negativer Effekt von verbalen Leistungen auf das mathematische Selbstkonzept und von mathematischen Leistungen auf das verbale Selbstkonzept. „For example, a high Math self-concept will be more likely when math skills are good (the external comparison) and when math skills are better than verbal skills (the internal comparison).“ (Marsh, 1986, S. 134) Ausschlaggebend für ein hohes mathematisches Selbstkonzept – vorausgesetzt höhere mathematischen Fähigkeiten sind vorhanden – ist der Unterschied zwischen mathematischen und verbalen Fähigkeiten. (Marsh, 1986) Es geht also um den Kontrasteffekt. (Möller & Trautwein, 2009) Diese Zusammenhänge werden in Abbildung 4.3 grafisch dargestellt. Erkennbar ist eine positive Korrelation zwischen den schulischen Leistungen in einem bestimmten Schulfach und den zugehörigen schulischen Selbstkonzepten. Ein positiver Einfluss von der mathematischen Leistung auf das mathematische Selbstkonzept ist nun beispielsweise erkennbar. (Möller & Trautwein, 2009) In folgender Abbildung 4.3 wird gezeigt, wie die einzelnen Faktoren einander beeinflussen. Erkennbar ist, dass der positive Effekt der mathematischen/verbalen Leistung auf das mathematische/verbale Selbstkonzept größer ist als der negative Effekt der verbalen Leistung auf das mathematische Selbstkonzept und der mathematischen Leistung auf das verbale Selbstkonzept.

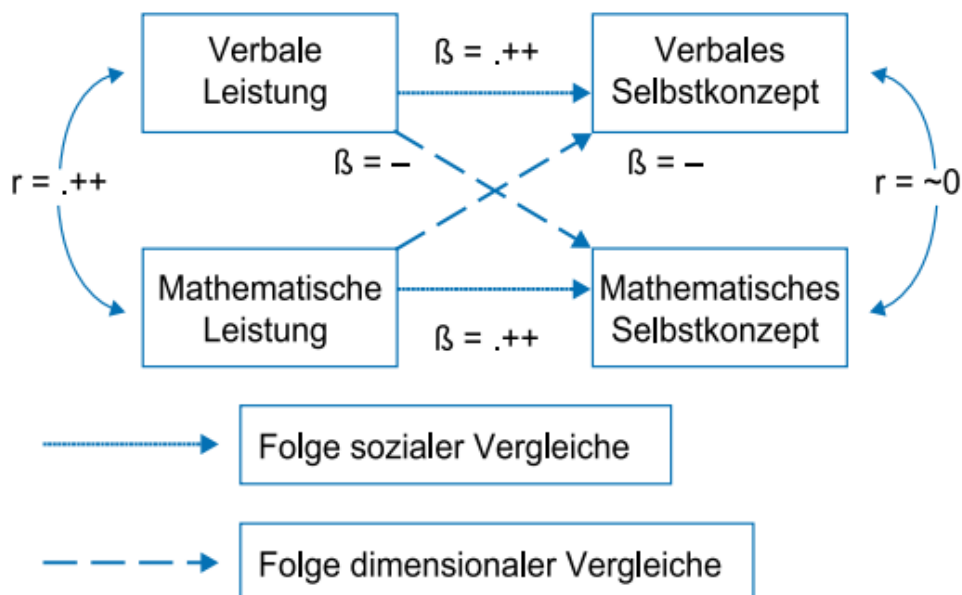


Abbildung 4.3

Internal/External Frame of Reference Modell: Zusammenhang zwischen verbalen Fähigkeiten, mathematischen Fähigkeiten, verbalem Selbstkonzept und mathematischem Selbstkonzept (Möller & Trautwein, 2009, S. 194)

Klar ist, dass dieses Modell seinen Anklang im schulischen Alltag findet. Dieses Modell besagt, dass das akademische Selbstkonzept in einem bestimmten Schulfach auf Basis von externalem und internalem Vergleich gebildet wird. Im Folgenden wird dies beispielhaft für das mathematische-akademische Selbstkonzept einer Schülerin betrachtet. Diese Schülerin schätzt ihre mathematischen und verbalen Fähigkeiten als unterdurchschnittlich ein. In Mathematik sind ihre Leistungen jedoch höher als im verbalen oder in anderen Schulfächern. Das mathematische-akademische Selbstkonzept wird nun auf Grundlage vom externalen und internalen Vergleich gebildet. Die mathematischen Fähigkeiten ihrer Mitschüler*innen sind höher als ihre eigenen – hier geschieht der externe Vergleich. Der interne Vergleich basiert auf ihren Leistungen in anderen Schulfächern, wobei in diesem Beispiel angenommen wird, dass sie in Mathematik höhere Fähigkeiten hat als in anderen Schulfächern. Je nachdem wie diese zwei Faktoren gewichtet werden, wird ein durchschnittliches oder gar überdurchschnittliches mathematisch-akademisches Selbstkonzept gebildet. Legt man das nun auf alle Schulfächer um, ist es demnach trotz niedriger Fähigkeiten in einem Schulfach möglich, ein durchschnittliches oder gar überdurchschnittliches Selbstkonzept darin zu haben. (Marsh, 1986)

Lange waren Wissenschaftler*innen (siehe etwa Calsyn und Kenny, 1977) der Überzeugung, dass die frühere Leistung das zukünftige Selbstkonzept (*skill development model*) oder das frühere Selbstkonzept die zukünftige Leistung (*self-enhancement model*) bestärkt. Ein wechselseitiger Zusammenhang wurde demnach ausgeschlossen. Sowohl die pädagogische Theorie als auch die pädagogische Praxis wurden von diesen entgegengesetzten Ansichten deutlich beeinflusst. Gehen Lehrende vom ersteren Modell aus, ist es sinnvoll, sich auf die Verbesserung der Leistung zu fokussieren. Umgekehrt – folgen Lehrende dem Modell, dass das frühere Selbstkonzept die spätere Leistung beeinflusst, so ist ein Arbeiten an der Verbesserung des Selbstkonzepts von Nutzen. (Seaton et al., 2015) Marsh und Craven (2005) argumentieren auf Basis einer Studie (Marsh & Yeung, 1997), dass die beiden oben genannten Modelle zu einfach seien und bestärkten damit das *Reciprocal Effects Modell (REM)*. Dieses Modell, das auch unter dem Begriff Wechselwirkungsmodell zu finden ist und ebenso einen Einfluss auf die Bildung des akademischen Selbstkonzepts hat, zeigt, dass diese ursprünglichen Annahmen von einer einseitigen Korrelation nicht korrekt sind. Das Reciprocal Effekt Modell beschreibt den kausalen wechselseitigen Zusammenhang zwischen der Leistung und dem akademischen Selbstkonzept: „The REM predicts that academic self-concept and academic achievement are reciprocally related in such a way that previous academic self-concept is positively related to future academic achievement, and previous academic achievement is positively related to future academic self-concept.“ (Seaton et al., 2015, S. 1) Diesem Modell nach ist der zu wählende Ansatz jener, bei dem sowohl das akademische Selbstkonzept als auch die akademische Leistung im Fokus sind und gleichermaßen thematisiert werden. (Seaton et al., 2015) Marsh (1990) halten fest,

dass das akademische Selbstkonzept und die Leistung zugleich Ursache und Wirkung voneinander sind. Nicht unerwähnt bleiben darf eine Metaanalyse von Valentine et al. (2004), die die Ergebnisse von 60 Längsschnittstudien zum Zusammenhang zwischen *self-beliefs* und Leistung vereinte. In dieser Metaanalyse wurden das Selbstkonzept, das Selbstwertgefühl und die Selbstwirksamkeit unter dem Begriff *self-beliefs* zusammengefasst wurde. Die Ergebnisse dieser Studie unterstreichen durchwegs das Reciprocal Effects Modell. Valentine und DuBois (2005) nehmen Bezug auf die Ergebnisse der Metaanalyse aus dem Jahr 2004 und halten fest, dass Lehrende gleichzeitig angehalten sind, die Leistung ihrer Schüler*innen zu steigern und die Wichtigkeit des akademischen *self-beliefs* zu beachten.

Zuletzt wird der *Big-Fish-Little-Pond Effekt* (BFLPE) vorgestellt. Dieser besagt, dass „[...] equally able students have lower academic self-concepts when attending schools where the average ability levels of classmates is high, and higher academic self-concepts when attending schools where the school-average ability is low.“ (Marsh, 2005, S. 119) Es wird also angenommen, dass ein Teil der Bildung des akademischen Selbstkonzepts darin besteht, dass Schüler*innen ihre eigene Leistung mit jener anderer Schüler*innen vergleichen. Vergleichen sich Schüler*innen mit anderen, die eine höhere Leistung erbringen, resultiert dies in einem niedrigen akademischen Selbstkonzept. Umgekehrt folgt ein höheres akademisches Selbstkonzept, wenn sich Schüler*innen mit anderen vergleichen, die weniger Leistung aufweisen als sie selbst. Daraus lässt sich schließen, dass das eigene akademische Selbstkonzept nicht nur von der eigenen Leistung abhängt, sondern auch von der Leistung des Umfelds beziehungsweise der der anderen Schüler*innen der Schule. (Marsh, 2005) Basierend auf denselben objektiven Leistungen variiert das Selbstkonzept abhängig vom Bezugsrahmen. Die Selbstwahrnehmung kann demnach nur angemessen verstanden werden, wenn der Bezugsrahmen beachtet wird. Dies beeinflusst im Weiteren zukünftige Leistungen und Verhaltensweisen. (Marsh et al., 2008) Forscher*innen fragten sich, ob der Big-Fish-Little-Pond-Effekt bei allen Leistungsniveaus auftritt. Marsh und Rowe konnten 1996 an einer großen Datenbank aus den USA zeigen, dass der Big-Fish-Little-Pond-Effekt bei allen Schüler*innen unabhängig vom Leistungsniveau erkennbar war. Die Größe des Effekts schwankte nur minimal mit der individuellen Leistung. Ähnliches konnten Marsh et al. (1995) in zwei Studien zeigen. Es konnte kein signifikanter Zusammenhang zwischen der Größe des Effekts und dem individuellen Leistungsniveau erhoben werden. Dies zeigt, dass sowohl bei Schüler*innen mit hoher Leistung als auch bei jenen mit geringer Leistung der Big-Fish-Little-Pond-Effekt auftritt.

Umgelegt auf den Mathematikunterricht bedeutet das, dass zwei Schüler*innen mit der gleichen Leistung aus zwei verschiedenen Klassen aufgrund der Leistung der Mitschüler*innen in der

Klasse ein unterschiedliches akademisches Selbstkonzept entwickeln können. Ist die Leistung der Mitschüler*innen in Mathematik in einer Klasse höher, so wird die Person ein eher niedrigeres akademisches Selbstkonzept bilden. Ist die Leistung der Mitschüler*innen jedoch niedriger, so wird die Person ein höheres akademisches Selbstkonzept entwickeln, obwohl die Leistung der beiden Schüler*innen gleich ist. Das nachstehende Flussdiagramm (Abbildung 4.4) zeigt den Effekt von individueller Leistung und durchschnittlicher Leistung auf das individuelle akademische Selbstkonzept.

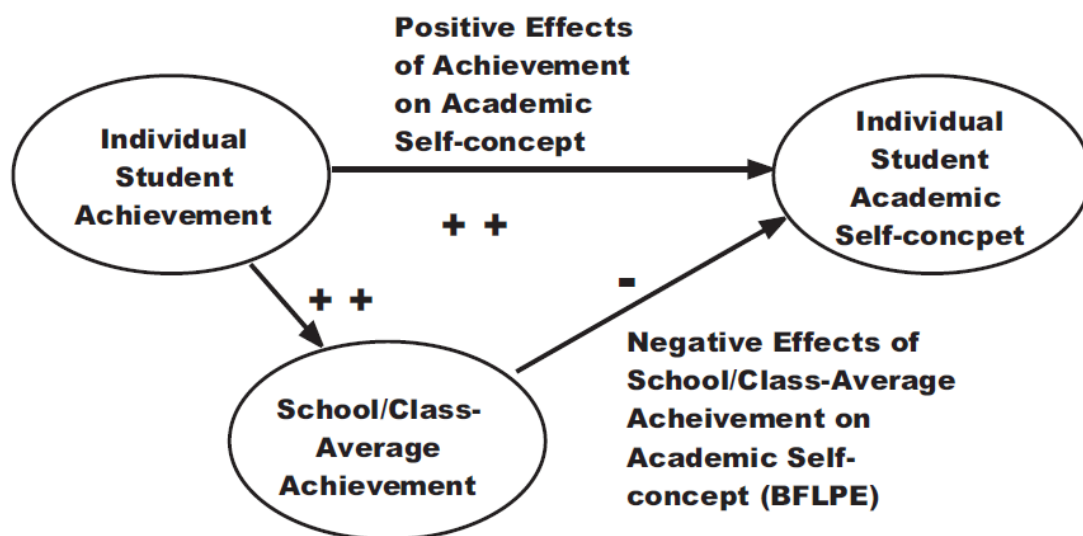


Abbildung 4.4

Big-fish-little-pond effect (Marsh, 2005, S. 120)

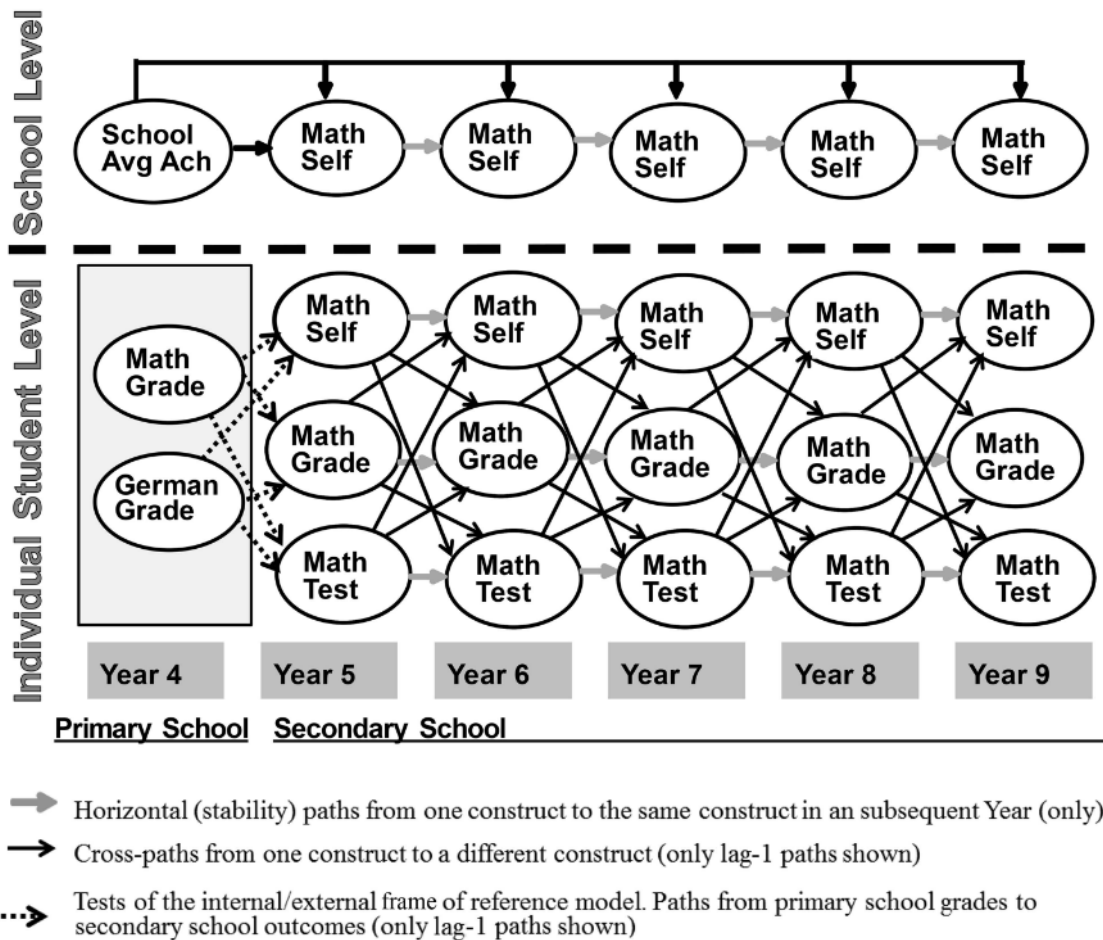
Nach Betrachtung dieser drei Modelle stellt sich die Frage, welche Bedeutung diese für den Mathematikunterricht beziehungsweise für Unterricht generell haben. Marsh (2005) schreibt in seinem Artikel, dass es einen Umschwung in der Psychologie gab und dass der Fokus heutzutage viel mehr darauf liegt „how normal individuals can get the most from life“ (S. 119). Diese Denkweise findet sich auch im schulischen Kontext wieder. Menschen erreichen mehr, wenn sie sich kompetent fühlen in dem, was sie tun und wenn sie positiv über sich selbst denken. Auch ein positives Selbstkonzept hat vor allem in schulischen Settings Anklang gefunden. Vor allem wenn das Selbstkonzept untergraben wird, wirkt sich dies negativ auf die Leistung aus. (Marsh, 2005) Fraglich ist, inwiefern das Schulsystem, die Schule und der Unterricht die Entwicklung des akademischen Selbstkonzepts beeinflussen und inwiefern Lehrpersonen darauf einwirken können. Es gibt in der Tat Studien, die zeigen, dass das Lernklima beziehungsweise soziale Klima in einer Klasse Einfluss auf das akademische

Selbstkonzept nehmen. Wichtig ist, dass Lehrpersonen Leistungen nicht nur auf Basis des sozialen Vergleichs bewerten (Kollektivnorm (Vierlinger, 1999) oder soziale Bezugsnorm (Rheinberg, 2014)). Dies würde einen sehr starken sozialen Bezugsrahmen schaffen, wodurch der intraindividuelle Bezug in den Hintergrund rückt und darunter vor allem schwächere Schüler*innen leiden. Diese haben die Möglichkeit höhere Selbstkonzepte zu entwickeln, wenn Lehrpersonen zusätzlich individuelle Bezugsrahmen heranziehen und dadurch die individuellen Leistungszuwächse anerkennen. Dies allein reicht jedoch nicht aus, damit alle Schüler*innen hohe Selbstkonzepte entwickeln. Weiters sind Leistungen in einer Klasse sehr heterogen. Viele Schüler*innen vergleichen sich mit „besseren“ Mitschüler*innen und dies wirkt sich negativ auf das Selbstkonzept aus. Um Schüler*innen in der Überzeugung ihrer Fähigkeiten zu unterstützen, ist neben dem Selbstkonzept auch die Selbstwirksamkeitsüberzeugung zu beachten. (Möller & Trautwein, 2009) Betrachtet man das Wechselwirkungsmodell, so ist unumstritten, dass eine simultane Thematisierung des akademischen Selbstkonzepts und der Leistung notwendig ist. (Marsh & Craven, 2005) Basierend auf den bisher vorgestellten Modellen lässt sich ebenso festhalten, dass die Schule an sich einen starken Einfluss auf das akademische Selbstkonzept von Schüler*innen haben kann. (Möller & Trautwein, 2009)

Nach Erläuterung der drei Modelle (Internal/External Frame of Reference Modell, Reciprocal-Effects-Modell, Big-Fish-Little-Pond-Effekt) wird nun das integrative akademische Selbstkonzeptmodell dargestellt. Dabei werden die drei Modelle vereint und in einer Längsschnittstudie untersucht. Für die Studie wurden Daten von 3.370 Schüler*innen aus 42 deutschen Schulen aus Bayern gesammelt. Diese beruhten auf dem Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA). Hinzugezogen wurden Daten beginnend mit den Noten des letzten Volksschuljahres hinweg über die weiteren 5 Schuljahre aus den Schulfächern Mathematik und Deutsch. Für die weitere Schullaufbahn nach der Volksschule gibt es in Deutschland je nach Leistung drei verschiedene Schultypen (Gymnasium, Realschule, Hauptschule). Ziel war es zu zeigen, dass die drei Modelle der Selbstkonzeptbildung, das Entwicklungsgleichgewicht⁴ und die passende statistische Methodik in einem Modell vereint werden können. Mit Entwicklungsgleichgewicht wird hierbei der Zustand des Gleichgewichts zwischen denkbar ausgleichenden und entgegengesetzten Kräften bezeichnet. Dafür beantworteten die Schüler*innen am Ende jedes Schuljahres einen Fragebogen, bei dem geschulte externe Testadministrator*innen an die Schule kamen. Ebenso wurden Schulnoten und die Beziehung zum mathematischen Selbstkonzept herangezogen. Die Ergebnisse der drei Modelle (I/E-Modell, REM, BFLPE) waren über die Jahre der Studie konsistent. Der Auswertung der Studie zufolge zeigten die Schulnoten in

⁴ freie Übersetzung der Autorin, Originalbegriff: developmental equilibrium

Mathematik einen positiven Effekt auf das mathematische Selbstkonzept, der Effekt der Deutschnoten war jedoch negativ. Eine weitere Begründung dessen wird in diesem Artikel (Marsh et al., 2018) nicht angeführt. Weiters besteht ein wechselseitiger Einfluss zwischen dem mathematischen Selbstkonzept, Testergebnissen und Schulnoten. Ein weiteres Resultat der Studie war, dass der Big-Fish-Little-Pond sehr robust ist und über die 5 Jahre einen fortdauernden negativen Effekt hatte. Ein weiteres Ergebnis, das nicht unerwähnt bleiben darf, ist der wechselseitige Zusammenhang von Schulnoten und der Entwicklung des akademischen Selbstkonzepts über diese fünf Jahre. Zusammenfassend sei gesagt, dass die hier erstmals vereinten erforschten Selbstkonzeptmodelle konsistente Ergebnisse über die fünf Jahre zeigten und auch sehr robust in dieser erforschten Entwicklungsphase waren. Das unterstreicht die Annahme des Entwicklungsgleichgewichts und ebenso die Vernetzung und Komplementarität der drei Modelle. (Marsh et al., 2018)

**Abbildung 4.5**

Integratives akademisches Selbstkonzeptmodell (Marsh et al., 2018, S. 265)

Selbsteinschätzung ist klarerweise subjektiv und nicht objektiv, wodurch die Möglichkeit zur Selbstunterschätzung oder Selbstüberschätzung besteht. Diese Phänomene werden im Folgenden genauer erläutert.

4.2 Selbstüberschätzung

Die Selbstüberschätzung (*overconfidence*) ist ein weit verbreitetes Phänomen, das erstmals in den 1960ern in der Psychologie auftauchte, aber später auch in der Wirtschaftswissenschaft Anklang fand. Oskamp (1965) beschreibt die Selbstüberschätzung prägnant als Übermaß an Zuversicht in Bezug auf die Richtigkeit. Überschätzt man also die Wahrscheinlichkeit, dass eigene Entscheidungen oder Urteile richtig sind, wird von Selbstüberschätzung gesprochen. (Dunning et al., 1990, 2004) Oskamp (1965) zeigte, dass dabei die Zuversicht und die Richtigkeit einer Entscheidung weit

auseinanderliegen. Dieser Abstand ist jedoch variabel und kann sich auch verstärken, wobei hier die Zuversicht im Verlauf steigen kann, die Richtigkeit hingegen kaum.

„Selbstüberschätzung ist keine generelle Persönlichkeitseigenschaft, sondern in erheblichem Umfang kontextabhängig.“ (Glaser, 2019, S.15) Personen überschätzen sich vor allem in sehr einfachen Situationen und im Alltag, wie etwa beim Autofahren. (Glaser, 2019) Gegenteiliges behaupten Lichtenstein et al. (1977), die den Hard-Easy-Effekt identifizierten. Dieser besagt, dass sich Personen vor allem in besonders schwierigen und komplexen Situationen überschätzen. Ihrer Ansicht nach unterschätzen sich Personen bei sehr einfachen Aufgaben.

Personen vergleichen sich in den unterschiedlichsten Kontexten, wie etwa in der Schule oder in der Arbeit mit anderen Menschen in ihrem sozialen Umfeld. (Suls et al., 2002) Die Verhaltensforschung der letzten Jahrzehnte hat gezeigt, dass diese Selbsteinschätzung vorhersagbar positiver ausfällt als bei einer Einschätzung von außen. Ebenso schmeicheln sich Personen oft selbst bei Selbsteinschätzungen. (Alicke, 1985; Dunning et al., 1989; Kruger & Dunning, 1999; Chambers & Windschitl, 2004) Demnach tendieren viele Menschen dazu, sich selbst sehr positiv zu bewerten, vor allem in „subjective, socially desirable, and common dimensions“ (Myers, 2009, S.64). Die eigenen Schwächen beziehungsweise negativen Eigenschaften werden als normal angesehen, während Stärken und positive Eigenschaften als etwas Besonderes wahrgenommen werden. (Dauenheimer et al., 2002) Ebenso denken viele, dass sie negative Ereignisse weniger wahrscheinlich treffen als andere Personen ihres Alters und sie mehr erwünschte und weniger unerwünschte Charakterzüge haben als andere Personen in ihrem Alter. Diese Neigung wird in der Sozialpsychologie als above-average effect oder auch better-than-average effect beschrieben und ist einer der am stärksten aufkommenden Effekte in der Thematik des Vergleichs mit anderen. Der above-average effect ist eine spezielle Art von sozialem Vergleich, bei dem Personen ihre Verhaltensweisen oder Merkmale mit einer Norm vergleichen. Bei dieser Norm handelt es sich gewöhnlich um das durchschnittliche Ansehen in Bezug auf ein bestimmtes Merkmal oder eine Verhaltensweise in der zugehörigen Altersgruppe. (Alicke & Govorun, 2005)

Chambers und Windschitl (2004) fassen mehrere Artikel zusammen und zählen unterschiedliche Beispiele auf, bei denen Personen angaben, besser zu sein als andere. Viele Personen meinten sie seien sportlicher, organisierter, bessere Autofahrer*innen, höflicher, fairer und gar bessere Führungskräfte. Derartige Listen können in den meisten Büchern über Sozialpsychologie gefunden werden, wie etwa in Myers' (2009) Buch über Sozialpsychologie: „Compared with people in general, most people see themselves as more ethical, more competent at their job, friendlier, more intelligent, better looking, less prejudiced, healthier, and even more insightful and less biased in their

self-assessments. “ (S.64) Anhand einer Datensammlung von circa 1 Million Studierenden, die am Scholastic Assessment Test (SAT) im Jahr 1976 teilnahmen, konnte der above-average effect das erste Mal und sehr markant gezeigt werden. 60% gaben an, in Bezug auf ihre Sportlichkeit über dem Median zu liegen, 70% in Bezug auf ihre Führungsfähigkeit und sogar 85% gaben an, sie können besser mit anderen auskommen. Bei vielen anderen Studien wurde nach einem Vergleich mit dem *average* gefragt, wodurch die Möglichkeit bestand, die Resultate auf ein falsches Verständnis des Begriffs *average* zurückzuführen. Anders war das bei dieser Studie, da hier eine Selbsteinschätzung im Bezug zu einem Median gefragt war. Das macht diese Studie zu einer besonderen und oft zitierten. (College Board, 1976-1977; Alicke & Govorun, 2005) Nicht unerwähnt darf eine Studie von Kruger und Dunning (1999) an der Cornell Universität bleiben. Diese gingen von der Annahme aus, dass der above-average effect durch Defizite in der Metakognition hervorgerufen werden können. Daraufhin führten diese vier Studien mit Psychologiestudierenden durch. Den Teilnehmer*innen wurden dabei Tests vorgelegt anhand derer ihre Fähigkeit in einem bestimmten Bereich gemessen wurde. Studie 1 erfragte den Humor, Studie 2 und 4 das logische Denkvermögen und Studie 3 die englische Grammatik. Im Anschluss mussten die Teilnehmer*innen ihre Fähigkeiten und Leistung im Test im Vergleich zu den anderen Psychologiestudierenden anhand eines Prozentranges einschätzen. Zusätzlich mussten diese im jeweiligen Bereich schätzen, wie viele Aufgaben sie richtig gelöst hatten. Die Ergebnisse zeigten, dass die Studierenden angaben, mehr Humor, logisches Denkvermögen und Kenntnisse der englischen Grammatik zu haben als durchschnittliche Studierende. Kruger und Dunning (1999) konnten mit diesen Studien den above-average effect nachweisen und zeigen, dass dieser Effekt unabhängig von der tatsächlichen Leistung auftritt.

4.3 Selbstunterschätzung

Neben der oft thematisierten Selbstüberschätzung wird nun auch der Selbstunterschätzung Beachtung geschenkt. Unterschätzt eine Person ihre Fähigkeiten, so besteht die Neigung dazu, bestimmte Handlungen nicht zu tätigen, obwohl diese die Fähigkeit dazu hätte. Dadurch können Handlungen unterlassen werden, die zum Beispiel zu Lösungen beitragen könnten. (Braun, 2003) Der below-average effect beschreibt nun den sozialen Vergleich, bei dem sich Personen in Bezug auf eine Fähigkeit als systematisch schlechter einstufen als ihre Peers. (Kruger, 1999) Dieser ist nun das Gegenteil des above-average effects, der oben beschrieben wurde. Nach Kruger (1999) beruhen beide Effekte auf dem gleichen Mechanismus und derartige Beurteilungen der eigenen Fähigkeiten sind laut dem Forscher äußerst egozentrisch. Wenn sich Personen nun mit anderen vergleichen, beachten diese hauptsächlich die eigenen Fähigkeiten, also wie fähig sie selbst sind und nur

nebensächlich beziehungsweise unzureichend die Fähigkeiten der anderen Person. Das führt zu einer Überschätzung in Bereichen, in denen die absoluten Fähigkeiten eher hoch sind, beziehungsweise die Grenze zu erfolgreicher Durchführung niedrig ist und zu einer Unterschätzung in Bereichen, in denen die absolute Fähigkeit eher niedrig ist, beziehungsweise die Grenze für erfolgreiche Durchführung hoch ist. (Kruger, 1999)

Beim below-average effect, der auch worse-than-average effect genannt wird, unterschätzen Personen ihre eigenen wünschenswerten Eigenschaften und Fähigkeiten. Dieser Effekt tritt auf, wenn die Chance auf Erfolg als sehr gering eingestuft wird. In dem Artikel werden hierfür mehrere Beispiele angeführt, unter anderem die Fähigkeit stundenlang zu lernen oder Kunst zu kreieren. Werden hierbei hauptsächlich die eigenen Fähigkeiten betrachtet, so würden sich die meisten als below-average einstufen. Es ist jedoch nicht so, dass man nicht nur weil man selbst bestimmte Fähigkeiten nicht besitzt, automatisch unter dem Durchschnitt liegt. (Kruger, 1999) Kruger untersuchte diesen Effekt anhand von drei Studien. Hinzugezogen wurden dafür Studierende der Cornell Universität in New York. Die erste Studie erfragte die Selbsteinschätzung im Vergleich zu den anderen Studierenden in Bezug auf acht Fähigkeiten. Davon wurden vier aufgrund deren Einfachheit und die übrigen vier Fähigkeiten aufgrund deren Schwierigkeit gewählt. Im Anschluss mussten sie die eigene absolute Fähigkeit angeben und ebenso die der anderen Studierenden. Zuletzt mussten die Studierenden die Erwünschtheit beziehungsweise Attraktivität und die Ambiguität der acht Fähigkeiten nennen. Die Proband*innen stufen sich als unterdurchschnittlich bei den schwierigeren Fähigkeiten ein und überdurchschnittlich bei den einfachen Fähigkeiten, wobei diese in Bezug auf die Erwünschtheit nur leicht höhere Werte erzielten. (Kruger, 1999)

Es darf nicht außer Betracht gelassen werden, dass sich Selbstunterschätzung negativ auf die Motivation auswirkt. (Bandura, 1997) Klar ist auch, dass Selbstunterschätzung einen Einfluss auf die Schule hat. Womöglich entscheiden Schüler*innen auf Basis ihrer Selbstunterschätzung anders und verpassen dadurch die Möglichkeit, etwas zu finden, indem sie sehr gut sind oder etwas, das ihnen Spaß macht. Sheldrake nennt hier zum Beispiel die Wahl verschiedener Unterrichtsfächer, die nicht verpflichtet sind. (Sheldrake, 2016) Auf das österreichische Schulsystem umgelegt, könnte die Selbstunterschätzung demnach die Schulwahl beeinflussen beziehungsweise später die Studienwahl.

4.4 Dunning-Kruger Effekt

Wie in Kapitel 4.2 bereits thematisiert, tendieren Personen dazu, sehr positive Ansichten über ihre Fähigkeiten im sozialen oder intellektuellen Bereich zu haben. Die Forscher Kruger und Dunning (1999) konnten feststellen, dass Personen dazu neigen, ihre eigenen Fähigkeiten in unterschiedlichen sozialen und kognitiven Bereichen zu überschätzen. Vor allem Personen mit geringen Fähigkeiten überschätzen ihre eigenen Fähigkeiten beim Problemlösen. Im Vergleich zum *above-average effect*, bei dem es generell um das zu hohe Einschätzen der eigenen Fähigkeiten im sozialen Vergleich geht, wird hier spezifisch auf Personen mit geringen Fähigkeiten eingegangen. Der Dunning-Kruger Effekt weist nun darauf hin, dass Personen mit geringeren Fähigkeiten oder Kenntnissen diese übermäßig und unrealistisch positiv einschätzen im Vergleich zu Gleichaltrigen. (Pavel et al., 2012) Kruger und Dunning (1999) deuten an, dass das Fehlen an Fertigkeit beziehungsweise Fähigkeit selbst, Personen daran hindert, das Ausmaß ihrer Inkompetenz zu erkennen, da diese nicht zwischen Kompetenz und Inkompetenz unterscheiden können. Sie sind also nicht in der Lage, eigene Fehler zu erkennen und gehen davon aus, dass sie weniger Fehler machen als tatsächlich, wodurch es zu einer Überschätzung kommt. Ebenso können diese nicht erkennen, wenn andere Personen in gewissen Fähigkeiten überlegen sind. Dieses Phänomen wird als *dual-burden account* (siehe McIntosh et al., 2019) beziehungsweise *double burden* (siehe Dunning, 2011) bezeichnet. Die beiden Wissenschaftler (Kruger & Dunning, 1999) formulierten dies wie folgt: „[...]people who are unskilled in these domains suffer a dual burden: Not only do these people reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the metacognitive ability to realize it.“ (S. 1121) Wichtig festzuhalten ist, dass hier mit dem Begriff Inkompetenz gemeint ist, dass eine Person weniger kompetent ist als ihre Peers. (Kruger & Dunning, 1999) Während sich die vier Studien von Kruger und Dunning (1999), die an der Cornell Universität durchgeführt wurden, auf logisches Denkvermögen, englische Grammatik und Humor fokussierten, wurden seither unter anderem auch Studien im Bereich der Wirtschaft, Politik, Medizin und Luftfahrt (siehe Pavel et al., 2012) durchgeführt.

Wie bereits erwähnt, definieren viele Forscher den Dunning-Kruger Effekt speziell für Personen mit niedriger Kompetenz. (Dunning, 2011; Schlösser et al., 2013; McIntosh et al., 2019) In einigen Studien wird jedoch auch das Gegenteil in die Definition inkludiert, also, dass sich nicht nur Personen mit niedriger Kompetenz überschätzen, sondern auch Personen, die tatsächlich über Kompetenzen verfügen unterschätzen. Ein Beispiel dafür stellt eine Studie von McIntosh et al. (2019), in der gezeigt wird, dass Unterschiede in der Metakognition zum Dunning-Kruger Effekt beitragen, aber dafür nicht notwendig und auch nicht ausreichend sind. Zudem findet sich eine Analyse dessen, wie Metakognition und andere Faktoren interagieren.

In diesen Kontext fällt auch eine weitere Studie von Sanchez und Dunning (2018). In dieser Studie wurde untersucht, wie sich die Zuversicht mit der tatsächlichen Leistung von Personen, die noch am Anfang ihres Lernprozesses stehen, ändert. Ziel war es zu beschreiben, wann die Zuversicht und die Leistung zueinanderpassen und ab wann in eine Selbstüberschätzung übergegangen wird. Eine ihrer Hypothesen war folgende: „Specifically, we suggest that people begin their career at some task by being quite cautious and unconfident in their decisions, but that they quickly become overconfident—the beginner’s bubble— before going through a “correction” phase in which confidence flattens while performance continues to improve.” (S. 11) Die Ergebnisse der Studie zeigten, dass sich Lernende ganz am Beginn des Lernprozesses noch nicht überschätzen. Mit ein wenig mehr Wissen oder Fähigkeit, holt die Zuversicht schnell auf und übersteigt die entsprechende Leistung (*beginner’s bubble*). Dieser Studie lässt sich nun entnehmen, dass der Dunning-Kruger Effekt nicht für komplette Anfänger*innen gilt. Lernen diese jedoch ein wenig dazu, fallen sie in die für den Dunning-Kruger Effekt anfälligste Gruppe. Im Folgenden wird nun der Bogen zur Schule gespannt und die Selbsteinschätzung in der Schule thematisiert.

4.5 Selbsteinschätzung in der Schule

Selbsteinschätzung ist längst kein unbekannter Begriff mehr in pädagogischen Kreisen. Durchsucht man das Internet nach diesem Begriff, finden sich etliche Arbeitsblätter, die Selbsteinschätzung in verschiedensten Formen beinhalten. Oft wird danach gefragt, wie einzelne Schüler*innen ihre Leistung in bestimmten Stoffgebieten oder einzelnen darin enthaltenen Kompetenzen einschätzen. Vermehrt finden sich auch Selbsteinschätzungsbögen, bei denen Schüler*innen ihr Arbeits- und Sozialverhalten bewerten müssen.

Nicht unerwähnt darf in diesem Kontext eine Studie von John Hattie bleiben, oft auch bekannt als Hattie-Studie. In dieser Studie wurden Metaanalysen gesammelt, verglichen und schlussendlich analysiert. Hattie führte zu Beginn 800 Metaanalysen zusammen und hielt die Ergebnisse in dem Buch *Visible Learning* (2009) fest. Eine Erweiterung der Studie wird in *Visible Learning for Teachers* dargestellt, in dem John Hattie (2012) die Ergebnisse von über 900 Metaanalysen vereinte und Umsetzungsmöglichkeiten für die Erkenntnisse seiner Studie bot. Inhaltlich befasste sich Hattie in den beiden Büchern mit den Einflussfaktoren auf Leistungen von Schüler*innen. Es kristallisierten sich 150 Einflussfaktoren heraus, für die jeweils Effektgrößen (*Cohens d*) berechnet wurden. Schlussendlich wurden diese der Reihe nach sortiert in einer Rangliste (siehe Hattie, 2012, S. 302ff) dargestellt. Das in Kapitel 4.1 thematisierte Selbstkonzept findet sich auf Rang 59 dieser Liste mit

einer Effektgröße von 0,47. Die Schwelle für erstrebenswerte Effekte setzte Hattie mit 0,4 an und stellte diese mit Hilfe eines Effektstärke-Barometers, das in Abbildung 4.6 abgebildet wird, dar. (Hattie, 2012)

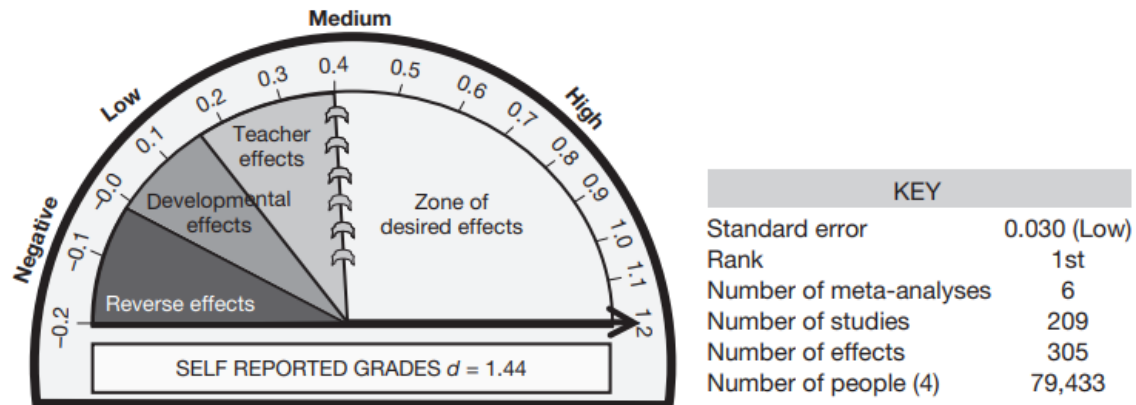


Abbildung 4.6

Effektstärke-Barometer für Selbstbeurteilung nach Hattie (2009, S. 44)

Ein besonderes Augenmerk wird hier nun auf jenen Einflussfaktor gelegt, der laut Hattie (2012) den größten Einfluss auf die Leistung von Schüler*innen nimmt. Mit einer Einflussgröße von 1,44 steht die Selbstbeurteilung⁵ an der Spitze der Rangliste (Abbildung 4.6). Als praxisbezogenes Beispiel wird angeführt, dass bevor ein Test oder eine Schularbeit geschrieben wird, Lehrende Schüler*innen auffordern sollen, eine Punktzahl oder eine Note vorherzusagen. Laut Hattie sind Schüler*innen sehr gut darin dies einzuschätzen. Der Forscher erklärt, warum diese genaue Einschätzung auch eine gewisse Problematik mit sich bringt. Oft basieren die Erwartungen der Schüler*innen auf der Grundlage, gerade genug zu tun. Dieses Prinzip wird auch *minimax* genannt, und meint damit „maximum grade return for minimal extra effort“ (Hattie, 2012, S. 137). Hier kommt nun die Aufgabe von Lehrenden zu tragen. Die Selbstbeurteilung der Schüler*innen bietet Informationen, die von Lehrer*innen genutzt werden können, um die Erwartungen der Schüler*innen zu steigern und sie dazu zu bringen ihr eigenes Potenzial zu übertreffen. Es gibt allerdings auch einzelne Gruppen, die ihre Leistung nicht gut einschätzen können. Schüler*innen mit niedrigeren Leistungen tendieren laut Hattie dazu, ihre Leistung zu unterschätzen und trauen sich nach einiger Zeit keine herausfordernden Aufgaben mehr zu. (Hattie, 2012) In diesem Zusammenhang geht es auch um die in Kapitel 3.3 thematisierte Kalibrierung. Ziel ist es, dass die Leistungen und die zugehörige Zuversicht von Schüler*innen zusammenpassen. Lehrer*innen werden an diesem Punkt aufgefordert, Schüler*innen Möglichkeiten zu bieten, ihre eigene Leistung einzuschätzen beziehungsweise vorauszusagen. Ebenso

⁵ Originalbegriff: self-reported grades – student expectations

sei es relevant, Lernziele und Beurteilungskriterien transparent zu machen und Erwartungen angemessen hoch anzusetzen. (Hattie, 2012) „Educating students to have high, challenging, appropriate expectations is among the most powerful influence in enhancing student achievement.“ (Hattie, 2012, S. 93) Diese Faktoren sind wichtig, um Zuversicht aufzubauen und dann zum Beispiel Schüler*innen mit niedrigerer Leistung wieder dazu zu bringen sich herausfordernden Aufgaben zu stellen.

Wie soeben thematisiert wurde, ist es nach Hattie (2012) von höchster Priorität, Schüler*innen die Möglichkeit zu geben, ihre eigenen Leistungen einzuschätzen. Mit den dadurch erhaltenen Informationen können Lehrer*innen im Anschluss arbeiten. Es stellt sich nun die Frage, in welcher Form diese Selbstbeurteilung, der eine Selbsteinschätzung zugrunde liegt, in den Unterricht implementiert werden kann, um so die Leistungen der Schüler*innen zu steigern. Es wurde bereits das Beispiel von Hattie genannt, bei dem er vorschlägt, dass Schüler*innen vor Tests und Schularbeiten eine Selbstbeurteilung vornehmen sollen.

Zentrales Thema dieses Textes ist Confidence-Based Learning/Assessment, bei der es nun grundlegend darum geht, die Komponente der Zuversicht in den Unterricht zu implementieren. Betrachtet man dies nun aus dem Blickwinkel der Selbsteinschätzung und behält dabei die Relevanz von Selbstbeurteilung im Hinterkopf, so lässt sich klar sagen, dass dies eine gute Möglichkeit bietet, um die Leistung von Schüler*innen zu steigern. Demnach wird hier nun diese Form der Implementierung von Selbstbeurteilung in den Unterricht vorgeschlagen. In den folgenden Kapiteln wird dies nun anhand einer Studie exemplarisch gezeigt, bei der die Zuversicht und Selbstbeurteilung in Form von Zuversichtsemojis in den Unterricht implementiert werden. Im Anschluss daran finden sich in der Diskussion, im Fazit beziehungsweise im Ausblick weitere Gedanken, wie mit den dadurch erhaltenen Informationen umgegangen werden kann und wie diese dann schlussendlich zu einer Steigerung der Zuversicht, einer höheren Genauigkeit bei der Selbsteinschätzung und zuletzt zu einer Leistungssteigerung bei Schüler*innen führen kann.

5 Methodik

Für den empirischen Teil dieser Arbeit wurde eine gemischte Methode herangezogen, um zur Beantwortung der Forschungsfrage die Vorteile der quantitativen und qualitativen Methode nutzen zu können. Dazu wurde eine Studie in einer multilingualen Klasse der Sekundarstufe 1 im Mathematikunterricht durchgeführt. Ziel war es herauszufinden, wo das Wissen der Schüler*innen zu einem bestimmten mathematischen Thema zu verorten ist (siehe „Knowledge Quadrants“, Bruno, 1987) und im selben Zuge wie genau sich diese selbst einschätzen können. Ebenso wurde untersucht, wie die Schüler*innen auf Zuversichtsemojis reagieren, um zu überprüfen, ob diese ein gutes Werkzeug zur Einbeziehung von Zuversicht in den Unterricht darstellen. Um diese Ziele zu erreichen, wurden Emojis entwickelt, die zur Angabe der Zuversicht in diesen drei Wochen eingeführt wurden. Diese Zuversichtsemojis wurden sowohl in Schulübungen, Hausübungen als auch bei einer Lernzielkontrolle verwendet. Für die Untersuchung der Verortung des Wissens wurde eine Lernzielkontrolle mit Einbezug der Zuversicht erstellt und durchgeführt. Innerhalb dieser drei Wochen wurden Feldnotizen von der Forscherin aufgezeichnet, um im Anschluss die einzelnen Unterrichtseinheiten reflektieren zu können. In Kapitel 5.4 werden Feldnotizen als Methode näher beleuchtet. Feedback seitens der Schüler*innen zu der Arbeit mit Zuversichtsemojis und der dreiwöchigen Intervention wurde über einen Feedbackbogen am Ende der Intervention eingeholt.

5.1 Setting und Teilnehmer*innen

Das Projekt umfasste drei Wochen im Mai 2022 an einer Wiener Mittelschule, an der die Forscherin zum Zeitpunkt des Projekts als Lehrerin in Mathematik und Bewegung und Sport tätig war. Die Teilnehmer*innen wurden nicht willkürlich gewählt, da die Forschung in einer der Klassen durchgeführt wurde, die die Autorin zu dem Zeitpunkt im Team-Teaching unterrichtete. Planmäßig hätte auch eine Kollegin Feldnotizen aufgezeichnet, diese ist jedoch am Tag vor Beginn der Studie erkrankt und erst nach Ende der Intervention zurückgekommen. Dadurch wurde die Klasse in dieser Zeit fast ausschließlich von der Forscherin selbst unterrichtet. Es handelte sich dabei um eine 7. Schulstufe mit 23 Schüler*innen. In dieser Klasse sind 10 Schülerinnen und 13 Schüler, die zu dem Zeitpunkt der Studie zwischen 12 und 15 Jahre alt waren. Nur wenige dieser Schüler*innen hatten Deutsch als Erstsprache und einige lernten Deutsch erst vor wenigen Jahren. Demnach wurden die Aufgaben sprachlich an die Voraussetzungen angepasst. Diese Klasse wurde gewählt, da die Problematik der fehlenden Zuversicht beziehungsweise zu hohen Zuversicht in dieser Klasse beobachtet werden konnte. Ebenso war es die einzige Klasse, die die Autorin zu diesem Zeitpunkt in Mathematik unterrichtete. Wichtig zu erwähnen ist, dass die Klassenräume über keinerlei Beamer

oder andere technische Ausstattung verfügten und somit viel analog ablaufen musste. Nachdem hier nun die Rahmenbedingungen geklärt wurden, wird im folgenden Kapitel die Entwicklung der Zuversichtsemojis erläutert.

5.2 Entwicklung Zuversichtsemojis

Bei der Wahl der *richtigen* Skala stößt man auf viele Fragen und muss die Vor- und Nachteile abwägen. Im Folgenden wird die Theorie dargelegt, die der Entscheidung der Autorin für die von ihr gewählte Skala zugrunde liegt. Wie in Kapitel 3.2 bereits thematisiert, finden sich in Studien unterschiedlichste Arten der Angabe der Zuversicht. Die Ratingskalen variieren dabei nicht nur in der Variante, sondern auch in der Anzahl der Skalenpunkte und der Skalenorientierung. Die Entscheidung für symbolische Marken fiel, da diese Art von Ratingskala vor allem bei Kindern und Jugendlichen verwendet wird. Ebenso sind diese noch anschaulicher als verbale Marken (z.B. Angabe einer Häufigkeit: nie – selten – gelegentlich – oft – immer), da diese auf einen Blick erkannt werden können. Auch bei Erhebungen mit Erwachsenen kommen diese teilweise zum Einsatz, um die Abgabe eines Urteils zu vereinfachen. Häufig kommen symbolische Marken in Form von Smileys vor. (Bortz & Döring, 2006) Dadurch können reliablere Daten erhoben werden, da die Emojis für Kinder und Jugendliche einfacher zu deuten sind. (Massey, 2022) Wichtig ist dabei auf die Äquidistanz und Skalensymmetrie zu achten, da es sonst zu ungewünschten Effekten kommen kann. (Menold & Bogner, 2015) Gerade bei längeren Fragebögen kann diese Art von Ratingskala eine Auflockerung bieten. (Bortz & Döring, 2006) Weiters verwendeten einige Forscher*innen (z.B. Davies, 2002; Barr & Burke, 2013) Deskriptoren in ihren Studien. So wie Gardner-Medwin und Gahan (2003) fiel die Entscheidung in dieser Studie auch aufgrund der möglichen unterschiedlichen Auffassungen gegen Deskriptoren. Menold & Bogner (2015) empfehlen auf Basis ihrer Recherchen die Ratingskala horizontal zu gestalten, um so den Primacy-Effekt – das Bevorzugte Wählen des linken Endes der Ratingskala – zu verringern. Es wurde ebenso erforscht, ob die Skalenorientierung, die entscheidet, ob bei der Ratingskala mit der höchsten oder niedrigsten Ausprägung begonnen wird, einen Einfluss auf den Primacy-Effekt nimmt. Da hierbei die Ergebnisse der Forschungen in Bezug auf den Primacy-Effekt beziehungsweise auf den Recency-Effekt – das Wählen der Kategorie am rechten Ende der Skala – verschieden sind, gibt es zur Wahl der Skalenorientierung keine Empfehlung. (Menold & Bogner, 2015) Wie bereits zuvor erwähnt, werden sehr unterschiedliche Arten von Ratingskalen und Anzahlen an Skalenpunkten verwendet. In Bezug auf die Wahl einer geraden oder ungeraden Anzahl an Stufen spalten sich die Meinungen der Wissenschaftler*innen. Sowohl eine gerade als auch eine ungerade Anzahl an Skalenpunkten bringt Vorteile mit sich. Bei der Wahl einer geradzahligen Ratingskala wird keine Mittelkategorie angeboten, wodurch eine Entscheidung in eine Richtung

gefordert wird. Dies wird empfohlen, wenn man Verzerrungen durch eine vermehrte Tendenz zur Mitte – das häufige Wählen der Mittelkategorie – verhindern möchte. Wird eine ungeradzahlige Ratingskala gewählt, so wird eine Mittelkategorie angeboten. Diese Mittelkategorie ist jedoch häufig schwierig zu interpretieren. Diese Problematik findet sich unter dem Begriff Ambivalenz-Indifferenz-Problem. Das Wählen der Kategorie kann demnach entweder bedeuten, dass keine dezidierte Meinung vertreten wird oder dass die Meinung zwiespältig ist. (Bortz & Döring, 2006) Menold und Bogner (2015) beschreiben in ihrem Text weitere mögliche Fehlerquellen in Bezug auf die Mittelkategorie. Diese Mittelkategorie kommt jenen Befragten entgegen, die den kognitiven Aufwand durch die Wahl dieser Kategorie reduzieren möchten und das Thema als unwichtig oder uninteressant einstufen. Demnach entspräche dies nicht ihrer tatsächlichen Meinung oder Entscheidung in Bezug auf die Fragestellungen. Im Gegensatz dazu gibt es tatsächlich Befragte, die eine mittlere beziehungsweise neutrale Meinung zu der Thematik haben. Durch das Weglassen einer Mittelkategorie können jene ihre tatsächliche Einstellung nicht angeben und werden dadurch zur Wahl einer Tendenz gedrängt. Problematisch ist auch, dass die Mittelkategorie in manchen Fällen aus sozialer Erwünschtheit gewählt wird, ohne dass eine eigene Einstellung vorhanden ist. Obwohl viele Gründe gegen eine Mittelkategorie sprechen, wird diese häufig trotzdem empfohlen. (Menold & Bogner, 2015) Ein weiteres Kriterium bei Ratingskalen ist die Skalenbreite. Diese bestimmt die Anzahl an Skalenpunkten. Einerseits führen zu breite Skalen zu Überforderung von Befragten, andererseits werden zu enge Skalen als zu wenig differenziert beschrieben. (Porst, 2008) Menold und Bogner (2015) empfehlen auf Basis einiger Studien, sich für 5- bis 7-stufige Skalen zu entscheiden.

Aufgrund dieser theoretischen Untermauerung wurden in dieser Studie symbolische Marken mit vier absteigenden Skalenpunkten verwendet. Abbildung 5.1 zeigt die von der Autorin gewählte und entwickelte Skala, die aufgrund der Thematik Zuversichtsemojis genannt wurde. Die Wahl für symbolische Marken fiel aufgrund der auflockernden Art und dem vielfachen Einsatz in Studien mit Kindern und Jugendlichen. Da auf Basis der derzeitigen Forschung keine Empfehlung der Skalenorientierung gegeben wird, fiel die Entscheidung aufgrund der möglicherweise motivierenden Wirkung auf eine absteigende Ratingskala. Es wurde eine gerade Anzahl an Skalenpunkten gewählt, da dadurch verhindert wird, dass Schüler*innen aus Desinteresse die Mittelkategorie wählen. Ebenso erscheint die Mittelkategorie aus Sicht der Autorin inhaltlich nicht sinnvoll, da es um die Angabe der Zuversicht über eine gegebene Antwort geht. Demnach könnte man vereinfacht folgende Stufen nennen, die im Weiteren genauer erläutert werden: sehr zuversichtlich – zuversichtlich – wenig zuversichtlich – gar nicht zuversichtlich. Die Angabe einer neutralen Skala erscheint in diesem Fall nicht sinnvoll, da es um konkrete Aufgaben geht und die Einschätzung in Bezug auf die Richtigkeit einer Antwort. Nach der Empfehlung für die Skalenbreite wäre eine 6-stufige Skala zu wählen. Eine

6-stufige Skala erscheint in Bezug auf das Thema und aufgrund des Alters der Proband*innen zu differenziert. Daraufhin wurde entgegen der Empfehlung von Menold und Bogner (2015) eine 4-stufige Skala verwendet. Die oben angeführten Argumente und Empfehlungen beziehen sich vor allem auf Studien, die Fragebögen nutzen. Da in dieser Studie die Skala in Form von Zuversichtsemojis allerdings in Lernsequenzen eingebettet wird und nicht unbedingt in Form eines Fragebogens, lässt sich aus pädagogischer Perspektive trotz der vielen Empfehlungen sagen, dass eine 4-stufige Skala als passend erscheint.

Bei der grafischen Gestaltung wurde darauf geachtet, die einzelnen Stufen möglichst gut zu repräsentierten. Das rechte Ende der Skala sollte dabei keinesfalls negativ wirken, da eine Angabe der beiden rechten beiden Skalenpunkte sogar sehr positiv sein kann in Hinblick auf die *knowledge quadrants* von Bruno (1987). In Kapitel 6 werden, die einzelnen Emojis näher erläutert, da diese im Rahmen der Studie mit den Schüler*innen gemeinsam erarbeitet wurden.

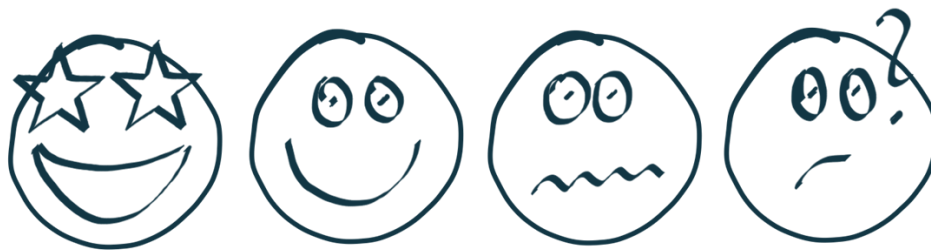


Abbildung 5.1
Zuversichtsemojis

Nach der Literaturrecherche und Planung dieser Studie stieß die Autorin auf einen zu dem Zeitpunkt kürzlich erschienenen Artikel (siehe Foster et al., 2022), in dem die Zuversicht mittels *confidence emojis* (siehe Abbildung 3.2) erhoben wird. Diese unterscheiden sich nicht nur in der Anzahl der Skalenpunkte und der grafischen Gestaltung, sondern auch in der Skalenorientierung. Trotz dieser Unterschiede wird dies als Bestätigung für die Wahl von Emojis als Skala für die Angabe der Zuversicht für diese Studie gesehen. Im Folgenden Kapitel werden nun das Studiendesign nähergebracht und die einzelnen Schritte beschrieben.

5.3 Studiendesign

Im Rahmen dieser Studie wurde die Angabe der Zuversicht mithilfe von Zuversichtsemojis für drei Wochen in den Unterricht implementiert. Einerseits wurde untersucht, wie Schüler*innen auf eine derartige Intervention mit Zuversichtsemojis und generell das Thema der Zuversicht reagieren. Andererseits war es das Ziel herauszufinden, wie gut kalibriert die Schüler*innen in Bezug auf das Thema Prozentrechnung sind und ob Schülerinnen anders ankreuzen als Schüler. Dieses Thema wurde gewählt, da es in diesem Zeitraum in der Jahresplanung vorgesehen war. Nach diesen drei Wochen wurden die Intervention und auch im speziellen die Zuversichtsemojis mittels Feedbackbogen evaluiert. Die dreiwöchige Intervention bestand aus mehreren Phasen. Zu Beginn gab es eine thematische Einführung und eine gemeinsame Erarbeitung der Zuversichtsemojis im Plenum. Für die anschließenden drei Wochen wurden Zuversichtsemojis sowohl bei Hausübungen und Schulübungen als auch bei einer Lernzielkontrolle verwendet, bevor diese im Anschluss evaluiert wurden. In Abbildung 5.2 wird das Studiendesign zur besseren Veranschaulichung grafisch dargestellt. Die folgenden Unterkapitel beschreiben nun die einzelnen Phasen genauer.

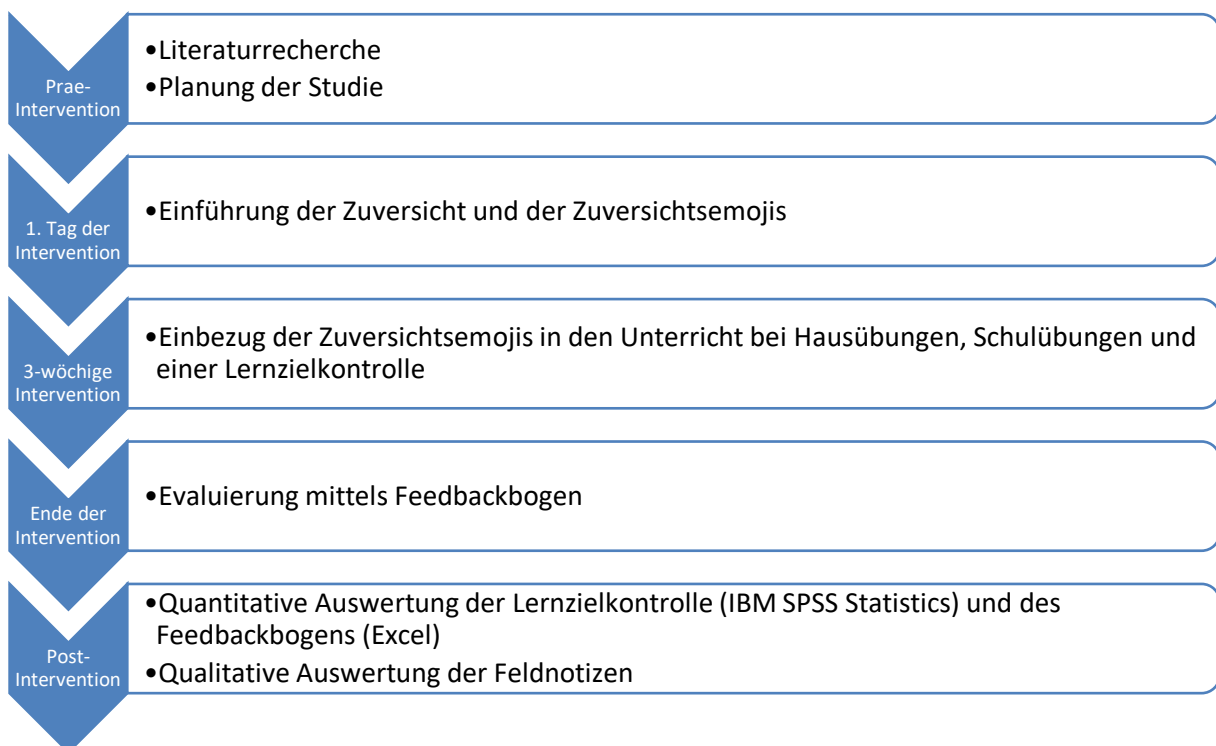


Abbildung 5.2
Studiendesign

5.3.1 Einführung der Zuversicht und der Zuversichtsemojis

Die dreiwöchige Intervention startete mit der Einführung in die Thematik der Zuversicht und der gemeinsamen Erarbeitung der Bedeutung beziehungsweise Beschreibung der Zuversichtsemojis. Das übergeordnete Ziel der ersten Unterrichtseinheit war es, den Schüler*innen das Konzept von Confidence-Based Learning näherzubringen und mit ihnen die Bedeutung der einzelnen Zuversichtsemojis zu erarbeiten. Den Schüler*innen wurde zu Beginn erklärt, dass in den nächsten drei Wochen ein Projekt durchgeführt wird. In dieser Zeit müssen sie Aufgaben nicht nur lösen, sondern auch angeben, wie zuversichtlich sie sind, dass sie die Aufgabe richtig gelöst haben. Um eine Beeinflussung der Ergebnisse zu vermeiden, erhielten die Schüler*innen keine weiteren Informationen zur Studie, wie etwa zu den Forschungsfragen.

Zu Beginn wurden die Schüler*innen gefragt, was sie unter dem Begriff Zuversicht verstehen, gefolgt von einer Erklärung der Lehrerin. Zur Angabe der Zuversicht wurden vier Zuversichtsemojis (siehe Abbildung 5.1) herangezogen. Diese wurden im Vorfeld einzeln auf A4-Blätter gedruckt und bei der Erarbeitung an die Tafel gehängt. Die Schüler*innen hatten nun ein paar Minuten Zeit, um selbstständig zu überlegen, was die einzelnen Zuversichtsemojis aussagen könnten, bevor es im Anschluss im Plenum diskutiert wurde. Ebenso sollten sie dazu Notizen in ihr Schulübungsheft niederschreiben. Die Schüler*innen bekamen nach dem Gespräch im Plenum jeweils alle vier Zuversichtsemojis, die sie untereinander in ihr Heft kleben sollten. Die Lehrerin einigte sich hier gemeinsam mit den Schüler*innen auf Sätze, die die Zuversichtsemojis am besten beschreiben (Kapitel 6.1.1). Diese wurden im Schulübungsheft und an der Tafel festgehalten. Es wurden im Anschluss ein paar Aufgaben unterschiedlicher Schwere an die Tafel geschrieben, beginnend mit $1+1=?$. Mit Hilfe dieser Aufgaben wurde die Wahl der Zuversichtsemojis erklärt, indem die Begründungen für die Auswahl besprochen wurden. Ziel dessen war es, Gedankengänge bei den Schüler*innen anzuregen. Immer wieder wurde betont, dass die Wahl für ein Zuversichtsemoji nicht willkürlich sein darf. Aufgrund dessen wurde ein großer Fokus auf die Begründung der Wahl eines Zuversichtsemojis gelegt. Der zweite Teil der Unterrichtseinheit wurde mit der Prozentrechnung verbracht und im Zuge dessen wurde die Verwendung der Zuversichtsemojis geübt. Da die Prozentrechnung bereits in der 6. Schulstufe thematisiert wurde, wurde zu Beginn wiederholt. Die Aufgabe, die in Abbildung 5.3 dargestellt wurde, diente dem erneuten Einstieg in die Thematik und ebenso der Erläuterung der Zuversichtsemojis. Die Schüler*innen lösten zu Beginn Aufgabe 1 in Einzelarbeit und wählten ein passendes Zuversichtsemoji, bevor diese im Plenum besprochen wurde. Ebenso wurde die Wahl der Zuversichtsemojis besprochen und einige Schüler*innen teilten mit, welches sie gewählt haben und begründeten ebenso ihre Wahl. Im Anschluss folgte der gleiche Ablauf

für Aufgabe 2. Die Lehrperson fragte nun, ob diesmal ein anderes Zuversichtsemoji gewählt wurde und ebenso nach einer Begründung.

Löse die Aufgabe zuerst und wähle dann ein Zuversichtsemoji!

1. Gib $\frac{11}{100}$ als Dezimalzahl und in Prozent an!



2. Gib $\frac{24}{100}$ als Dezimalzahl und in Prozent an!



Abbildung 5.3

Aufgabe zu Prozentrechnung mit Angabe der Zuversicht

5.3.2 Einbezug der Zuversichtsemojis in den Unterricht

In den drei Wochen der Intervention wurden die Zuversichtsemojis nicht nur in den Schulübungen implementiert, sondern auch bei den Hausübungen und in einer Lernzielkontrolle. Die Schüler*innen bekamen entweder Übungsblätter als Hausübungen, auf denen die Zuversichtsemojis bereits abgebildet waren (siehe Abbildung 5.4) oder sie bekamen diese für jede Aufgabe in Form eines Stickers. Auch für die Schulübungen erhielten die Schüler*innen die Zuversichtsemojis als Sticker. Im Unterricht, ebenso wie bei den Hausübungen, wurde weiterhin darauf geachtet, dass die Wahl der Zuversichtsemojis begründet wird. Dies wurde in den Schulübungen vermehrt mündlich besprochen.

Löse folgende Aufgaben und wähle jeweils eines der vier Zuversichtsemojis. Begründe, warum du es gewählt hast.

1. Schreibe als Prozentzahl, als Bruch und als Dezimalzahl!

Bsp.: $70\% = \frac{70}{100} = 0,7$

$77\% = \quad =$

$\frac{8}{100} = \quad =$

$1,15 = \quad =$



2. Wende die Begriffe Grundwert (G), Prozentsatz (p) und Prozentwert (W) richtig an. Setze die Zahlenwerte entsprechend ein.

70% von 760 Schüler*innen möchten gerne den Schultag erst um 9 Uhr beginnen. 532 Jugendliche würden gerne länger schlafen.

$G =$

$W =$

$p =$



Abbildung 5.4

Hausübung zur Prozentrechnung mit Angabe der Zuversicht und Begründung

5.3.3 Lernzielkontrolle

Um die zweite Forschungsfrage beantworten zu können, wurde eine Lernzielkontrolle (siehe Anhang 1) zum Thema Prozentrechnung erstellt und durchgeführt. Die Lernzielkontrolle bestand aus 11 Items, die verschiedene Aspekte der Prozentrechnung beinhalteten. Dabei mussten die Schüler*innen nicht nur die einzelnen Aufgaben beantworten, sondern bei jeder Aufgabe auch angeben, wie zuversichtlich sie sind, dass sie die Aufgabe richtig gelöst hatten. Dies erfolgte mittels Auswahl eines Zuversichtsemojis. Im Unterschied zu den Hausübungen und Schulübungen war hier nicht nach einer Begründung gefragt. Weiters bekamen die Schüler*innen für die Lernzielkontrolle mehr Zeit, um einerseits zu verhindern, dass ein Nachteil durch den Einbezug der Zuversichtsemojis entsteht und andererseits, damit die Wahl der Zuversichtsemojis auch gewissenhaft durchgeführt werden kann.

Im Folgenden möchte die Autorin hervorheben, warum die Lernzielkontrolle, die in diesem Fall auch dem Zweck einer Studie diene, nicht digital abgehalten wurde. Es ist klar, dass eine gewisse

Zeitersparnis mit der digitalen Abhaltung einer Studie einhergeht. In der bisherigen beruflichen Laufbahn als Lehrerin konnte die Autorin jedoch beobachten, dass es oftmals an der Praktikabilität scheitert, wenn Neues von Lehrer*innen nicht in den Unterricht implementiert wird. Da ein Ziel dieser Studie unter anderem ist, dass Confidence-Based Learning/Assessment mehr Aufmerksamkeit in Bildungseinrichtungen erhält und im Weiteren mehr zum Einsatz kommt, wurde hier versucht, dies so einfach umsetzbar wie möglich zu gestalten. Denkt man nun an den Unterricht, ist es nicht notwendig, die dadurch erhaltenen Daten statistisch auszuwerten. Auch ohne eine Auswertung erhalten Lehrer*innen Informationen über die Verortung des Wissens und wie gut sich Schüler*innen einschätzen können. Die dadurch enthaltenen Informationen können einerseits von Lehrer*innen genutzt werden, um beispielsweise Zuversichtsgespräche zu führen und andererseits für Lehrer*innen und Schüler*innen, um zu erkennen, wo noch Übungsbedarf besteht.

5.3.4 Evaluierung mittels Fragebogen

Nach der dreiwöchigen Intervention wurde diese mithilfe eines selbst erstellten Feedbackbogens (siehe Abbildung 5.5) evaluiert. Die Autorin nahm die Evaluationsstudie von Issroff und Gardner-Medwin (1998) wahr, bei der Medizinstudierende des University College London teilnahmen. Die dort verwendeten Items stufte die Autorin jedoch als inhaltlich zu komplex ein für Schüler*innen einer 7. Schulstufe. Weiters unterscheidet sich die zu beantwortende Forschungsfrage. Aufgrund dessen wurden die einzelnen Items von den sprachlichen Anforderungen und der Art der Fragestellungen an die Klasse und an die Forschungsfragen angepasst.

Liebe Schülerin! Lieber Schüler!
Bitte fülle diesen Fragebogen gewissenhaft aus!



- Fandest du das Arbeiten mit den Zuversichtsemojis...
 ...gut? Kreuze an! ☐ Ja ☐ Nein
 ...hilfreich? Kreuze an! ☐ Ja ☐ Nein
 ...nützlich? Kreuze an! ☐ Ja ☐ Nein
- Fandest du das Arbeiten mit den Zuversichtsemojis in folgenden Situationen gut/hilfreich?
 Kreuze an und begründe!

	Ja	Nein	Begründung
Hausübung	☐	☐	
Schulübung	☐	☐	
Lernzielkontrolle	☐	☐	

- Was hat dir am Arbeiten mit den Zuversichtsemojis gefallen? Kreuze Zutreffendes an oder/und schreibe selbst etwas dazu!
☐ Ich habe gelernt mich besser einzuschätzen.
☐ Ich habe erkannt, welche Themen ich noch üben sollte.
☐ Ich habe erkannt, welche Themen ich schon gut gefestigt habe.
☐
☐
- Was hat dir am Arbeiten mit den Zuversichtsemojis nicht gefallen?
- Wie ist es dir bei der Entscheidung für ein Zuversichtsemoji (Selbsteinschätzung) gegangen?
 Kreuze an!
☐ Es ist mir sehr leicht gefallen mich für ein Zuversichtsemoji zu entscheiden.
☐ Es ist mir eher leicht gefallen mich für ein Zuversichtsemoji zu entscheiden.
☐ Es ist mir eher schwer gefallen mich für ein Zuversichtsemoji zu entscheiden.
☐ Es ist mir sehr schwer gefallen mich für ein Zuversichtsemoji zu entscheiden.
- Hat dir das Begründen („weil:“) bei deiner Entscheidung für ein Emoji geholfen? Kreuze an! ☐ Ja ☐ Nein
- Würdest du wieder mit Zuversichtsemojis arbeiten wollen? Kreuze an und begründe! ☐ Ja ☐ Nein
 Begründung:
- Was würdest du am Arbeiten mit Zuversichtsemojis ändern?
- Hier ist Platz, um deine eigenen Zuversichtsemojis zu zeichnen.

Danke! 😊

Abbildung 5.5
Fragebogen zur Evaluierung der dreiwöchigen Intervention

5.4 Feldnotizen

Die qualitative Methode dieser Studie besteht aus Beobachtungen, die in Form von Feldnotizen verschriftlicht wurden. Feldnotizen können in einem sogenannten Feldtagebuch festgehalten werden. Dies erfolgt meist stichwortartig, da die Forschenden direkt im Forschungsfeld involviert sind und dadurch zumeist keine Zeit für das Niederschreiben vollständiger Sätze zur Verfügung steht. Es wird empfohlen, die stichwortartigen Feldnotizen direkt nach dem Verlassen des Forschungsfeldes auszuformulieren. (Bortz & Döring, 2006) Die Forschenden legen den Fokus ihrer Beobachtung passend zu ihrer Forschungsfrage. Es können dabei sowohl Eindrücke, Beobachtungen als auch Emotionen und Gedanken notiert werden. (PEARL Projekt-Team, o. D.) Die Feldnotizen enthalten im Kontext dieser Studie Notizen zu Ereignissen während der dreiwöchigen Intervention. Diese können demnach auch Interpretationen von bestimmten Situationen beinhalten, wie etwa der Vermutung über die Gründe von Reaktionen von Teilnehmer*innen. (Curd-Christiansen, 2020) Spradley (1980) empfiehlt auch Informationen über das Setting, die Aktivitäten, vorhandene Objekte, Handlungen und Aktivitäten, Ziele, Emotionen der Teilnehmenden und auch über die Zeit beziehungsweise Dauer zu notieren. Nach Curdt-Christiansen (2020) fehlt in dieser Aufzählung die Reflexion der Forschenden. Wichtig ist, dass dabei unterschieden wird zwischen objektiven Beobachtungen und subjektiven Gedanken der Forschenden und, dass diese nicht vermischt werden. (Bortz & Döring, 2006) Um diese Vermischung zu verhindern, haben Forschende unterschiedliche Methoden. Unter anderem werden Seiten in verschiedene Spalten unterteilt, von denen beispielsweise eine für Reflexionen der Forschenden genutzt wird. (Curd-Christiansen, 2020) Obwohl durch diese Methode viele Daten gesammelt werden können, bringt es natürlich auch Limitationen mit sich. Beispielsweise wählen die Forschenden „where to look, what to look at, when to look, how to look and whom to look at“ (Curd-Christiansen, 2020, S. 346). Demnach ist diese Art der Beobachtung sehr subjektiv und beeinflussbar von persönlichen Interessen und Erfahrungen. Ebenso kann die Anwesenheit der Forschenden die Teilnehmenden und vor allem ihr Verhalten in gewisser Weise beeinflussen. (Curd-Christiansen, 2020) Im Kontext dieser Studie muss der letzte Punkt jedoch modifiziert gesehen werden, da für die Teilnehmer*innen (Schüler*innen) durch die Anwesenheit der Forscherin (Lehrerin) in diesem Kontext kein ungewöhnliches Setting geschaffen wird. Trotz der vorhin genannten Limitierungen aufgrund der fehlenden Objektivität, entschied sich die Forscherin für den Einsatz von Feldnotizen in ihrer Studie.

5.5 Auswertung der Daten

Jene Daten, die durch die Hausübungen, Schulübungen sowie Feldnotizen gesammelt werden konnten, wurden qualitativ ausgewertet. Die quantitative Auswertung der Lernzielkontrolle und der Evaluierung am Ende der Intervention erfolgte mithilfe von Excel beziehungsweise IBM SPSS Statistics. Für die Auswertung der Evaluierung wurden die einzelnen Items passend codiert und händisch in Excel eingegeben. Die einzelnen offenen Fragen wurden im Anschluss separat betrachtet, wobei versucht wurde innerhalb der Daten Kategorien zu finden. Für die statistische Auswertung der Lernzielkontrolle wurde die Statistiksoftware IBM SPSS Statistics genutzt. Genauere Informationen über die Codierung der Daten erfolgt in Kapitel 6.

6 Ergebnisse

Dieses Kapitel befasst sich mit den Ergebnissen der Studie. Vorweg möchte die Forscherin anmerken, dass mündliche und schriftliche Zitate der Schüler*innen unverändert wiedergegeben werden, um ihre Stimmen nicht zu verfälschen. Auf Hinweise zu Fehlern in den Zitaten ([sic]) wird in diesem Kontext aufgrund der besseren Lesbarkeit verzichtet. Die einzelnen Aussagen der Schüler*innen wurden kodiert. Um zu erkennen von welchen Schüler*innen die Aussage stammt, wurde die Bezeichnung S1-S23 verwendet. Da die Aussagen in unterschiedlichen Kontexten vorgekommen sind, wurde dies ebenfalls entsprechend angemerkt. Aussagen in Hausübungen wurden mit *H* gekennzeichnet, von Schulübungen mit *Sch* und bei der Lernzielkontrolle mit *L*. Die Zahl im Anschluss bezeichnet die Anzahl der vorkommenden Aussage dieser Person. Anzumerken ist, dass die Evaluierung am Ende der drei-wöchigen Intervention anonym abgelaufen ist und somit die einzelnen Evaluierungen nummeriert (E1-E17) wurden. Diese Nummerierung stimmt demnach nicht mit der Nummerierung der Schüler*innen bei der Lernzielkontrolle beziehungsweise von Schulübungen und Hausübungen überein. Ebenso wurde angegeben, auf welches Item (I1-I9) die Antwort sich bezieht.

6.1 Feldnotizen und Ergebnisse der Schüler*innen

Im folgenden Kapitel werden die Feldnotizen festgehalten, die während des Projekts aufgezeichnet wurden. Ebenso werden einige der Ergebnisse der Schüler*innen beispielhaft präsentiert.

6.1.1 Einführung

Die erste Unterrichtseinheit des Projekts fand am 18. Mai 2022 statt. Die Schüler*innen wurden erst an diesem Tag in die Thematik eingeführt. Zuerst wurde den Schüler*innen erklärt, dass ein Projekt gestartet wird, bei dem ihre Mitarbeit sehr wichtig ist. Um die Schüler*innen auf das Thema vorzubereiten, begann die Forscherin über die Wörter „Zuversicht“ beziehungsweise „zuversichtlich sein“ zu sprechen, da sie sich nicht sicher war, ob sie diese Wörter kennen. Die Forscherin schrieb die Wörter an die Tafel und fragte in die Klasse, ob jemand erklären kann, was diese beiden Wörter bedeuten. Da diese Klasse für gewöhnlich sehr kommunikativ war und hier keine Antwort kam, ging sie davon aus, dass die Schüler*innen diese Wörter nicht kannten. Daher wurde versucht, diese Wörter in einen Kontext zu setzen. Die Forscherin sagte: „Ich bin sehr zuversichtlich, dass wir in der kommenden Unterrichtsstunde Bewegungs- und Sportunterricht haben.“ Daraufhin

meinte eine Schülerin: „Das heißt sicher sein, oder?“ (S12Sch1) Ein anderer meinte: „Ja, das ist, wenn man halt sagen mag, dass man sicher ist und so.“ (S2Sch1) Dabei fiel auf, dass die Schüler*innen mehr Bezug zu „sicher sein“ haben als zu „zuversichtlich sein“. Es benötigte ein wenig Zeit bis alle Schüler*innen die Bedeutung von „Zuversicht“ und „zuversichtlich sein“ verstanden oder zumindest behauptet hatten, sie würden es verstehen. Im Anschluss wurde den Schüler*innen erklärt, dass es wichtig ist, in den kommenden Wochen Aufgaben nicht nur zu lösen, sondern auch anzugeben, wie zuversichtlich sie sind, dass sie die Aufgabe richtig gelöst haben. Um zu zeigen, wie sie das angeben sollen, wurden die Zuversichtsemojis an die Tafel gehängt und eingeführt. Die Schüler*innen hatten Zeit sich zu überlegen, was die einzelnen Emojis aussagen könnten, und sollten sich dazu auch Notizen in ihr Schulübungsheft machen. Bei der gemeinsamen Besprechung herrschte rege Mitarbeit. Da in dieser Unterrichtseinheit auch eine zweite Kollegin anwesend war, konnte auch diese während der Besprechung im Plenum Notizen aufschreiben. Da in dieser Klasse in Mathematik für gewöhnlich immer im Team-Teaching unterrichtet wird und die Schüler*innen diese Kollegin kannten, stellte die Anwesenheit dieser Kollegin keine Beeinflussung der Schüler*innen dar. Im Folgenden werden die Feldnotizen wiedergegeben, die die Kollegin und die Forscherin während des Unterrichts notiert haben. Es handelt sich dabei um die Ideen der Schüler*innen zur Bedeutung der einzelnen Zuversichtsemojis. Die Aussagen der Schüler*innen wurden, wie bereits am Anfang dieses Kapitels erwähnt, direkt wiedergegeben. Aufgrund mangelnder Zeit während des Unterrichts, notierte die Forscherin nicht, welche*welcher Schüler*in die folgenden Aussagen tätigte. Die Schüler*innen äußerten sich zum ersten Zuversichtsemoji (Abbildung 6.1) wie folgt:

- „Dass man zuversichtlich ist.“
- „Halt, dass man ganz zuversichtlich ist.“
- „Ich glaube, das heißt, dass man sich sicher ist, dass die Aufgabe richtig ist.“
- „Wenn ich das ankreuze, bin ich mir richtig sicher.“
- „Das zeigt, dass ich mir richtig extrem sicher bin, dass es richtig ist.“



Abbildung 6.1

Erstes Zuversichtsemoji

Im Folgenden werden die Aussagen der Schüler*innen zum zweiten Zuversichtsemoji (siehe Abbildung 6.2) wiedergegeben:

- „Ich bin zuversichtlich.“
- „Ich bin mir nicht ganz sicher, aber schon eher sicher.“
- „Ich bin verwirrt.“
- „Ich bin unsicher, ob die Aufgabe richtig ist.“
- „Ich bin nicht sehr sicher, aber schon sicher.“
- „Dass es zuversichtlich ist.“
- „Ich bin mittelmäßig sicher.“



Abbildung 6.2

Zweites Zuversichtsemoji

Hier werden nun Ergebnisse für das dritte Zuversichtsemoji, das in Abbildung 6.3 dargestellt ist angeführt:

- „Wenig zuversichtlich, dass es richtig ist“
- „Ich weiß es nicht so wirklich.“
- „Ich bin verwirrt“
- „Ich bin unsicher.“

- „Ich bin irgendwie nervös.“
- „beides so, bisschen sicher, aber mehr unsicher“
- „Ich bin nicht so stark sicher.“
- „Das heißt normal sicher.“
- „Das Emoji zeigt, dass man sich nicht so viel sicher ist.“



Abbildung 6.3

Drittes Zuversichtsemoji

Zuletzt folgen die Ergebnisse bezüglich des vierten Zuversichtsemojis, das in Abbildung 6.4 dargestellt wird:

- „Ich bin verwirrt.“
- „Ich bin mir gar nicht sicher“
- „extrem unsicher“
- „Das 4.Smiley ist nicht so zuversichtlich.“
- „Es ist gar nicht zuversichtlich“
- „Das nimmt man, wenn man so gar nicht sicher ist.“



Abbildung 6.4

Viertes Zuversichtsemoji

Im Endeffekt wurden folgende Sätze mit den Schüler*innen (siehe auch Abbildung 6.5) erarbeitet, um die jeweiligen Zuversichtsemojis zu beschreiben:

- Ich bin mir sehr sicher, dass ich die Aufgabe richtig gelöst habe.
- Ich bin mir sicher, dass ich die Aufgabe richtig gelöst habe.
- Ich bin mir unsicher, ob ich die Aufgabe richtig gelöst habe.
- Ich bin extrem unsicher, ob ich die Aufgabe richtig gelöst habe.

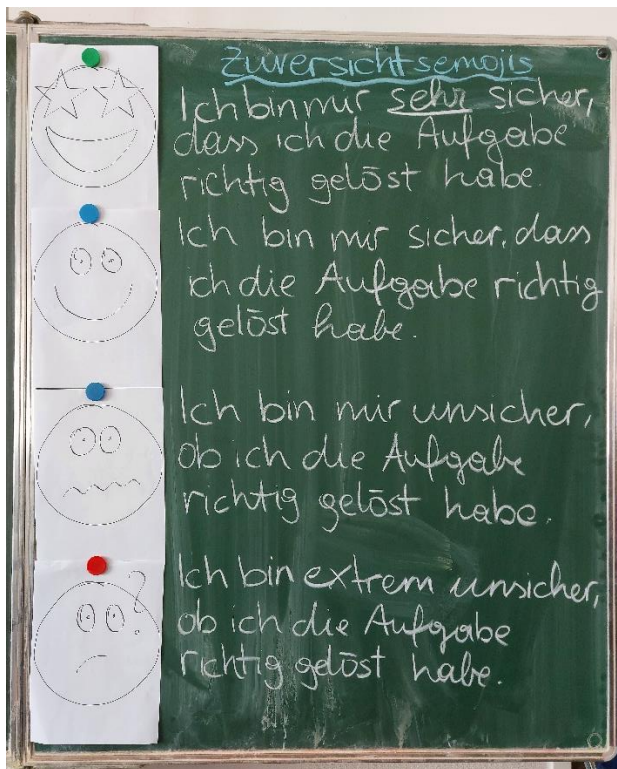


Abbildung 6.5

Beschreibung der Zuversichtsemojis an der Tafel

Nach der Erarbeitung der Zuversichtsemojis wurde geübt, wie man zur Wahl eines Zuversichtsemojis kommt. Bevor die Schüler*innen selbstständig Aufgaben bekamen, wurden einige gemeinsam gelöst. Wichtig war dabei, dass die Schüler*innen ehrlich sind und sich wirklich überlegen, welches Emoji sie ankreuzen. Ebenso wurde betont, dass die Wahl des Zuversichtsemojis ganz individuell ist und nicht in Partner*innenarbeit geschehen darf. Die erste Aufgabe, anhand derer die Wahl eines Emojis demonstriert wurde, war „ $1 + 1 = \underline{\quad}$ “. Bei der Frage, wer die Antwort kennt, zeigten gleich alle auf. Es wurde gefragt, welches Zuversichtsemoji sie wählen würden und warum. Eine Schülerin begründete: „Das erste mit den Sternaugen, weil das voll leicht ist und das jeder weiß.“ (S16Sch1) Als Übergang zum mathematischen Stoffgebiet, gab es ein kurzes Arbeitsblatt, das in Abbildung 5.3 zu sehen ist.

Die Schüler*innen mussten dabei zuerst die 1. Aufgabe (siehe Abbildung 5.3) lösen und ein Zuversichtsemoji auswählen und kurz begründen, warum sie ebendieses gewählt haben. Zur Überraschung der Forscherin konnten viele diese Aufgabe nicht richtig lösen, obwohl es sich dabei um eine Wiederholung aus der 6. Schulstufe handelt. Als Beispiel dienend werden die Ergebnisse von Schüler*innen angeführt. Schüler T. (S10) hat die Aufgabe falsch gelöst, kreuzte das vierte Zuversichtsemoji an und begründete es wie folgt: „Ich habe da letztes Jahr nicht aufgepasst. Ich habe keine Ahnung.“ (S10Sch1) Schülerin I. (S13) wählte das gleiche Zuversichtsemoji und begründete: „Weil ich’s komplett vergessen habe.“ (S13Sch1) „Weil ich mich nicht mehr so gut daran erinnern kann und ich mir nicht so sicher bin“ (S5Sch1), meinte Schüler M. (S5), der die Aufgabe richtig löste und sich für das dritte Zuversichtsemoji entschied. Das Beispiel wurde im Anschluss an der Tafel erklärt, dem folgte das selbstständige Bearbeiten der analogen 2. Aufgabe. Schüler A. (S2) löste die Aufgabe richtig und wählte das zweite Zuversichtsemoji und begründete seine Wahl: „Weil das vorherige Beispiel ähnlich war“ (S2Sch2) Ebenso konnte Schülerin M. (S3) die Aufgabe richtig lösen, entschied sich allerdings für das erste Zuversichtsemoji und sagte: „Jetzt weiß ich ja wie’s geht“ (S3Sch1) Die Forscherin fragte, wie viele von ihnen jetzt etwas anderes angekreuzt hatten als zuvor. 13 von 18 der anwesenden Schüler*innen hoben daraufhin ihre Hand. Die Schüler*innen teilten nun mit, warum sie welche Zuversichtsemojis gewählt haben und was sich geändert hatte. Schülerin M. (S19) erklärte: „Bei der ersten Aufgabe wusste ich gar nicht mehr was das überhaupt ist. Es war auch falsch und ich hab das dritte Emoji genommen, weil ich keine Ahnung hatte. Ja und beim zweiten habe ich’s dann richtig gehabt und habe das zweite Smiley genommen, weil ich da schon sicherer war, aber halt nicht ganz sicher.“ (S19Sch1)

6.1.2 Einbezug der Zuversichtsemojis in den Unterricht

In diesem Unterkapitel werden nun Beobachtungen und Ergebnisse aus der dreiwöchigen Intervention dargestellt. Dies beinhaltet Feldnotizen und Daten, die den Hausübungen der Schüler*innen entnommen wurden, dabei vor allem die Begründungen für die Wahl der Zuversichtsemojis. Die Schüler*innen begründeten ihre Wahl an Zuversichtsemojis grundsätzlich sehr unterschiedlich. Auffällig war, dass Schülerin M. (S19) immer das erste Zuversichtsemoji wählte und immer die gleichen Begründungen dazuschrieb („Weil ich es kann“ (S19H1) bzw. „Weil ich es gelernt habe.“ (S19H2)). Abbildung 6.6 zeigt ein Hausübungsbeispiel von Schülerin M. (S19). Ähnliches konnte bei Schüler M. (S6) beobachtet werden, der auch schrieb, dass er es könne, beziehungsweise, dass es sehr einfach sei. Bei einer Aufgabe notierte er: „Ich habe das erste gewählt, weil ich das heute gelernt und verstanden habe.“ (S6H1) Im Folgenden werden nun weitere Begründungen für die Wahl des ersten Zuversichtsemojis angeführt:

- „Weil wir es in der Schule gemacht haben“ (S7H1)
- „Weil es von der Lehrerin gut erklärt wurde“ (S7H2)
- „Mir die Schulübung geholfen hat“ (S9H1)
- „Weil wir es in der Schule gemacht haben“ (S17H1)
- „Weil mir die erste aufgabe [ein vorgerechnetes Beispiel als Hilfestellung] erklärt hat wie man es rechnet“ (S11H1)
- „Weil ich in Stunde aufgepasst“ (S4H1)
- „Weil mir die i Box [Infobox im Schulbuch] geholfen hat“ (S7H3)
- „Weil ich es kann“ (S5H1, S19H3)
- „Weil ich das in 2 Klasse gelernt“ (S20H1)
- „Weil es leicht war“ (S5H2, S6H2)
- „Weil ich das schon kann“ (S18H1)
- „Weil ich heute gelernt und verstanden hab“ (S16H1)
- „Weil ich den Formel auswendig kann und es nur eingesetzt hab“ (S12H1)

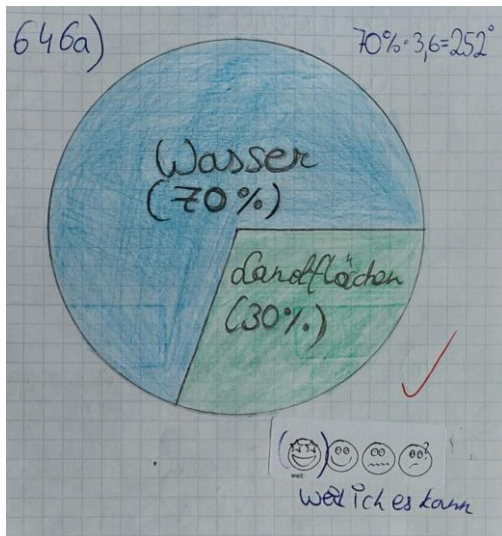


Abbildung 6.6

Hausübung von Schülerin M. (S19)

Schülerin W. (S12) wählte bei mehreren richtig gelösten Aufgaben das erste Zuversichtsemoji und schrieb unter anderem folgende Begründungen dazu: „Weil ich den Formel genommen habe wo ich mir sicher bin dass das der richtige Formel ist.“ (S12H2) und „Weil ich schon seit lange kan.“ (S12H3). Vor allem eine Hausübung der Schülerin Ma. (S3) zeigte, dass sie in vielen Bereichen der Prozentrechnung noch unsicher ist. Sie löste zwar alle Aufgaben richtig, wählte aber stets das dritte

oder vierte Zuversichtsemoji mit der Begründung, dass sie es „irgendwie vergessen“ (S3H1) hat. Ähnliches konnte bei Schüler A. (S2) festgestellt werden. Die Aufgabe ist richtig gelöst und die Wahl des dritten Zuversichtsemojis wurde mit den Worten „Weil ich mich nicht mehr bei Prozentrechnungen gut auskenne“ (S2H1) begründet. Andere Schüler*innen hingegen führten bei der Wahl des ersten Zuversichtsemojis aus, dass sie sicher sind und bei der Wahl des zweiten Zuversichtsemojis, dass sie unsicher sind. Schüler D. (S18) löste bei einer Hausübung eine Aufgabe falsch und wählte das zweite Zuversichtsemoji. Er schrieb er habe es „ein bisschen verstanden“ (S18H2). Dieselbe Aufgabe löste Schüler O. (S15) ebenso falsch und wählte das dritte Zuversichtsemoji begründet damit, dass er sich dabei schwergetan habe. Des Öfteren schrieben Schüler*innen, die das dritte oder vierte Zuversichtsemojis ankreuzten, dass sie verwirrt waren. Weitere Argumente waren „Ich habe es nicht verstanden“ (S11H2), „Ich bin mir nicht sicher“ (S21H1) und „Ich weiß nicht“ (S14H1). Auffallend war Schüler A. (S2), der zwei Aufgaben gar nicht löste und das vierte Zuversichtsemoji wählte und dazuschrieb, dass er es nicht versteht. Ebenso teilten manche Schüler*innen mit, dass sie nicht wissen, ob die Aufgabe richtig gelöst wäre und dass sie deshalb das vierte Zuversichtsemojis ankreuzten.

Die Forscherin bemerkte während der Intervention immer wieder, dass einige Schüler*innen die Richtigkeit ihrer Aufgaben gut einschätzen konnten. Dies wurde unabhängig davon beobachtet, ob die Aufgabe richtig oder falsch gelöst wurde. Jedoch konnten auch entgegengesetzte Beobachtungen gemacht werden. Schüler A. (S2) löste beispielsweise eine Aufgabe falsch, kreuzte das erste Zuversichtsemoji an und schrieb, dass er das schon in der 2.Klasse gelernt habe (siehe Abbildung 6.7). Zumindest bei dieser Aufgabe kam es zu einer Fehleinschätzung. Nur wenige Male konnte anhand der Hausübungen eine Unterschätzung erkannt werden, viel öfter kam es zu einer Überschätzung. Abbildung 6.8 zeigt eine Hausübung von Schülerin I. (S13), die sich bei der ersten Aufgabe unterschätzte und bei der dritten und vierten definitiv überschätzte.

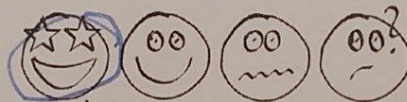
2. Wende die Begriffe Grundwert (G), Prozentsatz (p) und Prozentwert (W) richtig an. Setze die Zahlenwerte entsprechend ein.

70% von 760 Schüler*innen möchten gerne den Schultag erst um 9 Uhr beginnen. 532 Jugendliche würden gerne länger schlafen.

$$G = \cancel{1272}$$

$$W = \cancel{12,92}$$

$$p = \cancel{100\%}$$



weil: ich habe das in 2 Klasse gelernt

Abbildung 6.7

Hausübung von Schüler A. (S2)

635a) $G = 1200 \text{ kg}$ $p = 25\%$
 $W = G \cdot \frac{p}{100}$
 $W = 1200 \cdot \frac{25}{100}$
 $W = 300 \text{ kg}$

636a) 28% $G = 250$ $p = 28\%$
 $W = G \cdot \frac{p}{100}$
 $W = 250 \cdot \frac{28}{100}$
 $W = 70 \text{ kg}$

642a) $p = 30\%$ $W = 88 \text{ kg}$
 $G = \frac{W \cdot 100}{p}$
 $G = \frac{88 \cdot 100}{30}$
 $G = 8800 \div 30$
 $G = 293 \frac{1}{3}$

643a) 10% , 20%
 $G = \frac{W \cdot 100}{p}$
 $G = \frac{20 \cdot 100}{10}$
 $G = \frac{2000}{10}$
 $G = 200$

weil: Ich bin mir sehr unsicher weil es 2 verschiedene Lösungen gibt!

weil: Ich bin mir sicher das iches richtig hab weil es leicht ist

weil: Ich bin mir halb sicher das es richtig ist weil ich es ein bisschen vergessen habe.

weil: Ich bin mir ganz ganz sicher das es richtig ist!

Abbildung 6.8

Hausübung von Schülerin I. (S13)

Nicht unerwähnt bleiben soll an dieser Stelle eine Frage von Schüler Am. (S8). Er fragte: „Warum soll ich das [Wählen eines Zuversichtsemojis] machen? Das hat ja nichts mit Mathematik zu tun.“ (S8Sch1) Diese Frage kam vom gleichen Schüler in den drei Wochen öfter auf und wird in Kapitel

7 diskutiert. Weiters teilten Schüler*innen mit, dass sie das Einkleben der Zuversichtsemojis (in Form von Stickern) anstrengend fanden und ihnen das keinen Spaß machte.

Nach den Beobachtungen der dreiwöchigen Intervention und den Ergebnissen der Hausübungen der Schüler*inne, folgt nun die Auswertung der Daten der Lernzielkontrolle mittels IBM SPSS Statistics.

6.2 Ergebnisse der Lernzielkontrolle

6.2.1 Richtigkeit und Zuversicht

Für die statistische Auswertung der Lernzielkontrolle wurden die einzelnen Punkte der Items in IBM SPSS Statistics umcodiert. Wurde das Beispiel komplett richtig gelöst, was der vollen Punktezahl entspricht, wurde das Beispiel als richtig gewertet. In allen anderen Fällen wurde das Beispiel als falsch gelöst bewertet. Ebenso wurden die einzelnen Zuversichtsemojis umcodiert. Jedem Emoji wurde ein absteigender Wert beginnend mit 4 für das erste Zuversichtsemoji und 1 für das vierte zugeordnet. Wie in Tabelle 6.1 ersichtlich, stellten die Aufgaben 1 und 11 die größte Herausforderung dar, diese konnten nur von 10% der Schüler*innen korrekt gelöst werden. Aufgabe 6 hingegen konnten 16 von 20 Schüler*innen richtig lösen.

Item	Mittelwert	Standardabweichung
1	,10	,31
2	,65	,49
3	,50	,51
4	,15	,37
5	,35	,37
6	,80	,41
7	,30	,47
8	,20	,41
9	,45	,51
10	,25	,44
11	,10	,31

Tabelle 6.1

Mittelwert und Standardabweichung der richtigen Antworten für jedes Item

Der Zusammenhang zwischen der Anzahl an korrekten Items und der durchschnittlichen Zuversicht ist stark positiv ($r=,581$, $p<,01$). Aufgrund der geringen Aussagekraft dieses Zusammenhangs wurde eine neue Variable definiert, die angab, ob sich die Schüler*innen bei einem Item gut oder schlecht eingeschätzt hatten. In Tabelle 6.2 wird gezeigt, wie die neue Variable zustande kam. Zuerst wurde unterschieden, ob ein Item richtig oder falsch gelöst wurde und im Anschluss abhängig von der Angabe der Zuversicht bestimmt, ob der*die Schüler*in die Richtigkeit der Aufgabe gut oder schlecht eingeschätzt hatte. Zur besseren Auswertbarkeit wurden hierbei jeweils zwei Stufen der Zuversicht (3, 4 und 2, 1) zusammengefasst wurden.

Item	Angabe der Zuversicht	Einschätzung
richtig	4	Gut eingeschätzt
	3	
	2	Schlecht eingeschätzt
	1	
falsch	4	Schlecht eingeschätzt
	3	
	2	Gut eingeschätzt
	1	

Tabelle 6.2

Bewertung der Einschätzung abhängig von der Korrektheit des Items und der angegebenen Zuversicht

Folgende Tabelle 6.3 zeigt die Anzahl richtiger Antworten und die Anzahl gut eingeschätzter Items pro Person. Es lässt sich erkennen, dass zwei Schüler*innen nur eine Aufgabe richtig gelöst, sich aber bei 7 beziehungsweise 9 Aufgaben gut eingeschätzt hatten. Ein*eine Schüler*in konnte alle Aufgaben richtig lösen und schätzte sich auch bei allen Aufgaben gut ein. Die meisten Schüler*innen trafen in Bezug auf die Richtigkeit ihrer Antwort bei mehr als der Hälfte der Aufgaben eine gute Einschätzung. Im Schnitt lösten die Schüler*innen 3,85 von 11 Aufgaben richtig und schätzten sich bei 7,15 Items gut ein (siehe Tabelle 6.4).

Schüler*in	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Anzahl richtiger Antworten	1	6	3	1	3	2	3	8	4	5	3	6	5	5	2	2	2	2	11	3
Anzahl gut eingeschätzter Items	7	8	7	9	7	4	7	10	9	6	5	7	6	6	7	4	9	7	11	7

Tabelle 6.3

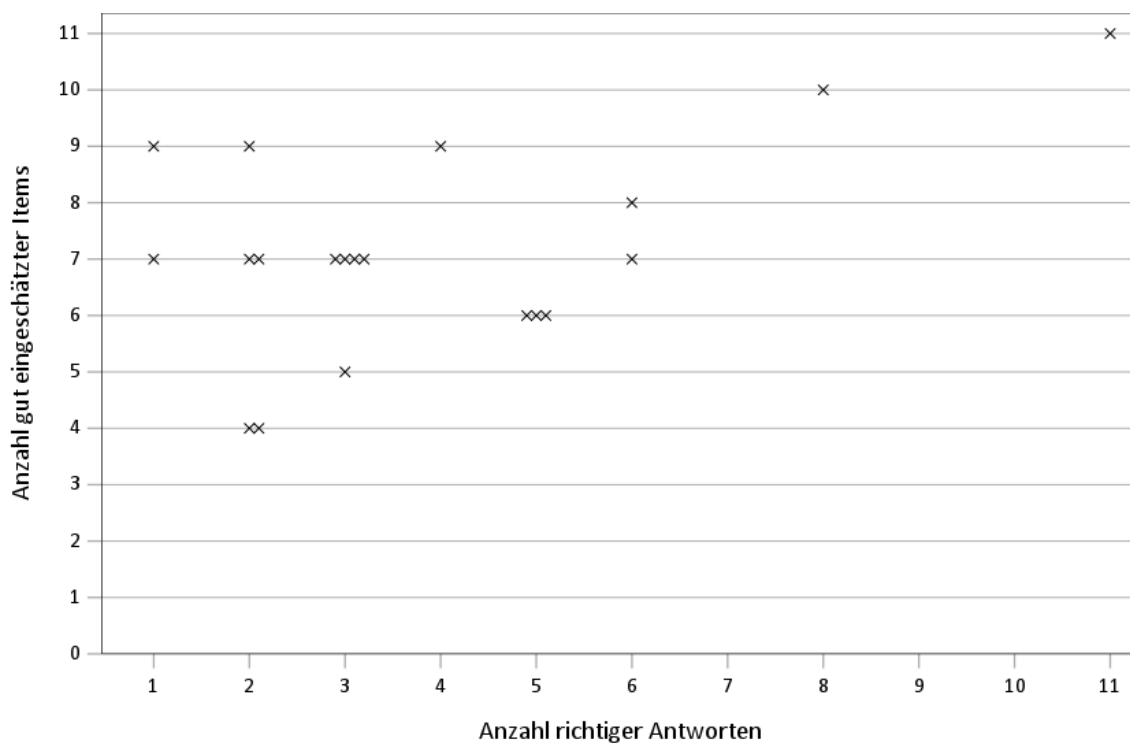
Anzahl richtiger Antworten und Anzahl an gut eingeschätzten Items pro Schüler*in

	Spannweite	Minimum	Maximum	Mittelwert	Std.- Abweichung
Mittlere Zuversicht pro Schüler*in	1,91	1,91	3,82	2,72	,53
Anzahl richtiger Antworten	10	1	11	3,85	2,50
Anzahl gut eingeschätzter Items pro Schüler*in	7	4	11	7,15	1,82

Tabelle 6.4

Deskriptive Statistik: mittlere Zuversicht pro Person, Anzahl richtiger Antworten und Anzahl gut eingeschätzter Items pro Schüler*in

Die Korrelation nach Pearson zeigt nach Cohen (1988) einen signifikanten hohen positiven Zusammenhang zwischen der Anzahl an richtigen Antworten und der Anzahl an gut eingeschätzten Items ($r = ,505$; $p < ,05$; $n=20$). Das Punktdiagramm (Abbildung 6.9) zeigt den positiven Zusammenhang der beiden Variablen. Auf die ganze Lernzielkontrolle bezogen kam es zu 77 Fehleinschätzungen. Dies entspricht 35% aller Einschätzungen. 69-mal wurde überschätzt und 8-mal unterschätzt. Mittels der Korrelation nach Pearson wurde überprüft, ob ein Zusammenhang zwischen der Anzahl an richtigen Antworten und der Überschätzung besteht. Zwischen der Anzahl richtiger Antworten und der Überschätzung konnte nach Cohen (1988) ein mittlerer negativer Zusammenhang nachgewiesen werden ($r = -.479$; $p < ,05$). Es konnte jedoch kein signifikanter Zusammenhang zwischen der Anzahl an richtigen Antworten und der Unterschätzung nachgewiesen werden.

**Abbildung 6.9**

*Zusammenhang zwischen der Anzahl richtiger Antworten und der Anzahl gut eingeschätzter Items pro Schüler*in*

Anmerkung. Da mehrfach die x- bzw. y-Komponente ident waren und diese im Punktdiagramm direkt übereinander liegen und nicht erkennbar wären, wurden die x-Komponenten zur besseren Veranschaulichung für die Darstellung angepasst.

6.2.2 Geschlechterunterschied

Mit einem Mittelwert von 3,91 ($SD=1,92$; $n=11$) beantworteten Schüler durchschnittlich mehr Aufgaben richtig als Schülerinnen ($M=3,78$; $SD=3,19$; $n=9$). Die durchschnittliche Anzahl an gut eingeschätzten Items bei Schülerinnen ($M=7,22$; $SD=2,17$) war höher als die der Schüler ($M=7,09$; $SD=1,58$). In Summe kam es zu 43 Fehleinschätzungen bei Schülern und zu 34 bei Schülerinnen (siehe Tabelle 6.5). Bemerkbar ist, dass sich Schüler öfter überschätzten als Schülerinnen. Weiters kam es zu 8 Unterschätzungen bei Schülerinnen und zu zwei bei Schülern. Sowohl bei der Überschätzung ($U=46,500$; $Z=-,232$; $p=,824$) als auch bei der Unterschätzung ($U=41,500$; $Z=-,758$; $p=,552$) konnte mittels Mann-Whitney-U-Test kein statistisch signifikanter Geschlechterunterschied nachgewiesen werden. Dabei wurde aufgrund der geringen Stichprobengröße die exakte Signifikanz betrachtet. Um

zu überprüfen, ob der Unterschied zwischen der Einschätzung von Schülerinnen und Schülern generell signifikant ist, wurde ein t-test herangezogen. Da die Stichprobengröße kleiner als 30 ist, muss die abhängige Variable als Voraussetzung normalverteilt sein. Dies wurde mittels Shapiro-Wilk-Test überprüft. Gemäß dem Shapiro-Wilk-Test war die Anzahl der Items, bei denen sich die Schüler*innen gut eingeschätzt haben, für die beiden Gruppen der unabhängigen Variablen (weiblich und männlich) normalverteilt ($p > ,05$). Da diese Voraussetzung erfüllt ist, wurde der t-test zur Berechnung des Unterschieds herangezogen. Nach dem Levene-Test der Varianzgleichheit wird angenommen, dass die Varianzen beider Gruppen gleich sind. Zwischen den beiden Gruppen konnte kein statistisch signifikanter Unterschied nachgewiesen werden ($t(18) = 0,157$; $p = ,877$).

	Überschätzung	Unterschätzung	gesamt
weiblich	32	8	34
männlich	37	2	43
gesamt	69	8	77

Tabelle 6.5

Anzahl der Überschätzungen und Unterschätzungen bei einer Gesamtanzahl von 220 Einschätzungen im Rahmen der Lernzielkontrolle

6.3 Ergebnisse der Evaluierung

Dieser Feedbackbogen behandelt die Erfahrungen der Schüler*innen, die sie in den drei Wochen während der Intervention gemacht haben. Ebenso wurde erfragt, in welchen Kontexten die Arbeit mit Zuversichtsemojis als gut beziehungsweise hilfreich empfunden wurde und wie es ihnen bei der Selbsteinschätzung ergangen ist. Am Tag der Erhebung waren nur 17 von 23 Schüler*innen anwesend. Aufgrund von Straightlining, dem konsequenten Auswählen der gleichen Antwortmöglichkeit im gesamten Fragebogen, wurde ein Fragebogen (E17) aus der Auswertung ausgeschlossen. Demnach werden Daten von 16 Schüler*innen analysiert. Beim ersten Item mussten die Schüler*innen angeben, wie sie das Arbeiten mit Zuversichtsemojis empfanden. Das gruppierte Säulendiagramm (siehe Abbildung 6.10) zeigt die Ergebnisse des ersten Items. 9 von 16 Schüler*innen gaben an, dass sie das Arbeiten als gut empfanden, dies entspricht in etwa 56% ($n=16$). Auffallend ist, dass 13 von 15 (eine Person hat hier nichts angekreuzt) Schüler*innen angaben, dass sie das Arbeiten mit Zuversichtsemojis als nicht nützlich einstufen. 60% ($n=15$) sagten aus, dass es auch nicht hilfreich war.

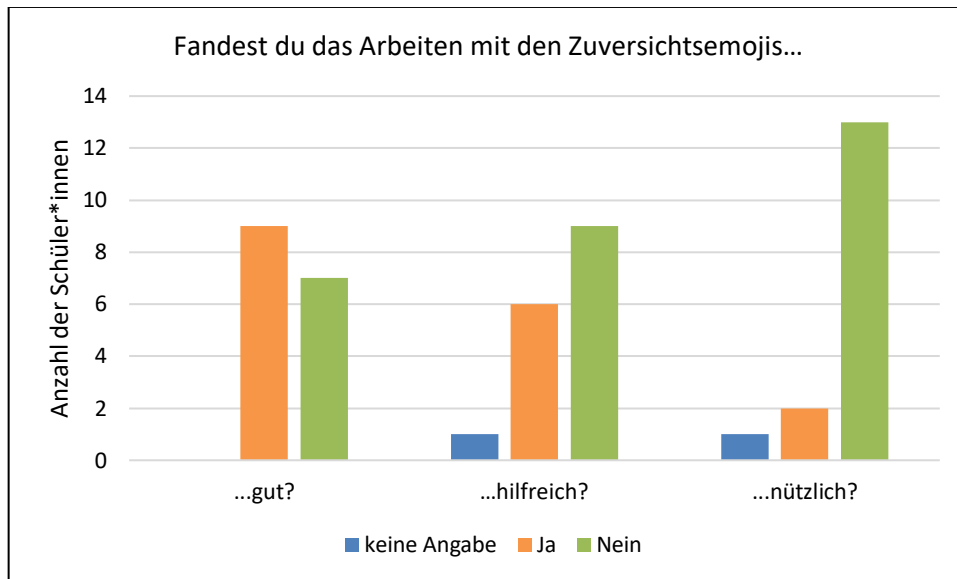


Abbildung 6.10

Meinung über das Arbeiten mit Zuversichtsemojis im Allgemeinen

Item 2 zielte darauf ab, herauszufinden in welchen Situationen die Schüler*innen das Arbeiten mit Zuversichtsemojis als gut beziehungsweise hilfreich beurteilten. Abbildungen 6.11, 6.12 und 6.13 illustrieren die Ergebnisse.

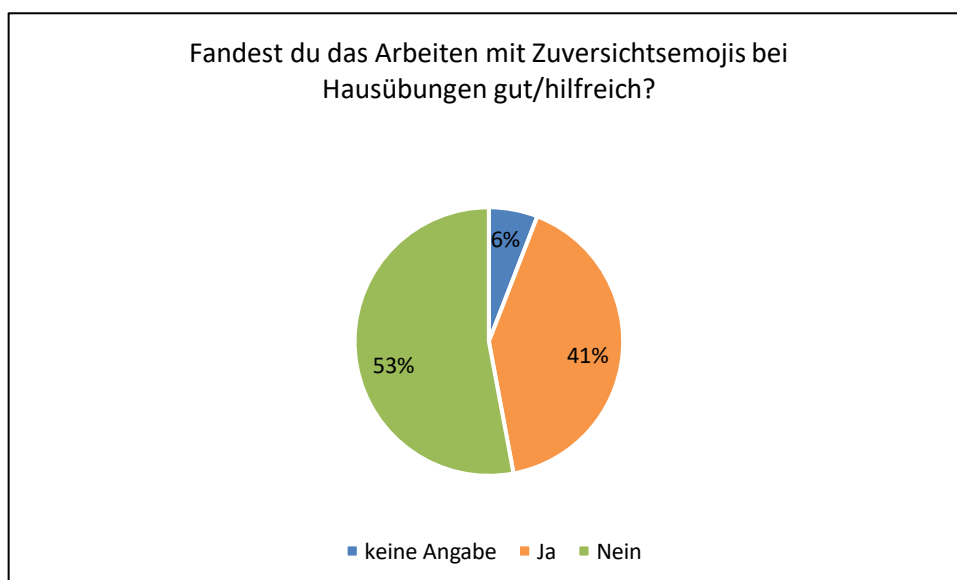


Abbildung 6.11

Einsatz von Zuversichtsemojis bei Hausübungen

Abbildung 6.12 zeigt, dass 53% angaben, dass sie den Einsatz der Zuversichtsemojis bei Hausübungen nicht gut oder hilfreich fanden. Sieben Schüler*innen merkten an, dass sie es nicht nützlich fanden, beziehungsweise, dass man es nicht brauche (z.B. „Hat nichts gebracht“ (E8I2.1), „Braucht man nicht“ (E11I2.1)). Ein*eine Schüler*in gab an, dass sie es nicht gut finde, weil „man manchmal drauf vergessen kann“ (E5I2.1). 41% der Klasse gaben an, dass sie es gut und hilfreich fanden und begründeten es damit, dass es hilfreich und leichter wäre und man sehen könne, ob der Stoff gut sitze (z.B. „So kann ich sehen ob ich es gut kann“ (E3I2.1), „Weil, wenn man Hü macht bringt es sehr man versteht alles“ (E10I2.1)). Weiters begründete eine Person, dass es sie an die Schulübung erinnere („Es hat mich an den Schulübungen erinnert. Also was wir in schule gemacht haben“ (E6I2.1)).

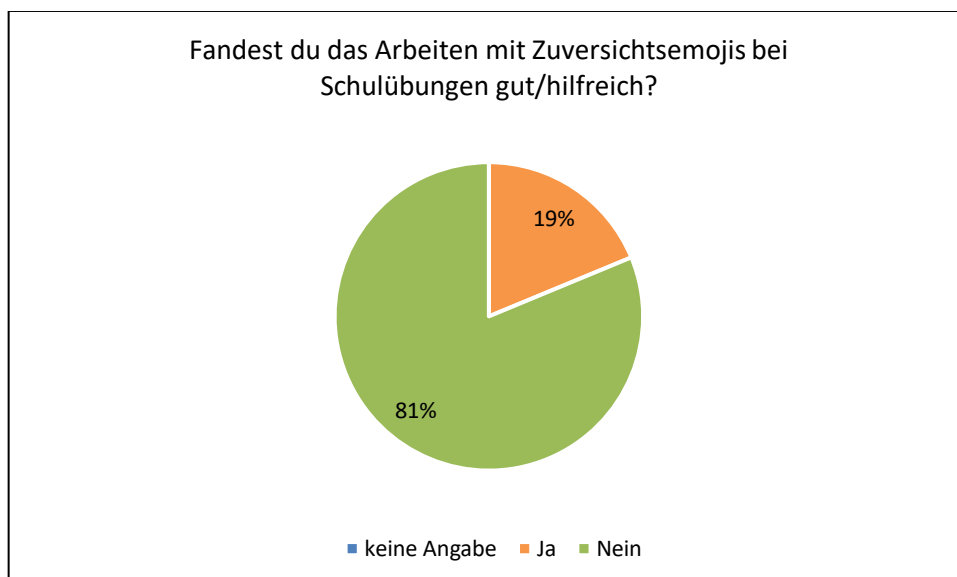
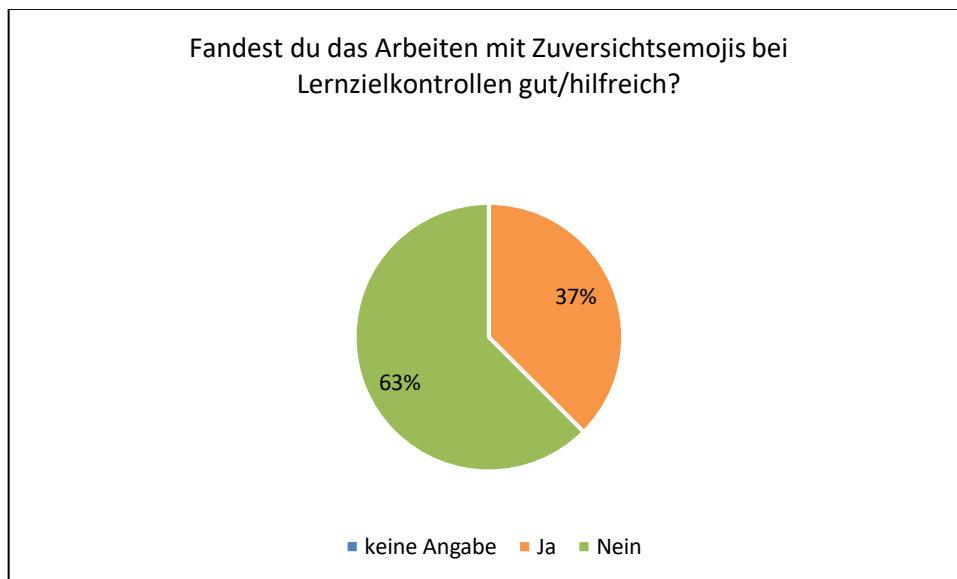


Abbildung 6.12

Einsatz von Zuversichtsemojis bei Schulübungen

Aus dem Kreisdiagramm in Abbildung 6.12 ergibt sich, dass ein Großteil (81%) das Arbeiten mit Zuversichtsemojis nicht gut fanden. Unter anderem schrieben die Schüler*innen, dass es unnötig sei bei Schulübungen (z.B. „Es ist unnötig wenn man alles zusammen macht“ (E1I2.2), „weil wir das eh mit der Lehrerin machen zsm. [zusammen]“ (E2I2.2)). Beim Lesen der Begründungen fällt auf, dass sich viele wiederholen – auch beim Betrachten der weiteren Items (v.a. „Hat nichts gebracht“ (E9I2.2)). Lediglich drei Schüler*innen (n=16) empfanden die Zuversichtsemojis als gut und hilfreich. Sie schrieben, dass es ihnen bei der Hausübung helfe und dass es Spaß mache. Folgendes Kreisdiagramm stellt nun die Ergebnisse für Lernzielkontrollen dar.

**Abbildung 6.13**

Einsatz von Zuversichtsemojis bei Lernzielkontrollen

Wie sich aus Abbildung 6.13 ablesen lässt, ist der Großteil (63%) der Klasse nicht der Meinung, dass der Einsatz von Zuversichtsemojis bei Lernzielkontrollen sinnvoll sei. Interessant ist, dass dies drei dieser Schüler*innen mit Zeitmangel begründeten (z.B. „weil man unnötig den zeit verschwindet“ (E4I2.3), „Man hat nicht genug Zeit“ (E5I2.3). Am Anfang der Lernzielkontrolle wurde mehrfach betont, dass für das Ankreuzen der Zuversichtsemojis zusätzlich Zeit zur Verfügung stehe, wodurch diese Begründungen und dadurch auch die Auswahl der Antwort obsolet ist. Nimmt man diese drei Personen in Bezug auf dieses Item aus der Wertung, sind es lediglich 53,8%, die den Einsatz bei Lernzielkontrollen als nicht nützlich empfinden. Ein paar Schüler*innen waren anderer Meinung, diese vertreten den Standpunkt, dass es sowohl ihnen (z.B. „Es hat mir dabei manchmal geholfen“ (E7I2.3) als auch der Lehrperson (z.B. „Es hilft Lehrerin.“ (E1I2.3)) helfe. Die Begründung einer weiteren Person wird hier ebenfalls zur Sprache gebracht: „Dann kann die Lehrerin sehen wo man sich schwer tut“. Item 3 erfragte, was den Schüler*innen am Arbeiten mit den Zuversichtsemojis gefallen hat. Hier hatten sie die Möglichkeit unter drei vorgegebenen Antworten zu wählen oder/und selbst etwas anzuführen, wobei zu beachten ist, dass Mehrfachnennungen möglich waren. 9 Schüler*innen gaben an, dass sie gelernt haben sich besser einzuschätzen („Ich habe gelernt mich besser einzuschätzen“ (I3)), 8 Schüler*innen, dass sie dadurch wussten, welches Thema sie noch üben sollten („Ich habe erkannt, welche Themen ich noch üben sollte.“ (I3)). 12 von 16 Schüler*innen, also 75%, entschieden sich unter anderem für die dritte Antwortmöglichkeit, die aussagt, dass sie dadurch erkannt haben, welche Themen sie schon gut können („Ich habe erkannt, welches Thema ich schon gut gefestigt habe.“ (I3)) Lediglich zwei Schüler*innen ergänzten die Antwortmöglichkeiten, wobei

eine Person die Frage offensichtlich nicht vollständig gelesen oder nicht verstanden hat. Die zweite Begründung darf an dieser Stelle jedoch nicht unerwähnt bleiben. Einem*einer Schüler*in gefällt, dass er*sie dadurch gemerkt habe, dass er*sie manchmal etwas nicht könne und dass es im Anschluss beim Lernen geholfen habe („Ich habe gemerkt, dass ich manchmal etwas nicht kann“, „Ich habe gemerkt, dass es mir dann leichter fehlt etwas zu lernen“ (E5I3).

Mithilfe von Item 4 wurde ermittelt, was den Schüler*innen an der Arbeit mit Zuversichtsemojis nicht gefallen hat, um diese im Anschluss überdenken zu können. Fünf Schüler*innen gaben an, dass sie nichts ändern würden. Eine Schülerin merkte an, dass ihr die Begründung nicht gefallen habe. Weiters schrieb ein Schüler, dass „es nichts nutzt also wir könnten einfach die leraren [Lehrerin] sagen was wir nicht verstehen“ (E6I4). Weitere Schüler*innen empfanden es als anstrengend, die Sticker einkleben zu müssen. Nachfolgendes Kreisdiagramm (siehe Abbildung 6.14) gibt Auskunft darüber, wie es den Schüler*innen bei der Wahl eines Zuversichtsemojis erging und zeigt zufolge, dass es ihnen überwiegend leicht- oder gar sehr leichtgefallen ist. 31% fiel die Wahl eher schwer und ausschließlich 6% sehr schwer.

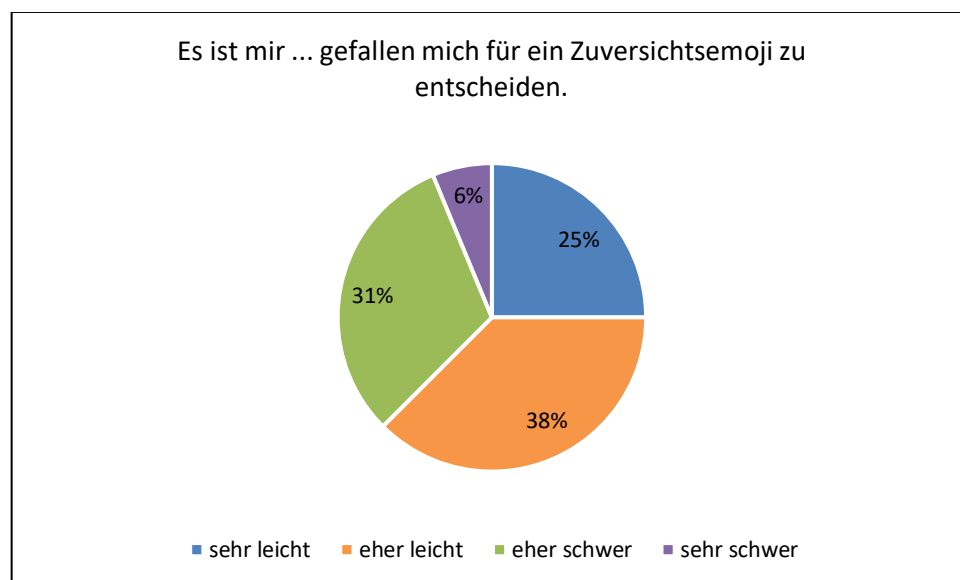


Abbildung 6.14

Entscheidung für ein Zuversichtsemoji (Item 5)

Thematisch schließt sich Item 6 an das vorherige an und erfragt mehr über die Entscheidung für ein Zuversichtsemoji. Es wurde befragt, ob das Begründen („weil:“ unterhalb der Zuversichtsemojis) bei der Entscheidung geholfen hat. 9 von 16 Schüler*innen gaben an, dass ihnen das Begründen bei der Entscheidung für ein Zuversichtsemoji geholfen habe. Item 7 ergab, dass 7 von

16 Schüler*innen wieder mit Zuversichtsemojis arbeiten wollen. Sie begründeten dies damit, dass es hilfreich für die Lehrerin sei und sie selbst dadurch einen Überblick erhalten, wie gut sie seien. Die Begründung einer Schülerin darf an dieser Stelle nicht unerwähnt bleiben: „weil ich erkennen kann ob ich Themen weiß“ (E4I7). 8 gaben gegenteiliges an und eine Person gab keine Angabe dazu. Dabei fielen Begründungen, dass es ihnen nicht helfe und dass „es nicht zu dem Thema Mathe passt“ (E7I7) 11 Schüler*innen würden nichts an der Arbeit mit Zuversichtsemojis ändern (Item 8). Zwei Schüler*innen würden die Begründung ändern, führten jedoch nicht aus, was sie daran ändern würden. An dieser Stelle wird auch die Idee einer Schülerin angeführt: „Das man sie [die Zuversichtsemojis] nicht selbständig einkleben soll sondern das die schon auf der Aufgabe drauf sind“ (E3I8) Im Anschluss werden nun ein paar Ergebnisse (siehe Abbildung 6.15) des letzten Items (I9) der Evaluierung angeführt, bei der die Schüler*innen eigene Zuversichtsemojis zeichnen konnten.

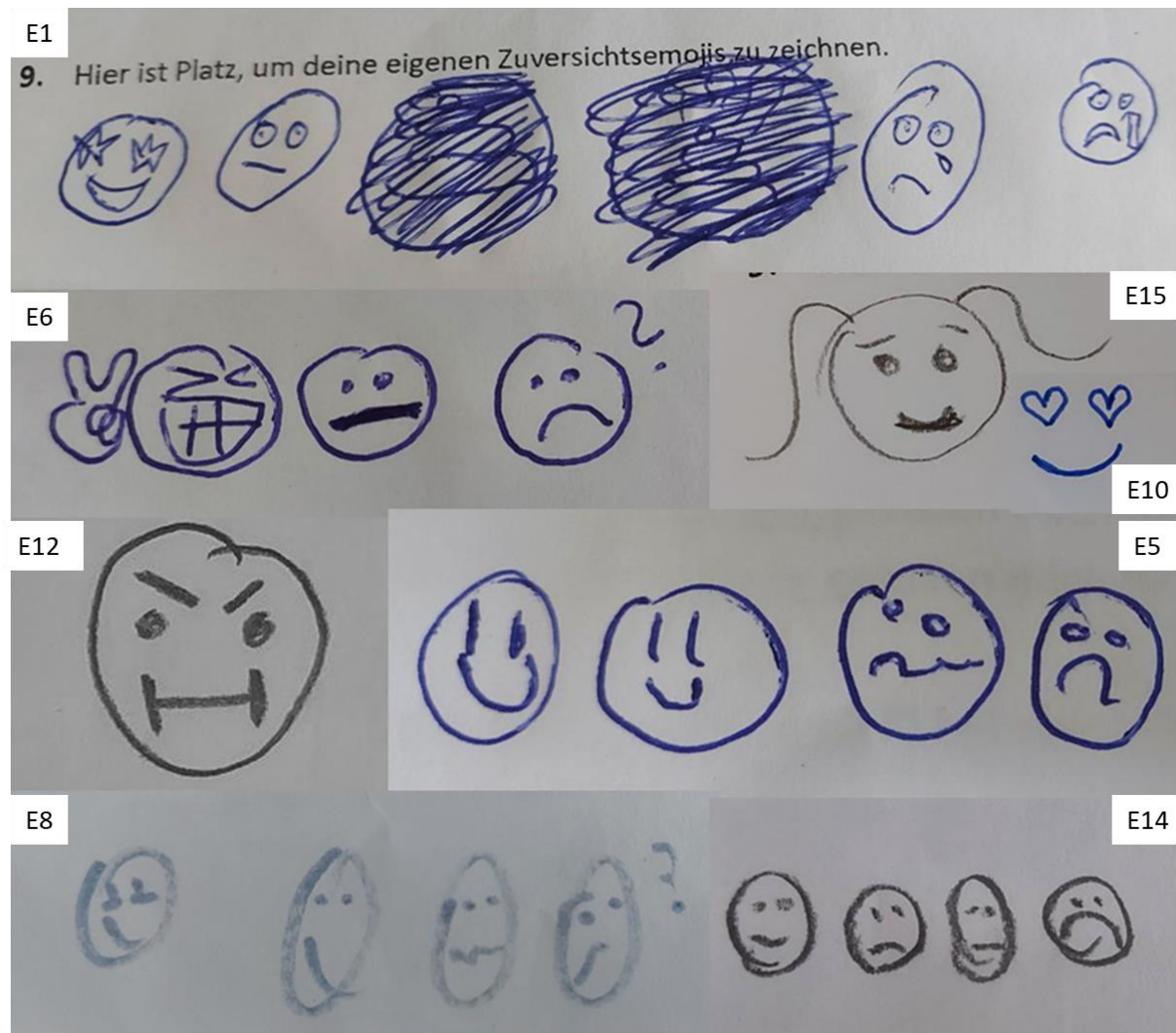


Abbildung 6.15
Zuversichtsemojis der Schüler*innen (Item 9)

7 Diskussion

Voriges Kapitel befasste sich mit der Darstellung der Ergebnisse, während dieses Kapitel die gewonnen Erkenntnisse diskutiert. Dies erfolgt unter Betrachtung der beiden Forschungsfragen.

7.1 Forschungsfrage 1

Die erste Forschungsfrage lautet wie folgt: Wie reagieren Schüler*innen auf den Einbezug von Zuversicht in Form von Zuversichtsemojis in den Unterricht? Wie fassen sie das auf und inwiefern sehen sie den Nutzen des Einbezugs bei Hausübungen, Schulübungen und Wissensüberprüfungen?

Die Schüler*innen haben die Bedeutung der einzelnen Zuversichtsemojis an sich schnell verstanden. Der Nutzen dessen war einigen Schüler*innen jedoch bis zum Ende der dreiwöchigen Intervention nicht klar und sie sahen darin bloß zusätzliche Arbeit. Dies wurde durch vereinzelte Fragen über den Bezug dessen zu Mathematik und anhand der Ergebnisse der Evaluierung erkennbar. Bei Schulübungen fanden die Zuversichtsemojis am wenigsten Anklang, da die Schüler*innen hier der Meinung waren, dass es nicht notwendig sei, wenn man viel gemeinsam arbeite. Auch bei Hausübungen waren 53% der Klasse der Ansicht, dass es nicht hilfreich wäre, da es nicht gebraucht werde. Dies könnte möglicherweise auf ein fehlendes Verständnis für den Nutzen der Zuversicht im Unterricht allgemein zurückzuführen sein. 41% hingegen gaben an, dass sie den Einbezug von Zuversichtsemojis bei Hausübungen gut und hilfreich fanden, da sie dadurch erkennen würden, was sie gut können und, dass sie es dadurch besser verstehen würden. Für die Forscherin sind diese positiven Kommentare Grund genug, um den Einsatz bei Hausübungen als positiv einzustufen. Auch der Einbezug bei Lernzielkontrollen wird positiv bewertet. Unter Berücksichtigung der Begründungen empfinden 53,8% den Einsatz von Zuversichtsemojis dabei als nicht hilfreich. Dazu muss gesagt werden, dass dieser Prozentsatz durch den Ausschluss von drei Angaben berechnet wurde. Dieser Ausschluss erfolgte aufgrund der Begründung, dass das Bearbeiten der Zuversichtsemojis ihnen Zeit der Lernzielkontrolle nehme. Da dieser Aspekt im Vorhinein mit der Klasse besprochen wurde und sie zusätzlich Zeit dafür bekamen, wurde beschlossen, diese aus dieser Statistik herauszunehmen. Positiv fielen hier die Begründungen von Schüler*innen auf, die die Meinung vertraten, dass es sowohl für sie als auch für die Lehrerin nützlich sei, und dadurch Wissenslücken erkennbar gemacht werden könnten. Weiters muss an dieser Stelle noch hervorgehoben werden, dass ein*eine Schüler*in schrieb, dass es ihm*ihr dadurch auch leichter falle zu lernen. 63% fiel es leicht, sich für ein

Zuversichtsemoji zu entscheiden, was die Wahl der Anzahl an Skalenpunkten bestätigt, da eine zu ausdifferenzierte Skala zu Schwierigkeiten führen könnte.

Aus den Ergebnissen lässt sich ableiten, dass die Schüler*innen unterschiedlichste Einstellungen zum Einbezug der Zuversicht in den Unterricht hatten. Die Tendenz ist beim Großteil der Fragen der Evaluierung jedoch negativ. Einigen negativen Kommentaren wie „Braucht man nicht.“ oder „Ist Papierverschwendung.“ wurde nicht viel Bedeutung zugemessen, da diese Kommentare von jenen Schüler*innen mehrmals bei verschiedenen Fragen geschrieben wurden. Es gab jedoch genauso Kritikpunkte, die in einer erneuten Studie oder für einen Einbezug in den Unterricht generell berücksichtigt werden könnten. Die Anzahl der positiven Kommentare über das Erkennen von Wissenslücken und das leichtere Lernen dadurch überwiegt in den Augen der Forscherin, trotz negativer Tendenz bei vielen Fragen der Evaluierung.

Die Forscherin ist der Ansicht, dass den zuvor negativen Kommentaren ein zu geringes Verständnis für den Nutzen von Zuversicht im Mathematikunterricht zugrunde liegt. Dies könnte dadurch verschuldet sein, dass nicht alle Schüler*innen bei der Einführung der Zuversichtsemojis anwesend waren. Ein weiterer Grund für die vielen negativen Antworten könnte die generelle negative Einstellung gegenüber der Schule sein, die möglicherweise bei manchen Schüler*innen durch die Pubertät verstärkt wird. Ein weiterer Punkt, der an dieser Stelle nicht unerwähnt bleiben darf, ist, dass die Schüler*innen das Wort Zuversicht vor der Einführung gar nicht kannten. Es wäre möglich, dass die Zeit nicht ausreichend war, um die Bedeutung des Wortes genau zu verstehen. Auffallend war auch, dass viele Schüler*innen „sicher sein“ verwendeten, anstatt „zuversichtlich sein“. Es stellt sich daher die Frage, ob die Einführung entsprechend anders gestaltet hätte werden müssen. Ebenso hätten Nutzungsgespräche während dieser Intervention hilfreich sein können, sodass die Schüler*innen mehr Verständnis erlangen. Als Idee könnte man dabei je zwei Schüler*innen über den Nutzen von Zuversicht im Unterricht sprechen und Notizen dazu aufzeichnen lassen. Weiters könnte ein Einbezug in mehreren Unterrichtsgegenständen von Vorteil sein, da dadurch klar werden würde, dass die Zuversicht in jedem Unterrichtsfach Platz finden könnte. Dies wäre auch hilfreich in Bezug auf eine von einem Schüler oftmals gestellte Frage, was diese Zuversichtsemojis mit Mathematik zu tun hätten. Dadurch wäre erkennbar, dass es sich dabei viel mehr um ein generelles didaktisches Prinzip handelt als um eine didaktische Handlungsstrategie, die ausschließlich für den Mathematikunterricht einsetzbar ist. An dieser Stelle muss jedoch gesagt werden, dass die Zuversicht auf verschiedene Art und Weisen in den Unterricht eingebaut werden kann und dies abhängig vom Unterrichtsfach ist.

Einer der Kritikpunkte von Schüler*innen war es, dass sie das Einkleben der Zuversichtsemojis in Form von Stickern als mühsam und anstrengend empfanden. Dies wurde von einer Schülerin auch konkret bei der Evaluierung erwähnt. Es besteht die Vermutung, dass abgesehen von dieser Schülerin, nicht das Einkleben per se als anstrengend empfunden wurde, sondern viel mehr der Umstand, dass sie daran denken mussten, die Zuversichtsemojis einzukleben. Bei der Entwicklung der Zuversichtsemojis überlegte die Forscherin auch, ob es sinnvoll wäre, diese in Form von Stempeln den Schüler*innen zur Verfügung zu stellen. Diese Idee wurde jedoch von den Stickern abgelöst, da die Forscherin die Befürchtung hatte, die Schüler*innen würden die Stempel verlieren oder in der Schule vergessen und diese dann nicht zur Verfügung haben. Eine Schülerin merkte an, dass sie es besser fände, wenn die Zuversichtsemojis bereits auf der Aufgabe drauf wären. Werden Arbeitsblätter verwendet, können die Zuversichtsemojis – wie auch in dieser Studie – verständlicherweise gleich draufgedruckt werden. Da es im realen Schulalltag nicht für jede Aufgabe ein eigenes Arbeitsblatt gibt, sondern auch Aufgaben aus dem Buch gelöst werden, benötigt es eine Möglichkeit, die Zuversichtsemojis zu inkludieren. Die Forscherin würde sich trotz des Kritikpunktes wieder für Zuversichtsemojis in Form von Stickern entscheiden. Allerdings würde sie diese nicht wie in dieser Studie jedes Mal in der entsprechenden Anzahl der Hausübungsbeispiele austeilen, sondern allen Schüler*innen am Beginn einen A4 Bogen mit Stickern zur Verfügung stellen. Dieser könnte je nach gewähltem Unterrichtsmittel in der entsprechenden Mappe oder im entsprechenden Heft gelagert werden und somit immer griffbereit sein, wenn die Schüler*innen ihre Aufgabe lösen. Zudem würde dadurch Zeit im Unterricht gespart werden.

7.2 Forschungsfrage 2

In diesem Unterkapitel werden nun Ergebnisse unter Betracht der zweiten Forschungsfrage diskutiert. Diese lautet wie folgt: Wie ist der Zusammenhang zwischen Zuversicht und Kompetenz von Schüler*innen im Bereich Prozentrechnung? Wie gut sind diese darin „kalibriert“? Haben Schülerinnen anders angekreuzt als Schüler?

Analog zu einer Studie von Foster (2016) wurde der Zusammenhang zwischen der Anzahl an korrekten Items und der durchschnittlichen Zuversicht berechnet. Dieser ist mit einem Korrelationskoeffizienten von $r=,581$ ($N=20$, $p<,01$) als stark einzustufen. In zuvor erwähnter Studie (Foster, 2016) wurde daraus geschlossen, dass die Schüler*innen gut kalibriert seien. Die Forscherin stellt die Aussagekraft dieser Korrelation jedoch in Frage, da die Anzahl an korrekten Items und die durchschnittliche Zuversicht ihrer Meinung nach keine Aussagekraft in Bezug darauf hat, wie gut

Schüler*innen in einem Bereich kalibriert sind. Dies begründet sie damit, dass bei dieser Berechnung der Zusammenhang von Kompetenz und Zuversicht nicht für jedes Item einzeln berücksichtigt wird. Dadurch wird nicht beachtet, ob Schüler*innen sich bei einzelnen Items gut eingeschätzt haben oder nicht. Der Forscherin geht es vielmehr darum, ob Kompetenz und Zuversicht bei einzelnen Items zusammenpassen (siehe Abbildung 3.3), woraus im Weiteren geschlussfolgert werden kann, ob Schüler*innen in einem gewissen Bereich gut kalibriert sind beziehungsweise, ob diese sich darin gut einschätzen können. Diese Analyse ergab, dass sich die Schüler*innen bei 7,15 von 11 Items gut einschätzten. Mit $r=,505$ ($N=20$, $p<,05$) besteht ein stark positiver Zusammenhang zwischen der Anzahl an richtigen Antworten und der Anzahl an gut eingeschätzten Items. In Bezug auf alle Einschätzungen bei dieser Lernzielkontrolle kam es bei 65% zu guten Einschätzungen. Aufgrund dessen lässt sich sagen, dass die Schüler*innen im Bereich Prozentrechnung gut kalibriert sind. Zuletzt wurde noch der Zusammenhang zwischen der Anzahl an richtigen Antworten und der Überschätzung überprüft. Dies ergab, dass es mit höherer Anzahl an richtigen Antworten zu weniger Überschätzungen kommt ($r=-,479$; $p<,05$). Schülerinnen schätzten sich bezogen auf den Mittelwert besser ein als Schüler, dieser Unterschied ist jedoch nicht signifikant ($t(18) = 0,157$; $p=,877$). Absolut gesehen, kam es bei Schülern öfter zu Überschätzungen als bei Schülerinnen und bei jenen öfter zu Unterschätzungen. Dies entspricht den Ergebnissen früherer Studien (Foster, 2016). Der Geschlechterunterschied in Bezug auf Unterschätzung ($U=41,500$; $Z=-,758$; $p=,552$) und Überschätzung ($U=46,500$; $Z=-,232$; $p=,824$) war jedoch nicht signifikant.

8 Fazit und Ausblick

Zusammenfassend lässt sich die erste Forschungsfrage nun damit beantworten, dass die Einstellungen der Schüler*innen in Bezug auf die Zuversichtsemojis unterschiedlich ausfielen, jedoch eine negative Tendenz erkennbar war. Die Vermutung besteht, dass dem das fehlende Verständnis des Nutzens des Einbezugs der Zuversicht in den Unterricht zugrunde liegt. Am wenigsten nützlich empfanden die Schüler*innen die Zuversichtsemojis bei Schulübungen. Bei Hausübungen und Wissensüberprüfungen waren die Meinungen sehr gespalten, jedoch stuft die Forscherin diese vor allem aufgrund der positiven Begründungen der Schüler*innen positiv ein. Die statistischen Auswertungen ergaben, dass sich die Schüler*innen der untersuchten Klasse im Bereich der Prozentrechnung gut einschätzen können, ihre Kompetenz und Zuversicht passt demnach gut zusammen. Weitere Untersuchungen in Bezug auf einen Geschlechterunterschied brachten keine signifikanten Ergebnisse. Absolut betrachtet unterschätzten sich Schülerinnen öfter, wohingegen Schüler sich öfter überschätzten.

Der Autorin ist bewusst, dass eine Studie mit einer einzigen Klasse wenig allgemeingültige Aussagekraft hat. Trotz dessen zeigt es, dass der Einbezug der Zuversicht in den Unterricht Vorteile für Schüler*innen und Lehrer*innen mit sich bringt. Es fragt sich nun, warum dies nicht schon längst Einzug in unserem Bildungssystem gefunden hat. Klar ist jedoch, dass mithilfe weiterer Studien überprüft werden müsste, wie sich Schüler*innen in anderen Bereichen der Mathematik einschätzen. Um allgemeingültigere Ergebnisse zu erhalten, bedarf es jedoch größerer Samples. Es erfordert daher vor allem interessierte Lehrer*innen, die bereit sind, den Einbezug der Zuversicht in ihrem Unterricht auszuprobieren. Um Lehrer*innen zu erreichen, benötigt es demnach Lehrer*innenfortbildungen, um diese in die Thematik einzuführen und Anregungen zu geben, wie eine Implementierung von Confidence-Based Learning machbar wäre. Aufgrund der hohen Relevanz des Themas sieht die Autorin auch die Notwendigkeit der Implementierung der Zuversicht in Schulbüchern. Ebenso wäre ein Forschungsprojekt zum Spannungsfeld Zuversicht und Prüfungsangst notwendig.

9 Zusammenfassung

Die Zuversicht ist eine Thematik, die bislang noch wenig Aufmerksamkeit im schulischen Kontext erhalten hat. Bei Confidence-Based Learning handelt es sich um ein Lehr- und Lernsystem, bei dem die Zuversicht in die Erfüllung einer Aufgabe einbezogen wird. Wenn eine Person demnach Vertrauen darin hat, dass sie eine Aufgabe richtig gelöst hat, ist diese zuversichtlich. Durch den Einbezug der Zuversicht in eine Aufgabe ist es möglich fehlerhaftes Wissen und Antworten, die auf reinem Raten basieren, zu identifizieren. Wird der Zusammenhang zwischen Wissen und Zuversicht betrachtet, lassen sich verschiedene Stufen von Wissen (knowledge quadrants) festmachen: Misinformed, Uninformed, Doubt und Mastery. Vor allem letztere Stufe befähigt Personen dazu, produktive Handlungen zu tätigen. Die erste Studie zu Confidence-Based Learning wurde bereits im Jahr 1932 durchgeführt, jedoch lag der Fokus erst später auf dem Wissenserwerb. Studien ergaben, dass Frauen sich besser einschätzen können und weniger dazu tendieren, sich zu überschätzen. Auch der Zusammenhang zwischen Zuversicht und dem Wissenserhalt über einen gewissen Zeitraum wurde bereits erforscht. Demnach zeigte eine Studie, dass Personen, die die höchste Kategorie der Zuversicht wählten, nach einer Woche noch immer über 90% richtig beantworten konnten. Grundsätzlich lässt sich Confidence-Based Learning in drei Phasen unterteilen. Die erste Phase befasst sich mit der Diagnose mittels Confidence-Based Assessment, gefolgt von der Bereitstellung von auf die lernende Person zugeschnittenen Lernunterlagen. Zuletzt geht es um den Lernprozess an sich und den Aufbau von Zuversicht.

Confidence-Based Assessment und Confidence-Based Marking beschreiben das Einbeziehen der Zuversicht in Leistungsüberprüfungen und im weiteren Sinne das entsprechende Beurteilen. Für den Einbezug bei Überprüfungen gibt es verschiedene Skalen, die sich unter anderem in der Anzahl der Skalenpunkte oder Verwendung von Deskriptoren unterscheiden. Abhängig vom Zusammenhang der Zuversicht und der Kompetenz sind Schüler*innen gut „kalibriert“ oder überschätzen beziehungsweise unterschätzen sich. Ziel ist es, dass die zwei Komponenten genau zusammenpassen. Für jene Beurteilung, bei der die Zuversicht hinzugezogen wird, gibt es verschiedene Beurteilungsschemata, die in dieser Arbeit genauer beschrieben werden. Doch auch ohne die entsprechende Beurteilung lassen sich nützliche Informationen anhand des Einbezugs der Zuversicht ableiten. Die Studienlage im schulischen Mathematikunterricht stellt sich als äußerst schwach dar. Vor allem ist bei allen vorhandenen Untersuchungen die Zuversicht ebenso in die Beurteilung eingeflossen. In Kapitel 3.7 versuchte die Forscherin Möglichkeiten zu sammeln, wie Confidence-Based Assessment ohne viel Aufwand in den Unterricht implementiert werden kann.

All diesen Angaben der Zuversicht liegt eine Selbsteinschätzung, die Einschätzung der eigenen Person in Hinblick auf gewisse Fähigkeiten, zugrunde. Ein wesentlicher Faktor, der die Selbsteinschätzung maßgebend beeinflusst, ist das Selbstkonzept. In dieser Arbeit findet sich ein kurzer Einblick in die Selbstkonzeptforschung, der Fokus liegt aber auf dem akademischen Selbstkonzept. Das hierarchische Selbstkonzeptmodell unterstreicht vor allem die Hierarchie und die Multidimensionalität des Selbstkonzepts. Da Studien zeigten, dass es eine klare Trennung zwischen dem verbalen und mathematischen Selbstkonzept gibt, wurde das zuvor genannte Modell revidiert. 2018 wird ein neues integratives akademisches Selbstkonzeptmodell vorgestellt, das sich aus drei Theorien zusammensetzt: Internal/External Frame of Reference Modell (I/E), Reciprocal Effect Modell (REM) und Big-Fish-Little-Pond Effekt (BFLPE).

Schätzt man sich nicht richtig ein, kommt es entweder zu einer Über- oder Unterschätzung. Personen überschätzen sich vor allem in einfachen alltäglichen Situationen. Im Vergleich mit anderen Personen kann es hierbei zum above-average effect kommen, der unabhängig von der tatsächlichen Leistung einer Person auftritt. Bei der Selbstunterschätzung besteht die Tendenz Handlungen aufgrund der Unterschätzung eigener Fähigkeiten zu unterlassen. Im Gegensatz zum above-average effect kommt es dabei eher zum below-average effect, bei dem man sich selbst im sozialen Vergleich als systematisch schlechter einstuft. Der Dunning-Kruger Effekt beschreibt das Überschätzen von Personen mit geringen Fähigkeiten. Es wird darauf hingedeutet, dass die geringen Fähigkeiten dazu führen, die eigene Inkompetenz nicht zu erkennen. Dies wird als dual-burden account bezeichnet. Selbsteinschätzungen finden sich in verschiedenen schulischen Kontexten wieder. Hattie erkannte beispielsweise die Wichtigkeit der Selbstbeurteilung und misst diesem sogar den größten Einfluss auf die Leistungen von Schüler*innen bei.

Der empirische Teil dieser Arbeit setzt sich aus einer gemischten Methode zusammen. Die Forscherin implementierte die Zuversicht über einen Zeitraum von drei Wochen in ihren eigenen Unterricht in einer 7.Schulstufe an einer Wiener Mittelschule. Dafür wurden Zuversichtsemojis entwickelt, die die Schüler*innen in Form von Stickern bekamen. Die Zuversicht wurde in dieser Zeit sowohl bei Schulübungen, Hausübungen als auch bei Wissensüberprüfungen integriert. In dieser Zeit zeichnete die Forscherin Feldnotizen auf. Diese drei-wöchige Intervention beinhaltete eine Lernzielkontrolle und eine Evaluierung der Arbeit mit Zuversichtsemojis am Ende. Um die Forschungsfragen zu beantworten, wurden die Daten der Lernzielkontrolle mittels IBM SPSS Statistics analysiert und die Feldnotizen sowie die Daten, die den Hausübungen entnommen werden konnten, qualitativ analysiert. Die Evaluierung am Ende wurde mithilfe von Excel ausgewertet. Die Ergebnisse der Evaluierung zeigten eine negative Tendenz, wobei aber die Begründungen der positiven

Rückmeldungen optimistisch für den Einsatz stimmen. Die vielen negativen Antworten könnten auf ein fehlendes Verständnis für den Nutzen und die Abwesenheit einiger Schüler*innen bei der Einführung in die Thematik zurückzuführen sein. Bei Hausübungen und Wissensüberprüfungen stuft die Forscherin die Ergebnisse trotz der negativen Tendenz als positiv ein. Die statistische Auswertung ergab, dass sich die Schüler*innen der beteiligten Klasse im Bereich der Prozentrechnung gut einschätzen können und, dass kein signifikanter Geschlechterunterschied besteht. In Zukunft bedarf es weiterer Forschung in diesem Bereich. Eine Erforschung des Spannungsfeldes Zuversicht und Prüfungsangst ist ebenso notwendig.

10 Literaturverzeichnis

- Adams T. M. & Gary W. E. (2009). The importance of Confidence in improving Educational Outcomes. 2009 25-Annual Conference on Distance Teaching and Learning.
- Ahlgren, A. (1969, 5-8 Februar). *Reliability, Predictive Validity, and Personality Bias of Confidence-Weighted Scores*. American Educational Research Association Convention, Los Angeles, California.
- Alicke, M.D., Govorun, O. (2005). The better-than-average effect. In M. Alicke, D. Dunning & J. Krueger (Hrsg.), *The self in social judgement*. (S.83-106). Psychology PressEditors.
- Bandura. (1997). *Self-efficacy: the exercise of control*. W.H. Freeman and Company.
- Barr, D. A., & Burke, J. R. (2013). *Using confidence-based marking in a laboratory setting: A tool for student self-assessment and learning*. Journal of Chiropractic Education, 27(1), 21–26. <https://doi.org/10.7899/jce-12-018>
- Beer, & Hughes, B. L. (2010). Neural systems of social comparison and the “above-average” effect. *NeuroImage*, 49(3), 2671–2679. <https://doi.org/10.1016/j.neuroimage.2009.10.075>
- Blanz, M. (2015). *Forschungsmethoden und Statistik für die Soziale Arbeit: Grundlagen und Anwendungen* (1. Aufl.). Kohlhammer.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler: mit 87 Tabellen* (4. überarb. Aufl.). Springer.
- Braun, M.W. (2003). *Genauigkeit der Selbsteinschätzung beim Erwerb neuer Kompetenzen in Abhängigkeit von Kontrollmeinung, Erfahrung, Selbstaufmerksamkeit, Ängstlichkeit und Geschlecht* [Dissertation, Universität Bern]. Edudoc.ch. <https://edudoc.ch/record/17440?ln=de>
- Bruno, J. (1995) Information Reference Testing (IRT) in Corporate and Technical Training Programs. UCLA.
- Bruno, J. E. (1993). Using Testing to Provide Feedback to Support Instruction: A Reexamination of the Role of Assessment in Educational Organizations. *Item Banking: Interactive Testing and Self-Assessment*, 190–209. https://doi.org/10.1007/978-3-642-58033-8_16

- Bruno, J.E. (1987). Admissible probability measurement in instructional management. *Journal of Computer Based Instruction*, 14(1), 95-107
- Butterfield, B. & Metcalfe, J. (2001). Errors Committed With High Confidence Are Hypercorrected. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 27(6), 1491–1494. <https://doi.org/10.1037/0278-7393.27.6.1491>
- Cabigon, H. J. (1993). *Effects of Self-Assessment on Retention of Training*. (Masterarbeit, Psychologie). Defense Technical Information Center. <https://apps.dtic.mil/sti/citations/ADA264483>
- Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, 69(2), 136–145. <https://doi.org/10.1037/0022-0663.69.2.136>
- Cambridge University Press & Assessment. (o. D.). confidence. In *Cambridge Dictionary*. Abgerufen am 13. August 2022, von <https://dictionary.cambridge.org/de/worterbuch/englisch/confidence>
- Chambers, & Gvozdenko, E. (2007). Applying Computerised Testing & Certainty Based Assessment to Reveal More about Student Learning: From Theory To Practice. *International Journal of Learning*, 13(12), 205–216. <https://doi.org/10.18848/1447-9494/CGP/v13i12/45114>
- Chambers, J. R., & Windschitl, P. D. (2004). Biases in Social Comparative Judgments: The Role of Nonmotivated Factors in Above-Average and Comparative-Optimism Effects. *Psychological Bulletin*, 130(5), 813–838. <https://doi.org/10.1037/0033-2909.130.5.813>
- Chatterjee, R., & Mandal, J. K. (2016). A novel Learning Object framework for Confidence Based Learning. 2016 International Conference on Information Science and Communications Technologies (ICISCT), 1–6. <https://doi.org/10.1109/ICISCT.2016.7777403>
- Clark, C. (2020). *The impact of confidence-based marking on unit exam achievement in a high school physical science course* [Masterarbeit, University of Northern Iowa]. UNI ScholarWorks. <https://scholarworks.uni.edu/grp/1449/>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Erlbaum.

College Board (1976–1977). Student descriptive questionnaire. Princeton, Educational Testing Service.

Curdt-Christiansen, X.L. (2020). Observations and field notes: recording lived experiences. In J. McKinley & H. Rose (Hrsg.), *The Routledge Handbook of Research Methods in Applied Linguistics* (S. 336–347). Routledge. <https://doi-org.uaccess.univie.ac.at/10.4324/9780367824471>

Dudenredaktion. (o. D.). Zuversicht. In Duden online. Abgerufen am 11. August 2022, von <https://www.duden.de/node/211971/revision/998694>

Dudenredaktion. (o. D.). Zuversicht. In Duden online. Abgerufen am 11. August 2022, von <https://www.duden.de/node/211971/revision/998694>

Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In J. M. Olson & M. P. Zanna (Hrsg.), *Advances in experimental social psychology*, (44. Ausgabe, S. 247–296). Academic Press. <https://doi.org/10.1016/B978-0-12-385522-0.00005-6>

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology*, 58, 568–581. <http://dx.doi.org/10.1037/0022-3514.58.4.568>

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5, 69–106. <http://dx.doi.org/10.1111/j.1529-1006.2004.00018.x>

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology*, 57(6), 1082–1090. <https://doi.org/10.1037/0022-3514.57.6.1082>

Erikson, E.H. (1974). *Jugend und Krise: die Psychodynamik im sozialen Wandel*. Ernst Klett Verlag

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 552–564. <https://doi.org/10.1037/0096-1523.3.4.552>

- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 552–564. <https://doi.org/10.1037/0096-1523.3.4.552>
- Florian, T.P. (2010). *Confidence-based assessment in Moodle: Insights from teachers, administrators, and programmers* (Publikation Nr. 814) [Dissertation, Walden University]. Walden Dissertation and Doctoral Studies. <https://scholarworks.waldenu.edu/dissertations/814/>
- Foster, C. (2016). Confidence and competence with mathematical procedures. *Educational Studies in Mathematics*, 91(2), 271–288. <https://doi.org/10.1007/s10649-015-9660-9>
- Foster, C. (2017). The guessing game. *Teach Secondary*, 6(8), 85. <https://www.foster77.co.uk/Foster,%20Teach%20Secondary,%20The%20guessing%20game.pdf>
- Foster, C., Woodhead, S., Barton, C. & Clark-Wilson, A. (2022). School students' confidence when answering diagnostic questions online. *Educational Studies in Mathematics*, 109(3), 491–521. <https://doi.org/10.1007/s10649-021-10084-7>
- Foster. (2021). Implementing Confidence Assessment in Low-Stakes, Formative Mathematics Assessments. *International Journal of Science and Mathematics Education*. <https://doi.org/10.1007/s10763-021-10207-9>
- Gardner-Medwin A.R. & Gahan M (2003, Juli). Formative and Summative Confidence-Based Assessment. In *Proceedings of the 7th International Computer-Aided Assessment Conference* (S.147-155), Loughborough, UK.
- Gardner-Medwin, A. R. (1995). Confidence assessment in the teaching of basic science. *Research in Learning Technology*, 3(1). <https://doi.org/10.3402/rlt.v3i1.9597>
- Gardner-Medwin, A. R. (o. D.). *CBM Help: Certainty-Based Marking*. University College London. <https://www.ucl.ac.uk/lapt/lpcfhelp.htm>
- Gardner-Medwin, A.R. & Curtin, N.A. (2007, 29.-31.Mai). *Certainty-based marking (CBM) for reflective learning and proper knowledge assessment*. [Konferenzbeitrag]. REAP Int. Online Conf. on Assessment Design for Learner Responsibility, online.

- Gardner-Medwin, A.R. (2006). Confidence-Based Marking – towards deeper learning and better exams.
- Gardner-Medwin, T. (2018). The value of self-tests and the acknowledgement of uncertainty. In Luckin, R. (Hrsg.), *Enhancing Learning and Teaching with Technology – What the research says* (Kapitel 1.2). UCL IOE Press.
- Gardner-Medwin, T. (2019a). Certainty-Based Marking: stimulating thinking and improving objective tests. In Bryan, C., & Clegg, K. (Hrsg.), *Innovative Assessment in Higher Education: A Handbook for Academic Practitioners (2nd ed.)*. (S. 141-150). Routledge.
<https://doi.org/10.4324/9780429506857>
- Gardner-Medwin, T. (2019b, 1. Mai). *Rewarding the identification of uncertain and confident answers* [Konferenzbeitrag]. APT2019, London, UK.
- Glaser. (2019). Selbstüberschätzung. In *Risiko im Management* (S. 13–16). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-25835-1_4
- Goetz, T., Cronjaeger, H., Frenzel, A. C., Lüdtke, O., & Hall, N. C. (2010). Academic self-concept and emotion relations: Domain specificity and age effects. *Contemporary Educational Psychology*, 35(1), 44–58. <https://doi.org/10.1016/j.cedpsych.2009.10.001>
- Groccia, J. E. & Buskist, W. (2011). Need for evidence-based teaching. *New Directions for Teaching and Learning*, 2011(128), 5–11. <https://doi.org/10.1002/tl.463>
- Gvozdenko, E. & Chambers, D. (2007). Applying Computerised Testing & Certainty Based Assessment to Reveal More about Student Learning: From Theory To Practice. *International Journal of Learning*, 13(12), 205–216. <https://doi.org/10.18848/1447-9494/CGP/v13i12/45114>
- Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52(1), 123-142.
<https://doi.org/10.2307/1170275>
- Hassmén, P., & Hunt, D. P. (1994). Human Self-Assessment in Multiple-Choice Testing. *Journal of Educational Measurement*, 31(2), 149–160. <https://doi.org/10.1111/j.1745-3984.1994.tb00440.x>

- Hassmén, P., Hunt, D. P., & Dybeck, C. (2002). Effects of self-assessment on retention in rule-based learning. *Perceptual and motor skills*, 94(1), 296–306.
<https://doi.org/10.2466/pms.2002.94.1.296>
- Hassmén, P., Sams, M. R. & Hunt, D. P. (1996). Self-Assessment Responding and Testing Methods: Effects on Performers and Observers. *Perceptual and Motor Skills*, 83(3_suppl), 1091–1104.
<https://doi.org/10.2466/pms.1996.83.3f.1091>
- Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- Hendriks, W. J. A. J., Bakker, N., Pluk, H., de Brouwer, A., Wieringa, B., Cambi, A., Zegers, M., Wansink, D. G., Leunissen, R. & Klaren, P. H. M. (2019). Certainty-based marking in a formative assessment improves student course appreciation but not summative examination scores. *BMC Medical Education*, 19(1), 1–11. <https://doi.org/10.1186/s12909-019-1610-2>
- Hevner, K. (1930). Tests for esthetic appreciation in the field of music. *Journal of Applied Psychology*, 14(5), 470–477. <https://doi.org/10.1037/h0071346>
- Hevner, K. (1932). A Method of Correcting for Guessing in True-False Tests and Empirical Evidence in Support of IT. *The Journal of Social Psychology*, 3(3), 359–362.
<https://doi.org/10.1080/00224545.1932.9919159>
- Hunt, D.P. (1993). Human Self-Assessment: Theory and Application to Learning and Testing. In: Leclercq, D.A., Bruno, J.E. (eds) *Item Banking: Interactive Testing and Self-Assessment*. NATO ASI Series, vol 112. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-58033-8_15
- Hunt, D. P. (1982). Effects of human self-assessment responding on learning. *Journal of Applied Psychology*, 67(1), 75–82. <https://doi.org/10.1037/0021-9010.67.1.75>
- Hunt, D. P. (2003). The concept of knowledge and how to measure it. *Journal of Intellectual Capital*, 4(1), 100–113. <https://doi.org/10.1108/14691930310455414>
- Hunt, D. P., & Furustig, H. (1989). Being informed, being misinformed and disinformation: a human learning and decision making approach. Technical Report PM 56: 238.

- Issroff K. & Gardner-Medwin A.R. (1998). Evaluation of Confidence Assessment within Optional Coursework. In M. Oliver (Hrsg.), *Innovation in the evaluation of Learning Technology* (S.169-179). Univ. N. London Press
- James. (1890). *The principles of psychology: in two volumes*. 1. Macmillan.
- Koivula, N., Hassmén, P., & Hunt, D. P. (2001). Performance on the Swedish Scholastic Aptitude Test: Effects of Self-Assessment and Gender. *Sex Roles*, 44(11), 629–645.
<https://doi.org/10.1023/A:1012203412708>
- Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77(2), 221–232. <https://doi.org/10.1037/0022-3514.77.2.221>
- Kruger, J., & Dunning, D. (1999). *Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments*. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Leclercq, D., & Poumay, M. (2005). *Three metacognitive indices for realism in self-assessment*. [Université de Liège]. ORBi. <https://orbi.uliege.be/handle/2268/14822>
- Lichtenstein, S., Fischhoff, B., Lawrence, P. D. (1982). Calibration of Probabilities. The State of the Art to 1980. In D. Kahneman, P. Slovic, A. Tversky (Hrsg.), *Judgement under Uncertainty, Heuristics and Biases*. (S.306-334) Press Syndicate of the University of Cambridge.
- Lohaus, A., Vierhaus, M., Maass, A. (2010). Selbstkonzept. In: *Entwicklungspsychologie des Kindes- und Jugendalters für Bachelor*. Springer-Lehrbuch. Springer, Berlin, Heidelberg.
https://doi.org/10.1007/978-3-642-03936-2_13
- Lundeberg, M. A., Fox, P. W. & Punčochař, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, 86(1), 114–121.
<https://doi.org/10.1037/0022-0663.86.1.114>
- Marcia, J. E. (1980). Identity in Adolescence. In J. Adelson (Hrsg.), *Handbook of Adolescent Psychology* (S. 110-137). Wiley.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *Educational Research Journal*, 23, 129–149. doi:10.3102/00028312023001129.

- Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology*, 82(4), 646–656. <https://doi.org/10.1037/0022-0663.82.4.646>
- Marsh, H. W. (2005). Big-Fish-Little-Pond Effect on Academic Self-Concept [Der "Big-Fish-Little-Pond"-Effekt und das akademische Selbstkonzept]. *Zeitschrift für Pädagogische Psychologie / German Journal of Educational Psychology*, 19(3), 119–127. <https://doi.org/10.1024/1010-0652.19.3.119>
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1(2), 133–163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>
- Marsh, H. W., & Rowe, K. J. (1996). The Negative Effects of School-Average Ability on Academic Self-Concept: An Application of Multilevel Modelling. *Australian Journal of Education*, 40(1), 65–87. <https://doi.org/10.1177/000494419604000105>
- Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, 20(3), 107–123. https://doi.org/10.1207/s15326985ep2003_1
- Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology*, 89(1), 41–54. <https://doi.org/10.1037/0022-0663.89.1.41>
- Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, 80(3), 366–380. <https://doi.org/10.1037/0022-0663.80.3.366>
- Marsh, H. W., Chessor, D., Craven, R., & Roche, L. (1995). The Effects of Gifted and Talented Programs on Academic Self-Concept: The Big Fish Strikes Again. *American Educational Research Journal*, 32(2), 285–319. <https://doi.org/10.2307/1163433>
- Marsh, H. W., Parker, J. W., & Smith, I. D. (1983). Preadolescent self-concept: Its relation to self-concept as inferred by teachers and to academic ability. *British Journal of Educational Psychology*, 53(1), 60–78. <https://doi.org/10.1111/j.2044-8279.1983.tb02536.x>

- Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*, 54(2), 263–280. <https://doi.org/10.1037/dev0000393>
- Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish-little-pond effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, 20, 319–350. doi:10.1007/s10648-008-9075-6
- Marsh, H.W. (2007). Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. In R.P. Perry, J.C. Smart (Hrsg.) *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (S. 319–383). Springer, Dordrecht. https://doi.org/10.1007/1-4020-5742-3_9
- Marsh. (1990). The structure of academic self-concept : The Marsh/Shavelson model. *The Journal of Educational Psychology*, 82(4), 623–636.
- Marsh. (1990). The Structure of Academic Self-Concept. *Journal of Educational Psychology*, 82(4), 623–636. <https://doi.org/10.1037/0022-0663.82.4.623>
- Massey. (2022). Using Emojis and drawings in surveys to measure children's attitudes to mathematics. *International Journal of Social Research Methodology*, 25(6), 877–889. <https://doi.org/10.1080/13645579.2021.1940774>
- McIntosh, R.D., Fowler, E.A., Lyu, T., Della Sala, S. (2019). Wise up: Clarifying the role of metacognition in the Dunning-Kruger effect. In *Journal of Experimental Psychology: General*. (Band 148, Nr. 11, S. 1882–1897). <https://doi.org/10.1037/xge0000579>
- Menold, N. & Bogner, K. (2015). *Gestaltung von Ratingskalen in Fragebögen*. SDM-Survey Guidelines (GESIS Leibniz Institute for the Social Sciences). https://doi.org/10.15465/gesis-sg_015
- Möller, J., Trautwein, U. (2009). Selbstkonzept. In E. Wild, J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 179–203). Springer Medizin Verlag. https://doi.org/10.1007/978-3-540-88573-3_8
- Myers, D. B. (2009). *Social psychology* (10.Auflage). McGraw Hill.

- Novacek, P. (2013). Confidence-based assessments within an adult learning environment. In Sampson, D.G., Spector, J.M., Ifenthaler, D., Isaías, P. (Hrsg. Proceedings of the IADIS International Conference on Cognition and Exploratory Learning in the Digital Age (S. 403-406). IADIS Press
- Novacek, P. (2015). *The implementation of a confidence-based assessment tool within an aviation training program*. [Dissertation, University of North Texas]. UNT Digital Library. <https://digital.library.unt.edu/ark:/67531/metadc804869/>
- Omidbakhsh, M. (2016). *A Methodology for Confidence-based Adaptive Numeracy Skill Assessment in Healthcare* (Publikation Nr. 981832) [Dissertation, Concordia University]. Spectrum Research Repository
- Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 29(3), 261–265. <https://doi.org/10.1037/h0022125>
- Paterson, D. & Langlie, T. (1925). Empirical data on the scoring of true-false tests. *Journal of Applied Psychology*, 9(4), 339–348. <https://doi.org/10.1037/h0069813>
- Pavel, S., Robertson, M. F., & Harrison, B. T. (2012). The Dunning-Kruger Effect and SIUC University's Aviation Students. *Journal of Aviation Technology and Engineering*, 2(1), 125–129. <https://doi.org/10.5703/1288284314864>
- PEARL Projekt-Team. (o. D.). *Tools und Methoden der qualitativen Sozialforschung - einsetzbar zur Erforschung des Praktikums im Rahmen des Projekts PEARL* -. uibk. https://www.uibk.ac.at/projects/pearl/downloads/toolbox_1.pdf
- Porst, R. (2008). Fragebogen. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-90897-7>
- Rheinberg, F. (2014) Bezugsnormen und schulische Leistungsbeurteilungen. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 59 – 71). Beltz Verlagsgruppe

- Rodgers, Carol R. , Scott, Katherine H. (2008): The development of the personal self and professional identity in learning to teach. In Cochran-Smith, Marylin/ Feiman-Nemser, Sharon/ McIntyre, D. John and Demers, Kelly E. (Hg): Handbook of Research on Teacher Education. New York/London : Routledge and the Association of Teacher Educators. S. 732-755.
- Salehi, M., Sadighi, F., & Bagheri, M.S. (2015). *Comparing confidence-based and conventional scoring methods: The case of an English grammar class*. Journal of Teaching Language Skills, 33, 123-152.
- Sanchez, C. & Dunning, D. (2018, Jänner). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, 114(1), 10–28.
<https://doi.org/10.1037/pspa0000102>
- Schlösser, T., Dunning, D., Johnson, K. L. & Kruger, J. (2013, Dezember). How unaware are the unskilled? Empirical tests of the “signal extraction” counterexplanation for the Dunning–Kruger effect in self-evaluation of performance. *Journal of Economic Psychology*, 39, 85–100. <https://doi.org/10.1016/j.joep.2013.07.004>
- Seashore, C. E. (1929). MEIER-SEASHORE ART JUDGMENT TEST. *Science*, 69(1788), 380.
<https://doi.org/10.1126/science.69.1788.380>
- Seaton, M., Marsh, H. W., Parker, P. D., Craven, R. G., & Yeung, A. S. (2015). The Reciprocal Effects Model Revisited: Extending Its Reach to Gifted Students Attending Academically Selective Schools. *Gifted Child Quarterly*, 59(3), 143–156.
<https://doi.org/10.1177/0016986215583870>
- Shavelson, Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research*, 46(3), 407–441.
<https://doi.org/10.3102/00346543046003407>
- Sheldrake. (2016). Confidence as motivational expressions of interest, utility, and other influences: Exploring under-confidence and over-confidence in science students at secondary school. *International Journal of Educational Research*, 76, 50–65.
<https://doi.org/10.1016/j.ijer.2015.12.001>

- Slavin, R. E. (2008). Perspectives on Evidence-Based Research in Education—What Works? Issues in Synthesizing Educational Program Evaluations. *Educational Researcher*, 37(1), 5–14.
<https://doi.org/10.3102/0013189X08314117>
- Smrkolj, P., Bančov, E. & Smrkolj, V. (2022). The Reliability and Medical Students' Appreciation of Certainty-Based Marking. *International Journal of Environmental Research and Public Health*, 19(3), 1706. <https://doi.org/10.3390/ijerph19031706>
- Snow, E. (2019). *Effects Of A Confidence-Based, Individualized Remediation Strategy On Student Learning And Final Grades In A Multi-Campus Human Anatomy Curriculum*. (Publikation Nr. 2578) [Dissertation, University of North Dakota]. Theses and Dissertations.
<https://commons.und.edu/theses/2578>
- Spradley, J. (1980). Participant observation. Holt, Rinehart and Winston.
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39(2), 111–133.
https://doi.org/10.1207/s15326985ep3902_3
- Valentine, J.C., & DuBois, D.L. (2005). Effects of self-beliefs on academic achievement and vice versa. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Hrsg.), *International advances in self research: The new frontiers of self research* (Band 2, S.53-78). Information Age Publishing
- Vierlinger, R. (1999). *Leistung spricht für sich selbst: "direkte Leistungsvorlage" (Portfolios) statt Ziffernzensuren und Notenfetischismus*. Dieck-Verlag.
- Webb, James M.; Stock, William A.; McCarthy, Michael T. (1994). The Effects of Feedback Timing on Learning Facts: The Role of Response Confidence. *Contemporary Educational Psychology*, 19(3), 251–265. doi:10.1006/ceps.1994.1020

11 Abbildungsverzeichnis

Abbildung 2.1 Zusammenhang von Richtigkeit, Zuversicht, dass eine Antwort richtig ist und Nützlichkeit des Wissens (Hunt, 2003).....	6
Abbildung 2.2 Knowledge quadrants (modifiziert nach Bruno, 1987)	7
Abbildung 3.1 Interface einer Frage mit 5-Punkte-Skala (Gvozdenko & Chamber, 2007, S.207)	19
Abbildung 3.2 Angabe der Zuversicht mittels „Confidence emojis“ (modifiziert nach Foster et al., 2022, S. 493).....	20
Abbildung 3.3 Zusammenhang zwischen Zuversicht und Kompetenz (modifiziert nach Foster, 2016, S.274)	21
Abbildung 3.4 Unterschiedliche Level an Zuversicht über eine korrekte Aussage (Gardner-Medwin, 2019b, S.12)	23
Abbildung 3.5 UCL Beurteilungsschema für Richtig/Falsch-Fragen und die Bereiche, wann welcher Grad an Zuversicht optimal ist (Gardner-Medwin & Gahan, 2003, S. 3)	25
Abbildung 3.6 Durchschnittlicher Prozentsatz an richtig beantworteten Fragen für Richtig/Falsch-Fragen für die unterschiedlichen Level an Zuversicht (C=1, 2 oder 3) und für die unterschiedlichen Geschlechter (Gardner-Medwin & Gahan, 2003, S. 5)	31
Abbildung 3.7 Aufgabenblatt der Schüler*innen (Foster, 2016, S. 278)	34
Abbildung 3.8 Typische Frage von Eedi (Foster et al., 2022, S.493)	37
Abbildung 4.1 Das hierarchische Selbstkonzeptmodell (modifiziert nach Möller & Trautwein, 2009, S. 182)	45
Abbildung 4.2 Struktur des revidierten akademischen Selbstkonzeptmodells (modifiziert nach Marsh et al., 1988)	46
Abbildung 4.3 Internal/External Frame of Reference Modell: Zusammenhang zwischen verbalen Fähigkeiten, mathematischen Fähigkeiten, verbalem Selbstkonzept und mathematischem Selbstkonzept (Möller & Trautwein, 2009, S. 194)	48
Abbildung 4.4 Big-fish-little-pond effect (Marsh, 2005, S. 120)	51
Abbildung 4.5 Integratives akademisches Selbstkonzeptmodell (Marsh et al., 2018, S. 265)	54
Abbildung 4.6 Effektstärke-Barometer für Selbstbeurteilung nach Hattie (2009, S. 44)	60
Abbildung 5.1 Zuversichtsemojis	65
Abbildung 5.2 Studiendesign	66
Abbildung 5.3 Aufgabe zu Prozentrechnung mit Angabe der Zuversicht.....	68
Abbildung 5.4 Hausübung zur Prozentrechnung mit Angabe der Zuversicht und Begründung	69
Abbildung 5.5 Fragebogen zur Evaluierung der dreiwöchigen Intervention	71
Abbildung 6.1 Erstes Zuversichtsemoji.....	76
Abbildung 6.2 Zweites Zuversichtsemoji.....	76
Abbildung 6.3 Drittes Zuversichtsemoji	77
Abbildung 6.4 Viertes Zuversichtsemoji	77
Abbildung 6.5 Beschreibung der Zuversichtsemojis an der Tafel	78

Abbildung 6.6	Hausübung von Schülerin M. (S19)	80
Abbildung 6.7	Hausübung von Schüler A. (S2)	82
Abbildung 6.8	Hausübung von Schülerin I. (S13)	82
Abbildung 6.9	Zusammenhang zwischen der Anzahl richtiger Antworten und der Anzahl gut eingeschätzter Items pro Schüler*in.....	86
Abbildung 6.10	Meinung über das Arbeiten mit Zuversichtsemojis im Allgemeinen	88
Abbildung 6.11	Einsatz von Zuversichtsemojis bei Hausübungen	88
Abbildung 6.12	Einsatz von Zuversichtsemojis bei Schulübungen.....	89
Abbildung 6.13	Einsatz von Zuversichtsemojis bei Lernzielkontrollen.....	90
Abbildung 6.14	Entscheidung für ein Zuversichtsemoji (Item 5)	91
Abbildung 6.15	Zuversichtsemojis der Schüler*innen (Item 9)	93

12 Tabellenverzeichnis

Tabelle 2.1 Zusammenhang zwischen Wissen und Zuversicht (in Anlehnung an Hunt & Furustig, 1989)	8
Tabelle 2.2 durchschnittlicher Prozentsatz richtiger Antworten (und zugehörige Standardabweichungen) bei Multiple-Choice Tests (MC) und jenen mit Selbsteinschätzung (SA) in den Subtests WORD und DS für Frauen und Männer (modifiziert nach Koivula et al, 2001)	12
Tabelle 2.3 Prozentsatz, der eine Woche später noch korrekt beantworteten Fragen abhängig vom Grad an Zuversicht (in Anlehnung an Cabigon, 1993, S.28).....	13
Tabelle 2.4 durchschnittlicher Prozentsatz an richtigen Antworten nach einer bestimmten Zeit abhängig vom Grad an Zuversicht	14
Tabelle 3.1 Beurteilungsschema (UCL scoring scheme) für Richtig/Falsch-Fragen nach Gardner-Medwin (modifiziert nach Gardner-Medwin, 2006, S.2)	24
Tabelle 3.2 Beurteilungsschema für Multiple-Choice-Fragen nach Gardner-Medwin (2006)	26
Tabelle 3.3 Beurteilungsschema für Multiple-Choice-Fragen von Hassmén und Hunt (in Anlehnung an Hassmén und Hunt, 1994).....	27
Tabelle 3.4 Beurteilungsschema für Multiple-Choice-Fragen nach Davies (2002)	27
Tabelle 4.1 Vier Identitätsstadien nach Marcia (modifiziert nach Lohaus et al., 2010, S. 166)	43
Tabelle 6.1 Mittelwert und Standardabweichung der richtigen Antworten für jedes Item	83
Tabelle 6.2 Bewertung der Einschätzung abhängig von der Korrektheit des Items und der angegebenen Zuversicht	84
Tabelle 6.3 Anzahl richtiger Antworten und Anzahl an gut eingeschätzten Items pro Schüler*in	85
Tabelle 6.4 Deskriptive Statistik: mittlere Zuversicht pro Person, Anzahl richtiger Antworten und Anzahl gut eingeschätzter Items pro Schüler*in	85
Tabelle 6.5 Anzahl der Überschätzungen und Unterschätzungen bei einer Gesamtanzahl von 220 Einschätzungen im Rahmen der Lernzielkontrolle.....	87

13 Anhang

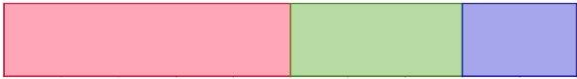
13.1 Anhang 1: Lernzielkontrolle

Name: _____

LERNZIELKONTROLLE – PROZENTRECHNUNG

1. Ergänze!																			
Bruchschreibweise	Prozentschreibweise	Dezimalschreibweise																	
	25%																		
		0,1																	
$\frac{1}{2}$																			
			__/6																
2. Kreuze an, welcher der Begriffe im Text unterstrichen ist – Grundwert (G), Prozentwert (W) oder Prozentsatz (p)?																			
			<table border="1"> <thead> <tr> <th></th> <th>G</th> <th>W</th> <th>p</th> </tr> </thead> <tbody> <tr> <td>Die gesamte Apfelernte betrug <u>650 kg</u>. Familie Huber hat davon 80% verkauft. Das sind 520 kg.</td> <td></td> <td></td> <td></td> </tr> <tr> <td>25% einer Klasse mit 20 Schüler*innen tragen eine Brille. Das sind <u>5 Schüler*innen</u>.</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Thomas ist schon <u>30%</u> von 900 km gefahren.</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>		G	W	p	Die gesamte Apfelernte betrug <u>650 kg</u> . Familie Huber hat davon 80% verkauft. Das sind 520 kg.				25% einer Klasse mit 20 Schüler*innen tragen eine Brille. Das sind <u>5 Schüler*innen</u> .				Thomas ist schon <u>30%</u> von 900 km gefahren.			
	G	W	p																
Die gesamte Apfelernte betrug <u>650 kg</u> . Familie Huber hat davon 80% verkauft. Das sind 520 kg.																			
25% einer Klasse mit 20 Schüler*innen tragen eine Brille. Das sind <u>5 Schüler*innen</u> .																			
Thomas ist schon <u>30%</u> von 900 km gefahren.																			
																			
3. Herr Sommer zeigt im Chemieunterricht insgesamt 120 Versuche. 90% davon laufen wie geplant. Über wie viele gelungene Versuche darf er sich freuen?																			
4. Berechne den Grundwert! $p = 50\%$ $W = 88 \text{ kg}$																			
5. Der Grundwert beträgt 30 000€. a. Berechne 1 ‰! b. Berechne 10 ‰!																			

Name: _____

6. In folgendem Prozentstreifen ($100\% \triangleq 100\text{ mm}$) sind die Lieblingsfarben von 100 befragten Personen dargestellt.  a. Wie viel Prozent nannten rot als ihre Lieblingsfarbe? _____ b. Wie viele Personen nannten grün als ihre Lieblingsfarbe? _____	____/2
7. Bei einem Baumarkt gibt es an einem Tag 30% Rabatt. Wie viel bezahlt man an diesem Tag für einen Akkubohrschrauber, der normalerweise 240 € kostet?	____/4
8. Von 25 Personen spielen 35% Fußball, 40% spielen Volleyball und 25% spielen Tennis. a. Welche Sportart ist am beliebtesten? _____ b. Stelle die Prozentsätze in einem Kreisdiagramm dar mit $r = 3\text{ cm}$! Denke an die Beschriftung!	____/1 ____/6

Name: _____

9. Kreuze an, ob die Aussagen richtig (r) oder falsch (f) sind.			
Aussage	r	f	
Die Mehrwertsteuer beträgt immer 10%.			
Die Mehrwertsteuer ist ein Preisnachlass.			
Die Mehrwertsteuer wird an den Staat gezahlt.			
			___/3
10. Sara kauft ein Fahrrad, das ohne Mehrwertsteuer 799€ kostet. Wie viel bezahlt sie für das Fahrrad mit 20% Mehrwertsteuer?			
			___/4
11. Erkläre den Begriff „Skonto“ in eigenen Worten!			
			___/2
Gutes Gelingen! 			
Erreichte Punkte: _____ von 40 $\hat{=}$ _____ %			