

Pascal Klein · Nicole Graulich
Jochen Kuhn · Maike Schindler *Hrsg.*

Eye-Tracking in der Mathematik- und Naturwissenschaftsdidaktik

Forschung und Praxis



Springer Spektrum

Eye-Tracking in der Mathematik- und Naturwissenschaftsdidaktik

Pascal Klein · Nicole Graulich · Jochen Kuhn ·
Maike Schindler
(Hrsg.)

Eye-Tracking in der Mathematik- und Naturwissenschafts- didaktik

Forschung und Praxis

Hrsg.

Pascal Klein
Fakultät für Physik, Institut für Dynamik
komplexer Systeme, Physik und ihre
Didaktik, Universität Göttingen
Göttingen, Niedersachsen, Deutschland

Nicole Graulich
Institut für Didaktik der Chemie
Justus-Liebig Universität Gießen
Gießen, Hessen, Deutschland

Jochen Kuhn
Fakultät für Physik
Lehrstuhl für Didaktik der Physik
Ludwig-Maximilians-Universität
München München, Bayern, Deutschland

Maike Schindler
Department Heilpädagogik
Universität zu Köln
Köln, Nordrhein-Westfalen
Deutschland

ISBN 978-3-662-63213-0 ISBN 978-3-662-63214-7 (eBook)

<https://doi.org/10.1007/978-3-662-63214-7>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert an Springer-Verlag GmbH, DE, ein
Teil von Springer Nature 2022

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht
ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags.
Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und
die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in
diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung
zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die
Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in
diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch
die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des
Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen
und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: Caroline Strunz

Springer Spektrum ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist
ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Vorwort der Herausgeber*innen

Liebe Leserinnen, liebe Leser,

die Zahl an Eye-Tracking-Studien in der fachdidaktischen Forschung nimmt seit einigen Jahren spürbar zu. Was können wir aus der Rekonstruktion von Blickbewegungen über das Lernen und Problemlösen aussagen? Wie bereichert die Erfassung visueller Aufmerksamkeit die bisherige mathematik- und naturwissenschaftsdidaktische Forschung? Inwiefern sind die Erkenntnisse auch für die Unterrichtspraxis relevant? In diesem Buch werden Beiträge einschlägiger Studien aus den verschiedenen Disziplinen aufgeführt, die einer Interessensgruppe die Anwendungsmöglichkeiten des Eye-Trackings in der didaktischen Forschung aufzeigen und einen Überblick über die Vielfalt der methodischen Herangehensweisen bei Eye-Tracking-Studien geben. Die Anwendungsszenarien reichen dabei von der Aufforderung, visuelle Stimuli sorgfältig zu betrachten und Muster zu erkennen, über das Lösen von Aufgaben bis zu Lernszenarien in Multimedia-Umgebungen. Die dabei verwendeten visuellen Stimuli spiegeln typische Unterrichtsmaterialien oder Repräsentationen der Biologie, Chemie, Informatik, Mathematik und Physik wider, wie z. B. Rechenrahmen, Textaufgaben, Programmcodes, Strukturformeln, elektrische Schaltungen, Diagramme und Versuchsaufbauten. In den Kapiteln wird dargestellt, wie die jeweilige Untersuchung umgesetzt und die Stimuli gestaltet und für die Analyse vorbereitet werden müssen, um aus den Eye-Tracking-Daten bedeutungsvolle Informationen zu gewinnen. Die tabellarische Übersicht am Ende dieses Vorworts charakterisiert das zentrale Forschungsanliegen der Kapitel kurz, nennt die disziplinäre Verortung und das untersuchte Lehr-/Lernszenario. Die Kapitel durchliefen vor der Publikation in diesem Herausgeberwerk einen anonymisierten Peer-Review-Prozess.

In Kap. 1 untersuchen **Nicole Graulich et al.** „idealtypische“ Blickmuster, die Expert*innen beim Problemlösen mit organisch-chemischen Repräsentationen zeigen. Diese Blickdaten werden dann zur Förderung schwächerer Studierender beim Lernen des gleichen Inhalts eingesetzt. Dazu wurde die visuelle Aufmerksamkeit von Studierenden und Expert*innen beim Lösen von Fallvergleichsaufgaben in der Organischen Chemie gemessen.

Wie das Blickverhalten beim Lernen in einer Multimedia-Umgebung überhaupt qualitativ beschrieben werden kann, untersuchen **Marit Kastaun** und **Monique Meier** in Kap. 2. Den Schüler*innen wurden bei der Planungsphase eines Biologie-experiments Unterstützungen durch statische und dynamische Repräsentationen

geboten. Die Eye-Tracking-Daten geben einen Aufschluss darüber, wie mit diesen verschiedenen Repräsentationstypen beim Lernen umgegangen wird.

Dinah Reuter und **Julia Bruns** kombinieren in Kap. 3 Verbal- und Eye-Tracking-Daten, um Einblicke in die Muster- und Strukturerkennung potenziell mathematisch begabter Kinder zu gewinnen. Dabei wird die methodische Frage gestellt, welche Rückschlüsse die Eye-Tracking-Methode auf die Erkennung von Mustern und Strukturen erlaubt und ob diese kongruent mit den verbalen Erklärungen der Lernenden sind.

In Kap. 4 nutzen **Bianca Watzka et al.** die Eye-Tracking-Methode, um den Umgang mit visuell angebotenen Hilfen in verschiedenen Lernszenarien zu physikalischen Themen zu erforschen. In drei separaten Studien werden die Blickdaten der Lernenden analysiert, um den Einfluss des Vorwissens, der Darbietungsart und der Interaktivität auf das Lernen und Problemlösen zu untersuchen.

Kap. 5 von **Maike Schindler** und **Florian Schindler** widmet sich der Frage, wie die Förderung von Grundschüler*innen mit Schwierigkeiten im Rechnenlernen in der Schulpraxis gelingen kann. Dabei wird Eye-Tracking als Instrument genutzt, um die Wirkung einer schuljahresbegleitenden Intervention auf Prozessebene zu evaluieren. Dafür wird die visuelle Aufmerksamkeit bei der Bearbeitung verschiedener Aufgaben am Rechenrahmen erfasst.

In Kap. 6 von **Sina Lenski** und **Jörg Großschedl** wird die Informationsaufnahme aus Concept Maps zu den Themen Zellen und Viren im Biologieunterricht mit unterschiedlicher Map-Struktur untersucht. Das Hauptaugenmerk liegt dabei auf dem Einfluss eines intensiven Trainings zum Umgang mit dieser Darstellungsform auf die visuelle Aufmerksamkeit beim Arbeiten mit Concept Maps.

Anna Lisa Simon et al. untersuchen in Kap. 7, ob sich Schüler*innen mit Rechenschwierigkeiten von solchen ohne auch hinsichtlich der kognitiven Anforderungen eines anderen mathematischen Inhaltsbereichs – der Geometrie – unterscheiden. Dafür zeigten die Autor*innen den Kindern eine Serie geometrischer Figuren, bei denen die Lernenden beurteilen sollten, ob es sich um Dreiecke handelt oder nicht. Eye-Tracking wurde bei diesem Entscheidungsprozess genutzt, um neben der Fehlerrate Aufschluss über die Informationsentnahme zu erhalten.

In Kap. 8 von **Kathrin Kennel et al.** werden die Blickbewegungen von Schüler*innen der gymnasialen Oberstufe beim Bearbeiten von Aufgaben zum grafischen Ableiten untersucht. Mit dem Ziel, Blickdaten in adaptiven Lernszenarien zu nutzen, wird in der beschriebenen Studie eine Grundlage gelegt, indem Blickunterschiede zwischen korrekten und inkorrekten Lösungen analysiert werden. Ebenso versuchen die Autor*innen, typische Fehler und Lernschwierigkeiten mit den Blickdaten in Verbindung zu bringen.

Kap. 9 von **Stefanie Lenzer** und **Andreas Nehring** liefert zunächst einen Vorschlag, wie hypothesenprüfende Eye-Tracking-Studien in den Naturwissenschafts-didaktiken konzipiert werden können. In einem zweiten Schritt wird vor dieser methodisch-konzeptionellen Arbeit eine konfirmatorische Eye-Tracking-Studie beschrieben, die untersucht, ob chemische Demonstrationsexperimente – mit oder ohne Berücksichtigung der Gestealtgesetze – von Schüler*innen visuell effektiver wahrgenommen werden.

Saskia Schreiter et al. (Kap. 10) untersuchten Lehramtsstudierende, die die Schwierigkeit von Bruchrechenaufgaben beurteilen sollten. Zur Schwierigkeitsbeurteilung wurden die beiden Kategorien fachlich und instruktional unterschieden, wobei einer Studierendengruppe diese Kategorien explizit mitgeteilt wurden. Eye-Tracking wurde nun in Kombination mit Post-hoc-Interviews genutzt, um einerseits diejenigen Aufgabenmerkmale zu identifizieren, die die angehenden Lehrkräfte zur Beurteilung heranzogen, und um andererseits Unterschiede zwischen den beiden Studierendengruppen zu finden.

In Kap. 11 stellen **Stefan Küchemann et al.** einen Literaturüberblick zu Eye-Tracking-Studien mit Graphen dar. Darin werden 27 Publikationen identifiziert, die das Blickverhalten beim Lernen und Problemlösen mit Graphen thematisieren, wobei die Schwerpunkte auf drei Aspekten liegen: der Kombination von Graphen mit anderen Darstellungsformen, dem Einfluss von Kontext und Darstellungsart auf das Lernen und Problemlösen sowie den Einflüssen weiterer Charakteristika beim Lernen und Problemlösen mit Graphen.

Larissa Hahn et al. (Kap. 12) erfassten die visuelle Aufmerksamkeit von Studierenden, während sie Aufgaben zu Vektorfeldern lösten. Die Studierenden sahen Vektorfelddiagramme, denen die Autor*innen verschiedene Komplexitätsgrade zuschrieben, und mussten die jeweils passende symbolisch-formale Darstellung (Formel) identifizieren. Eye-Tracking wurde in diesem Szenario als ein Instrument zur Test- und Itemanalyse genutzt, indem die a priori definierten Aufgabenmerkmale mit den Indikatoren der Blickbewegungen abgeglichen wurden.

Jennifer Dröse et al. untersuchten, wie Fünftklässler*innen mathematische Textaufgaben lesen. Dabei wurde Eye-Tracking genutzt, um speziell die Verarbeitungsprozesse syntaktischer Strukturen in Textaufgaben zu untersuchen. In Kap. 13 wird die Annahme untersucht, ob die Bearbeitung von Textaufgaben im Allgemeinen sowie die Wahrnehmung und Interpretation syntaktischer Strukturen im Speziellen mit der Lesekompetenz und der Mehrsprachigkeit der Lernenden zusammenhängen.

Antje Boomgarden et al. (Kap. 14) ließen Fünftklässler*innen zunächst Aufgaben zum Bruchrechnen lösen und zeigten ihnen anschließend eine Musterlösung, die sie nachvollziehen und mit der sie ihr Ergebnis vergleichen konnten. Eye-Tracking wurde in dieser Phase eingesetzt, um Aufschluss über diese Fehlerverarbeitungsprozesse zu erhalten. Die aufgezeichneten Blickmuster wurden dann als Grundlage für ein stimulierte Recall-Interview genutzt, um die Verbaldaten und die Blickdaten gemeinsam hinsichtlich der Verarbeitungsprozesse zu interpretieren.

In Kap. 15 untersuchen **Phillip Kater** und **Jan Vahrenhold** das Leseverhalten fortgeschrittener Studierender beim Arbeiten mit Programmen und Algorithmen. Die Autoren nutzen Eye-Tracking, um die Verweildauer auf Programmabschnitten unterschiedlicher Komplexität sowie auf Beweisskizzen zu untersuchen. Dies liefert Erkenntnisse über den Leseprozess von und über die Bearbeitungsunterschiede zwischen Programmen und Algorithmen.

Die Vielfalt der Kapitel zeigt auf, wie Eye-Tracking die bereits etablierten Forschungsmethoden ergänzen und welche Einblicke Eye-Tracking in die Lernprozesse ermöglichen kann, die in verschiedenen Altersstufen verortet sind: von der Grundschule über die Sekundarstufen 1 und 2 bis hin zu Studierenden.

Kapitel	Disziplin und Zielgruppe	Szenario/ Anforderung	Aufgabentyp und Domäne	Zentrales Forschungsanliegen
1	Chemiedidaktik, Studierende	Repräsentationen vergleichen	Fallvergleiche; Organische Chemie	Identifikation von Blickmustern
2	Biologiedidaktik, Sek. 1	Lernen in einer Multimedia-Umgebung	Verarbeiten verbaler/schriftlicher Information, Planungsphase Experiment (Variablenkontrolle)	Inhaltliche Beschreibung von Blickmustern
3	Mathematikdidaktik, Grundschule	Stimuli anschauen und erinnern	Indikatoraufgaben zur Struktur- und Mustererkennung, Arithmetik	Erkennbarkeit von Muster- und Strukturidentifikation
4	Physikdidaktik, Studierende und Sek. 2	1) Repräsentationen interpretieren und verknüpfen, 2) Videosexperiment, 3) Datenaufnahme	1) Elektrische Schaltungen manipulieren, 2) Fachwissen zum Experiment, 3) Übertragen von Messwerten in Diagramm	Einfluss verschiedener Hilfen auf Aufmerksamkeit und Wirksamkeit der Hilfen
5	Mathematikdidaktik, Übergang Grundschule/ Sek 1	Aufgaben lösen	Anzahlerfassung, Addition und Subtraktion am Rechenrahmen	Förderung bei Schwierigkeiten im Rechnenlernen
6	Biologiedidaktik, Sek. 1	Visuelle Stimuli betrachten, Replikation von Inhalt	Concept Maps zu Viren und Zellen	Einfluss eines Trainings auf den Umgang mit Concept Maps
7	Mathematikdidaktik, Übergang Grundschule/ Sek 1	Beurteilungsaufgabe lösen	Geometrie, Identifikation von Dreiecken	Unterschiede zwischen Kindern mit und ohne Rechenschwierigkeiten identifizieren
8	Mathematikdidaktik, Sek. 2	Aufgabe lösen	Grafisches Ableiten, Analysis	Identifikation von Blickunterschieden zwischen korrekten und inkorrekten Lösungen
9	Chemiedidaktik, Sek. 1 und 2	Stimuli betrachten (Momentaufnahme zum Versuchsaufbau)	Beobachtungsauftrag, Versuch zum Boyle-Experiment	Prüfung der Wirksamkeit von Gestaltprinzipien auf Indikatoren der Blickbewegung

Kapitel	Disziplin und Zielgruppe	Szenario/ Anforderung	Aufgabentyp und Domäne	Zentrales Forschungsanliegen
10	Mathematikdidaktik, Lehramtsstudierende	Diagnostik von Aufgabenschwierigkeit mittels visueller Stimuli	Schwierigkeits-erzeugende Merkmale identifizieren, Bruchrechenaufgaben	Aufmerksamkeit beim Beurteilen von Aufgaben; Unterschiede zwischen informierten und nichtinformierten Studierenden
11	Übergreifend	Literaturüberblick über die visuelle Aufmerksamkeit beim Arbeiten mit Graphen		
12	Physik-/ Mathematikdidaktik, Studierende	Aufgaben lösen	Formel einem Diagramm zuordnen, Vektorfelder	Einfluss des Komplexitätsgrades von Vektorfeldern auf die Blickdaten
13	Mathematikdidaktik, Sek. 1	Aufgaben lösen	Textaufgaben, Arithmetik	Wahrnehmen und Interpretieren syntaktischer Strukturen bei Kindern mit verschiedenen Sprachprofilen
14	Mathematikdidaktik, Sek. 1	Vergleich eigener Lösung mit Musterlösung	Aufgabe zum Bruchrechnen lösen, mit Lösung abgleichen	Identifikation von Fehlerbearbeitungsprozessen (nachvollziehen, vergleichen)
15	Informatikdidaktik, Studierende	Umgang mit Programmen, Programmieren und Beweisen, Fehlersuche	Input-Output (Programm); Lesen und Fehlersuche (Algorithmus)	Unterschiede im Leseverhalten zwischen Programm und Algorithmus

Eye-Tracking – ein Versprechen für die MINT-Bildung?

„Nihil est in intellectu quod non prius sit in sensu“, sagte schon der Theologe und Philosoph Thomas von Aquin und erklärte damit, dass wir das Wesen einer Sache nur dann erforschen können, wenn wir beachten, welche Sinnesempfindungen zu ihr geführt haben (Aquinas, n. d.). Möchten wir also den menschlichen Geist mitsamt seiner Entwicklung erforschen – kurzum das Lernen selbst –, so müssen wir unser Augenmerk darauf richten, welche Informationen die oder der Lernende aufnimmt und wie genau das vonstattengeht. Als Menschen verlassen wir uns in großem Maße auf den visuellen Sinn: unsere Augen. Die Informationsaufnahme über die Augen ist somit von zentraler Bedeutung für die Erforschung des Lernens an sich. Dies trifft besonders für MINT-Fächer zu, bei denen komplexe, wissenschaftliche Abbildungen eine große Rolle spielen. Die Methode zur Messung dieser visuellen Informationsaufnahme schlechthin ist Eye-Tracking (Blickerfassung; Holmqvist et al., 2011). Im Folgenden gehe ich kurz auf die wichtigsten Versprechen dieser Methode für die Zukunft der Bildung in MINT-Fächern ein, aber auch auf die Frage, wovor wir uns in Acht nehmen sollten.

Das am häufigsten verwendete Eye-Tracking-System besteht aus den folgenden Hardware- und Softwarekomponenten: Infrarotlicht(ern) mit Infrarotkamera(s) inklusive einer Bildverarbeitungssoftware. Des Weiteren werden oftmals Softwarepakete zur Darbietung von Stimuli sowie zur Weiterverarbeitung, Visualisierung und Analyse von Eye-Tracking-Daten angeboten. Diese Technologie dient der Messung von Bewegungen unserer Augäpfel im Verhältnis zu einem Stimulus – sei es eine digitale Mathematiklernumgebung auf dem Laptop, ein Informatikquiz auf dem Smartphone, eine Abbildung der Mechanik des Vogelfluges in einem Buch oder Schaltkreise, mit denen die Schüler*innen im Klassenraum arbeiten. Verschiedene Arten von Eye-Trackern erlauben es uns, in all diesen unterschiedlichen Situationen zu erfassen, wo eine Person hinschaut und wie lange ihr Blick dort verweilt. Darüber hinaus können wir erfassen, in welcher Reihenfolge diese Person verschiedene Elemente des Stimulus (z. B. bestimmte Teile einer Grafik sowie Teile eines dazugehörigen Textes) betrachtet und zu welchem Zeitpunkt das geschieht (z. B. im Verhältnis zur mündlichen Erklärung der Lehrkraft). Wir erfahren ebenfalls, welche Informationen der oder dem Lernenden entgangen sind. Basierend auf diesen Erkenntnissen können wir das Lern- bzw. Testmaterial oder gar den Kontext an sich so anpassen, dass es optimal der

menschlichen Informationsverarbeitung entspricht und somit zu effizientem Lernen oder aber validem Testen führt. Zu diesen Zwecken wird Eye-Tracking nunmehr seit über 20 Jahren erfolgreich in der Bildungsforschung eingesetzt (Alemdag & Cagiltay, 2018; Jarodzka et al., 2017; Lai et al., 2013). Dabei haben sich insbesondere vier Forschungsbereiche herauskristallisiert:

1. Eye-Tracking kann zur Beschreibung von Expertiseunterschieden in bestimmten Berufen und deren Ausbildung verwendet werden (Gruber & Jarodzka, 2018). Dazu zählen u. a. die Medizin (Gegenfurtner et al., 2011; Sheridan & Reingold, 2017), die Softwareprogrammierung (Emhardt et al., 2020), die Flugsicherung (Van Meeuwen et al., 2014) und die Zoologie (Jarodzka et al., 2010). In letzter Zeit eröffnet sich ein neues, besonders interessantes Forschungsgebiet für die Bildungswissenschaften, nämlich die visuelle Expertiseentwicklung von Lehrkräften (für zwei aktuelle Sonderhefte zu diesem Thema s. Jarodzka et al., 2021; Lachner et al., 2016). Dabei wird erforscht, welche Rolle die Fähigkeit die Schüler*innen zu beobachten, diese Beobachtungen im Kontext der Lehrsituation zu interpretieren und entsprechend zu reagieren, für die Klassenführung spielt.
2. Mittels Eye-Tracking können wir verstehen, wie Lernende die im Lern- oder Testmaterial angebotenen Informationen verarbeiten. Dies ist von zentraler Bedeutung für unseren Forschungszweig, da es dazu beitragen kann, bestehende Lehr- und Lerntheorien (Cognitive Theory of Multimedia Learning: Mayer, 2005; Cognitive Load Theory: Sweller et al., 2019) zu überprüfen und weiterzuentwickeln (Kirschner et al., 2016). Das wiederum erlaubt es uns, jenes Material dementsprechend anzupassen und somit die konkreten Designrichtlinien, die sich aus diesen Theorien ergeben, direkt auf ihre Wirksamkeit zu überprüfen (Alemdag & Cagiltay, 2018; Lindner, 2020).
3. Eye-Tracking kann dazu verwendet werden, die momentane Auslastung des Arbeitsgedächtnisses beim Lernen und Testen objektiv zu erfassen, z. B. durch Pupillenweitenmessung (Korbach et al., 2017; Szulewski et al., 2017).
4. Eye-Tracking ermöglicht es, direkt durch die Augen eines Experten zu schauen mittels sog. Eye Movement Modeling Examples (Van Gog et al., 2009). Dabei werden die Blickbewegungen einer Lehrperson auf Lehrmaterial übertragen, während sie die Aufgabe erklärt, damit die Lernenden wissen, worauf die Lehrperson sich bei der mündlichen Erklärung bezieht (z. B. in der Zoologie: Jarodzka et al., 2013; in der Biologie: Mason et al., 2015; in der Informationssuche im WWW: Salmerón et al., 2020; in der Geometrie: Van Marlen et al., 2018).

Trotz aller Chancen dürfen wir die Augen vor Herausforderungen des Eye-Trackings nicht verschließen. Insbesondere möchte ich auf zwei Punkte eingehen. Erstens dürfen wir nicht blind der Technik vertrauen (Kok & Jarodzka, 2017a, b). Eye-Tracking ist zum einen – trotz anders lautender Beteuerungen der Verkäufer – eine sehr komplizierte Apparatur, die nur von geschultem

Personal so bedient werden kann, dass die erhobenen Daten tatsächlich reliabel und valide sind. Zum anderen können durch mangelnde Erfahrung sehr leicht falsche Schlüsse aus der Erhebung gezogen werden, selbst wenn Daten durch erfahrene Versuchsleiter erhoben wurden. Wir müssen uns immer wieder vor Augen führen, dass Eye-Tracking wirklich nur die Bewegung der Augäpfel, der Pupillen sowie der Lider misst. Jedwede weiterführende Interpretation der Daten bezüglich der kognitiven Weiterverarbeitung der angeschauten Informationen muss mit entsprechender Vorsicht vorgenommen werden (Orquin & Holmqvist, 2018). Zweitens steht die Frage im Raum, inwiefern es sich um personenbezogene Daten handelt und wie mit diesen umgegangen werden kann. Bei Eye-Tracking-Daten handelt es sich z. T. durchaus um biometrische Daten, die auf eine Person zurückführbar sind. Eindeutige Fälle sind, wenn ein Eye-Tracker ein Bild oder gar ein Video der Iris der Versuchsperson erfasst oder wenn mittels einer mobilen Eye-Tracking-Brille ein Video von einem Klassenraum mit (gar minderjährigen) Schülern aufgenommen wird. Weniger eindeutig ist die Lage bei den Rohdaten im Textformat. Auf den ersten Blick scheint es sich um anonymisierte Datensätze zu handeln (vorausgesetzt, es werden keine Klarnamen der Versuchspersonen verwendet). Wenn man allerdings bedenkt, dass es sich um unglaublich reiche Datensätze handelt (zwischen 30 und 1000 Hz binokularer x- und y-Koordinaten inklusive Pupillenweite), und man um die idiosynkratischen Eigenschaften von Blickbewegungen weiß (Andrews & Coppola, 1999) und dabei noch die wachsenden Möglichkeiten der Big Data und AI-Algorithmen berücksichtigt, kann die Anonymität dieser Datensätze durchaus in Zweifel gezogen werden. Darüber hinaus muss man beachten, dass Tech-Giganten wie Facebook, Microsoft, Apple etc. seit einigen Jahren in Eye-Tracking investieren. Deswegen plädiere ich dafür, Eye-Tracking-Daten als personenbezogene Daten zu betrachten, was sich zwar nachteilig auf einen Open-Science-Ansatz zum Datensharing auswirkt, aber unsere Versuchspersonen und deren Daten maximal schützt.

Zusammenfassend kann man sagen, dass Eye-Tracking bereits jetzt viele bisher der Forschung nicht zugängliche, Forschungsfragen rund um individuelle Lern-, Lehr- und Testprozesse beantworten kann. In Anbetracht der schnellen technologischen Entwicklungen und des rasanten Anstiegs der Eye-Tracking-Nutzung in der MINT-Bildungsforschung (aber nicht nur!) ist zu erwarten, dass uns diese Technologie noch viele interessante Einsichten bringen wird. Gerade wegen dieses Potenzials müssen wir aber den Einsatz des Eye-Trackings in der Praxis stets kritisch reflektieren.

Halszka Jarodzka
Departement für Online-Lehre und
Instruktion an der Open Universiteit
Heerlen, Niederlande

Literatur

- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, 125, 413–428. <https://doi.org/10.1016/j.compedu.2018.06.023>.
- Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, 39(17), 2947–2953.
- Aquinas, T. (n. d.). *Quaestiones disputatae de veritate*. <http://www.corpusthomicum.org/qdv02.html>.
- Emhardt, S. N., Kok, E. M., Jarodzka, H., Brand-Gruwel, S., Drumm, C., & Gog, T. (2020). How experts adapt their gaze behavior when modeling a task to novices. *Cognitive Science*, 44(9). <https://doi.org/10.1111/cogs.12893>.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552. <https://doi.org/10.1007/s10648-011-9174-7>.
- Gruber, H., & Jarodzka, H. (2018). Expertise als Forschungsgebiet. In D. H. Rost, J. R. Sparfeldt, & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 185–192). Beltz.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154. <https://doi.org/10.1016/j.learninstruc.2009.02.019>.
- Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70. <https://doi.org/10.1016/j.learninstruc.2012.11.004>.
- Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, 10(1), 1–18. <https://doi.org/10.16910/jemr.10.1.3>.
- Jarodzka, H., Skuballa, I., & Gruber, H. (2021). Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom. *Educational Psychology Review*, 33(1), 1–10. <https://doi.org/10.1007/s10648-020-09565-7>.
- Kirschner, P. A., Park, B., Malone, S., & Jarodzka, H. (2016). Toward a Cognitive Theory of Multimedia Assessment (CTMMA). In M. J. Spector, B. B. Lockee, & M. D. Childress (Hrsg.), *Learning, Design, and Technology* (S. 1–23). Springer. https://doi.org/10.1007/978-3-319-17727-4_53-1.
- Kok, E. M., & Jarodzka, H. (2017a). Before your very eyes: The value and limitations of eye tracking in medical education. *Medical Education*, 51(1), 114–122. <https://doi.org/10.1111/medu.13066>.
- Kok, E. M., & Jarodzka, H. (2017b). Beyond your very eyes: Eye movements are necessary, not sufficient. *Medical Education*, 51(11), 1190–1190. <https://doi.org/10.1111/medu.13384>.
- Korbach, A., Brünken, R., & Park, B. (2017). Measurement of cognitive load in multimedia learning: A comparison of different objective measures. *Instructional Science*, 45(4), 515–536. <https://doi.org/10.1007/s11251-017-9413-5>.
- Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities (editorial to Special Issue). *Instructional Science*, 44(3), 197–203. <https://doi.org/10.1007/s11251-016-9376-y>.
- Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10, 90–115. <https://doi.org/10.1016/j.edurev.2013.10.001>.

- Lindner, M. A. (2020). Representational and decorative pictures in science and mathematics tests: Do they make a difference? *Learning and Instruction*, 68, 101345. <https://doi.org/10.1016/j.learninstruc.2020.101345>.
- Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology*, 41, 172–187. <https://doi.org/10.1016/j.cedpsych.2015.01.004>.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge Handbook of Multimedia Learning*, 41, 31–48.
- Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, 50(4), 1645–1656. <https://doi.org/10.3758/s13428-017-0998-z>.
- Salmerón, L., Delgado, P., & Mason, L. (2020). Using eye-movement modelling examples to improve critical reading of multiple webpages on a conflicting topic. *Journal of Computer Assisted Learning*. <https://doi.org/10.1111/jcal.12458>.
- Sheridan, H., & Reingold, E. M. (2017). The holistic processing account of visual expertise in medical image perception: A review. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01620>.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>.
- Szulewski, A., Kelton, D., & Howes, D. (2017). Pupillometry as a tool to study expertise in medicine. *Frontline Learning Research*, 5(3), 55–65. <https://doi.org/10.14786/flr.v5i3.256>.
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25(3), 785–791. <https://doi.org/10.1016/j.chb.2009.02.007>.
- Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2018). Effectiveness of eye movement modeling examples in problem solving: The role of verbal ambiguity and prior knowledge. *Learning and Instruction*, 58, 274–283. <https://doi.org/10.1016/j.learninstruc.2018.07.005>.
- Van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., de Bock, J. J. P. R., & Van Merriënboer, J. J. G. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, 32, 10–21. <https://doi.org/10.1016/j.learninstruc.2014.01.004>.

Inhaltsverzeichnis

1	Gibt es ideale Blickbewegungsmodelle zur Förderung der Lernenden beim Lösen organisch-chemischer Aufgaben?	1
	Nicole Graulich, Marc Rodemer, Julia Eckhard und Sascha Bernholt	
1.1	Einleitung	1
1.2	Theoretischer Hintergrund	2
1.3	Forschungsfragen	5
1.4	Material und Methode	6
1.5	Ergebnisse	9
1.6	Diskussion	14
	Literatur	16
2	Eine qualitative Analyse von Blickdaten bei statischen und dynamischen Repräsentationen im naturwissenschaftlichen Erkenntnisprozess	19
	Marit Kastaun und Monique Meier	
2.1	Einleitung	19
2.2	Theoretischer Hintergrund	20
2.3	Zielsetzung und Forschungsfragen	22
2.4	Material und Methode	22
2.5	Ergebnisse	27
2.6	Diskussion	33
	Literatur	36
3	Muster- und Strukturerkennungsprozesse bei potenziell mathematisch begabten Grundschulkindern untersuchen – Das Potenzial von Eye-Tracking und retrospektivem Interview im Vergleich	41
	Dinah Reuter und Julia Bruns	
3.1	Einleitung	41
3.2	Muster- und Strukturerkennung im Kontext von Begabung	42
3.3	Erfassung der Muster- und Strukturerkennung	43
3.4	Forschungsfragen	44
3.5	Material und Methode	44
3.6	Ergebnisse	47

3.7	Diskussion	54
	Literatur.	56
4	Eye-Tracking als Methode zur Analyse der Wirkung unterstützender Hinweise in digitalen Lernumgebungen	59
	Christoph Hoyer, Bianca Watzka, Maleen Hurzlmeier, Bernhard Ertl und Raimund Girwidz	
4.1	Einleitung und Forschungslage	60
4.2	Theoretischer Hintergrund	60
4.3	Beschreibung der Experimente	62
4.4	Diskussion	70
4.5	Resümee	72
	Literatur.	72
5	Förderung der strukturierten Anzahlerfassung bei Kindern mit Schwierigkeiten im Rechnenlernen: Eine explorative Eye-Tracking Studie zur Evaluation schulischer Förderung in Klasse 5.	75
	Maike Schindler und Florian Schindler	
5.1	Einleitung	75
5.2	Theoretischer Hintergrund und vorherige Forschung	76
5.3	Das Förderprojekt	77
5.4	Die Begleitforschung	79
5.5	Ergebnisse	84
5.6	Diskussion und Ausblick	85
	Literatur.	89
6	Biologie lernen mit Concept Maps – Lässt sich die Expertise im Umgang mit Concept Maps von den Augen ablesen?	91
	Sina Lenski und Jörg Großschedl	
6.1	Einleitung	91
6.2	Theoretischer Hintergrund	92
6.3	Hypothesen	95
6.4	Material und Methode	95
6.5	Ergebnisse	100
6.6	Diskussion	101
	Literatur.	104
7	Haben Kinder mit Rechenschwierigkeiten auch Schwierigkeiten im Bereich Geometrie? – Eine Eye-Tracking-Studie zur Identifikation von Dreiecken bei Schüler*innen im inklusiven Kontext	107
	Anna Lisa Simon, Benjamin Rott und Maike Schindler	
7.1	Einleitung	107
7.2	Theoretischer Hintergrund und Forschungsfrage	108
7.3	Studie	112
7.4	Ergebnisse	115

7.5	Diskussion und Ausblick	120
	Literatur.	122
8	Blickbewegungen beim grafischen Ableiten – Lassen sich Fehler durch Eye-Tracking-Daten vorhersagen und elaborieren?	125
	Kathrin Kennel, Sebastian Becker, Pascal Klein, Stefan Küchemann, Jochen Kuhn und Stefan Ruzika	
8.1	Einleitung und Forschungslage	125
8.2	Theoretischer Hintergrund	126
8.3	Forschungsfragen und Hypothesen	129
8.4	Material und Methode	129
8.5	Ergebnisse	132
8.6	Diskussion	135
	Literatur.	141
9	Eye-Tracking als Methode hypothesenprüfender Forschung in den Naturwissenschaftsdidaktiken – Neue Einblicke in die Effekte von Demonstrationsexperimenten	145
	Stefanie Lenzer und Andreas Nehring	
9.1	Einleitung	145
9.2	Planung und Durchführung einer konfirmatorischen Eye-Tracking-Studie – Von der Fragestellung bis zur Auswertung. . .	147
9.3	Ausblick: Eye-Tracking in der konfirmatorischen Modellierung von Lernangebot, -prozessen und -produkten. . .	154
	Literatur.	156
10	Fachlich vs. instruktional: Welche Aufgabenmerkmale werden bei der Schwierigkeitseinschätzung von Bruchrechenaufgaben identifiziert und evaluiert? – Eine Studie mit Eye-Tracking Stimulated Recall Interviews	159
	Saskia Schreiter, Markus Vogel, Markus Rehm und Tobias Dörfler	
10.1	Einleitung	159
10.2	Theoretischer Hintergrund	160
10.3	Forschungsfragen	164
10.4	Material und Methode	165
10.5	Ergebnisse	168
10.6	Diskussion	171
10.7	Fazit	173
	Literatur.	174
11	Blickverhalten beim Lernen und Problemlösen mit Graphen – Ein Literaturüberblick bis 2020	177
	Stefan Küchemann, Nils Cullmann, Sarah Kovac, Sebastian Becker, Pascal Klein, Kathrin Kennel, Stefan Ruzika und Jochen Kuhn	
11.1	Einleitung	178
11.2	Theorie	178

11.3	Forschungsfragen	180
11.4	Ergebnisse	181
11.5	Diskussion	188
	Literatur	190
12	Wechsel zwischen Diagramm und Formel im Kontext von Vektorfeldern – Einfluss der Aufgabenkomplexität auf Indikatoren visueller Aufmerksamkeit	193
	Larissa Hahn, Stefan Halverscheid, Jochen Kuhn und Pascal Klein	
12.1	Einleitung	194
12.2	Theoretischer Hintergrund	195
12.3	Forschungsfrage	198
12.4	Material und Methode	198
12.5	Ergebnisse	201
12.6	Diskussion	203
	Literatur	206
13	Eye-Tracking in der Mathematik- und Naturwissenschaftsdidaktik	209
	Jennifer Dröse, Philipp Neugebauer, Renate Delucchi Danhier und Barbara Mertins	
13.1	Einleitung und Forschungslage	209
13.2	Theoretischer Hintergrund	211
13.3	Forschungsfragen	215
13.4	Material und Methode	215
13.5	Ergebnisse	220
13.6	Diskussion	222
	Literatur	224
14	Blickbewegungen als Indikator für den Lernerfolg? – Nachvollziehen und Vergleichen beim Lösungsvergleich	227
	Antje Boomgaarden, Timo Leuders und Katharina Loibl	
14.1	Einleitung	227
14.2	Theoretischer Hintergrund	229
14.3	Forschungsfragen	231
14.4	Material und Methode	231
14.5	Ergebnisse	234
14.6	Diskussion	241
	Literatur	242
15	Unterschiede im Leseverhalten von Algorithmen und Programmen	245
	Philipp Kather und Jan Vahrenhold	
15.1	Einleitung und Forschungslage	245
15.2	Theoretischer Hintergrund	246

15.3	Forschungsfragen und Hypothesen	247
15.4	Material und Methodik	250
15.5	Ergebnisse	255
15.6	Diskussion	258
	Literatur.	259

Herausgeber- und Autorenverzeichnis

Über die Herausgeber

Dr. Pascal Klein ist Professor für Physik und ihre Didaktik an der Georg-August-Universität Göttingen. Er nutzt Eye-Tracking als Forschungsmethode, um das Lernen von Physik empirisch zu untersuchen. Seine Schwerpunkte liegen dabei in der Studieneingangsphase Physik, multiplen Repräsentation und digitalen Medien.

Dr. Nicole Graulich ist Professorin für Chemiedidaktik am Institut für Didaktik der Chemie der Justus-Liebig-Universität Gießen. Sie forscht zum fachlichen Lernen mit Repräsentationen in der Chemie an der Hochschule und nutzt Eye-Tracking Methoden zur qualitativen Beschreibung von Lernprozessen.

Dr. Jochen Kuhn ist Universitätsprofessor und Leiter der Arbeitsgruppe Didaktik der Physik an der TU Kaiserslautern. Seine Forschungsschwerpunkte sind Lernen mit multiplen Repräsentationen in multimedialen Lernumgebungen in Physik und prozessbasierte Analyseverfahren – wie z. B. Eye-Tracking-Analysen beim Lernen und Problemlösen.

Dr. Maike Maike ist Universitätsprofessorin an der Universität zu Köln. Ihre Forschung hat zwei zentrale Schwerpunkte: Empirisch wird das Mathematiklernen – v. a. bei Kindern und Jugendlichen mit Lernschwierigkeiten und sonderpädagogischen Förderbedarfen – untersucht, methodisch steht die Nutzung und Entwicklung neuer Technologien (z. B. Eye-Tracking) im Kontext von Mathematiklernen im Fokus.

Autorenverzeichnis

Jun.-Prof. Dr. Sebastian Becker-Genschow Digitale Bildung, Department Didaktiken der Mathematik und der Naturwissenschaften, Universität zu Köln, Köln, Deutschland

Dr. Sascha Bernholt Didaktik der Chemie, IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, Kiel, Deutschland

Antje Boomgaarden Institut für Mathematische Bildung, Pädagogische Hochschule Freiburg, Freiburg, Deutschland

Jun.-Prof. Julia Bruns Institut für Mathematik (Mathematikdidaktik), Universität Paderborn, Paderborn, Deutschland

Nils Cullmann Didaktik der Physik, Fachbereich Physik, Technische Universität Kaiserslautern, Kaiserslautern, Deutschland

Dr. Renate Delucchi Danhier IdSuL, Technische Universität Dortmund, Dortmund, Deutschland

Dr. Jennifer Dröse Fachgruppe Didaktik der Mathematik, Universität Paderborn, Paderborn, Deutschland

Prof. Dr. Tobias Dörfler Pädagogische Hochschule Heidelberg, Heidelberg, Deutschland

Julia Eckhard Institut für Didaktik der Chemie, Justus-Liebig-Universität Gießen, Gießen, Deutschland

Prof. Dr. Bernhard Ertl Erziehungswissenschaft mit dem Schwerpunkt Lernen und Lehren mit Medien, Universität der Bundeswehr München, München, Deutschland

Prof. Dr. Raimund Girwidz Lehrstuhl für Didaktik der Physik, Ludwig-Maximilians-Universität München, München, Deutschland

Prof. Dr. Nicole Graulich Institut für Didaktik der Chemie, Justus-Liebig-Universität Gießen, Gießen, Deutschland

J-Prof. Dr. Jörg Großschedl Institut für Biologiedidaktik, Universität zu Köln, Köln, Deutschland

Larissa Hahn Physik und ihre Didaktik, Fakultät für Physik, Georg-August-Universität Göttingen, Göttingen, Deutschland

Prof. Dr. Stefan Halverscheid Fachbereich Mathematik und Informatik, Georg-August-Universität Göttingen, Göttingen, Deutschland

Christoph Hoyer Lehrstuhl für Didaktik der Physik, Ludwig-Maximilians-Universität München, München, Deutschland

Maleen Hurzlmeier Erziehungswissenschaft mit dem Schwerpunkt Lernen und Lehren mit Medien, Universität der Bundeswehr München, München, Deutschland

Marit Kastaun Didaktik der Biologie, Universität Kassel, Kassel, Deutschland

Philipp Kather M.Sc. Institut für Informatik, Westfälische Wilhelms-Universität Münster, Münster, Deutschland

Kathrin Kennel Wirtschafts- und Schulmathematik, Fachbereich Mathematik, Technische Universität Kaiserslautern, Kaiserslautern, Deutschland

Prof. Dr. Pascal Klein Physik und ihre Didaktik, Fakultät für Physik, Georg-August-Universität Göttingen, Göttingen, Deutschland

Sarah Kovac Didaktik der Physik, Fachbereich Physik, Technische Universität Kaiserslautern, Kaiserslautern, Deutschland

Prof. Dr. Jochen Kuhn Lehrstuhl für Didaktik der Physik, Fakultät für Physik, Ludwig-Maximilians-Universität München, München, Deutschland

Dr. Stefan Küchemann Lehrstuhl für Didaktik der Physik, Fakultät für Physik, Ludwig-Maximilians-Universität München, München, Deutschland

Sina Lenski M.Sc. Institut für Biologiedidaktik, Universität zu Köln, Köln, Deutschland

Stefanie Lenzer Institut für Didaktik der Naturwissenschaften, Fachgebiet Didaktik der Chemie, Leibniz Universität Hannover, Hannover, Deutschland

Dr. Timo Leuders Institut für Mathematische Bildung, Pädagogische Hochschule Freiburg, Freiburg, Deutschland

Dr. Katharina Loibl Institut für Mathematische Bildung, Pädagogische Hochschule Freiburg, Freiburg, Deutschland

Dr. Monique Meier Didaktik der Biologie, Universität Kassel, Kassel, Deutschland

Prof. Dr. Barbara Mertins IdSuL, Technische Universität Dortmund, Dortmund, Deutschland

Prof. Dr. Andreas Nehring Institut für Didaktik der Naturwissenschaften, Fachgebiet Didaktik der Chemie, Leibniz Universität Hannover, Hannover, Deutschland

Philipp Neugebauer IEEM, Technische Universität Dortmund, Dortmund, Deutschland

Prof. Dr. Markus Rehm Pädagogische Hochschule Heidelberg, Heidelberg, Deutschland

Dr. Dinah Reuter Institut für Mathematische Bildung Freiburg, Pädagogische Hochschule Freiburg, Freiburg im Breisgau, Deutschland

Marc Rodemer Didaktik der Chemie, Universität Duisburg-Essen, North Rhine-Westphalia, Deutschland

Prof. Dr. Benjamin Rott Universität zu Köln, Köln, Deutschland

Prof. Dr. Stefan Ruzika Wirtschafts- und Schulmathematik, Fachbereich Mathematik, Technische Universität Kaiserslautern, Kaiserslautern, Deutschland

Florian Schindler Gesamtschule Wulfen, Dorsten, Deutschland

Prof. Dr. Maike Schindler Department Heilpädagogik und Rehabilitation, Universität zu Köln, Köln, Deutschland

Saskia Schreiter Pädagogische Hochschule Heidelberg, Heidelberg, Deutschland

Anna Lisa Simon Universität zu Köln, Köln, Deutschland

Prof. Dr. Jan Vahrenhold Institut für Informatik, Westfälische Wilhelms-Universität Münster, Münster, Deutschland

Prof. Dr. Markus Vogel Pädagogische Hochschule Heidelberg, Heidelberg, Deutschland

Dr. Bianca Watzka Didaktik der Physik, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Deutschland



Gibt es ideale Blickbewegungsmodelle zur Förderung der Lernenden beim Lösen organisch-chemischer Aufgaben?

Nicole Graulich, Marc Rodemer, Julia Eckhard und Sascha Bernholt

Inhaltsverzeichnis

1.1	Einleitung	1
1.2	Theoretischer Hintergrund	2
1.3	Forschungsfragen	5
1.4	Material und Methode	6
1.5	Ergebnisse	9
1.6	Diskussion	14
	Literatur	16

1.1 Einleitung

Das Lernen in visuell komplexen Domänen stellt Lernende häufig vor Herausforderungen. So zeigen AnfängerInnen häufig ein langsames und weniger zielorientiertes Erfassen der Informationen, während ExpertInnen schnell

N. Graulich (✉)

Institut für Didaktik der Chemie, Justus-Liebig-Universität Gießen, Gießen, Deutschland
E-Mail: Nicole.Graulich@didaktik.Chemie.uni-giessen.de

M. Rodemer

Didaktik der Chemie, Universität Duisburg-Essen, North Rhine-Westphalia, Deutschland
E-Mail: marc.rodemer@uni-due.de

J. Eckhard

Institut für Didaktik der Chemie, Justus-Liebig-Universität Gießen, Gießen, Deutschland
E-Mail: Julia.Eckhard@didaktik.Chemie.uni-giessen.de

S. Bernholt

Didaktik der Chemie, IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, Kiel, Deutschland
E-Mail: bernholt@ipn.uni-kiel.de

Entscheidungen auf Grundlage ihrer Interpretation komplexer Visualisierungen treffen und in ihre Problemlösung einbinden (vgl. Grant & Spivey, 2003; van Meeuwen et al., 2014). In der Chemie müssen beispielsweise Spektren analysiert und der passenden chemischen Struktur zugeordnet werden (Cullipher & Sevia, 2015; Topczewski et al., 2017). Bei der Interpretation dieser Repräsentationen spielen unterschiedliche Wahrnehmungs- und Interpretationsprozesse eine Rolle, um die gegebenen Informationen adäquat zu verarbeiten. Mithilfe von Eye-Tracking lässt sich untersuchen, inwiefern sich ExpertInnen und AnfängerInnen in ihrem Wahrnehmungsverhalten bei der Verarbeitung visueller Repräsentationen beim Problemlösen unterscheiden. Im Vergleich zu AnfängerInnen haben ExpertInnen im Allgemeinen kürzere Fixationsdauern und häufige Fixationen auf aufgabenrelevante Bereiche. ExpertInnen zeigen zudem längere Sakkaden und kürzere Zeiten, um relevante Informationen zu fixieren (Gegenfurtner et al., 2011). Aus diesen Beobachtungen heraus lässt sich die Annahme ableiten, dass Lernende durch die Blickbewegungen von ExpertInnen in einer Domäne geleitet und damit beim Erschließen visuell komplexer Repräsentation unterstützt werden könnten (vgl. Gegenfurtner et al., 2011, 2013; Grant & Spivey, 2003). Dabei stellt sich jedoch die Frage, ob und wie sich aus unterschiedlichen Expertisegruppen (z. B. Dozierenden, fortgeschrittenen Studierenden) ein bestimmtes Blickbewegungsmodell als ideales Modell für AnfängerInnen identifizieren lässt.

1.2 Theoretischer Hintergrund

Blickverhalten von Lernenden beim Lösen von Aufgaben in der Chemie

Lernende zeigen, je nach Vorwissen, sowohl unterschiedliche Problemlösestrategien als auch unterschiedliche Blickbewegungsmuster und unterscheiden sich letztlich auch im Erfolg beim Lösen von Aufgaben (vgl. Gegenfurtner et al., 2011, 2013; Grant & Spivey, 2003; Tang et al., 2016; Topczewski et al., 2017). In der Chemie steht die Analyse der Blickbewegungen von Lernenden, meist Studierenden, noch am Anfang. Studien zum Umgang mit chemischen Repräsentationen fokussierten vorrangig auf die Integration mehrerer Repräsentationsarten wie submikroskopisch-symbolischer oder strukturell-diagrammatischer Abbildungen. Tang und Pienta (2012) beobachteten, dass beim Lösen stöchiometrischer Aufgaben (textlicher Aufgabenstamm und Symboldarstellung) nicht erfolgreiche Lernende länger auf dem Text und der Aufgabenstellung verweilen. Die Autoren verknüpften die höhere Fixationsdauer des Aufgabenstammes mit der empfundenen kognitiven Belastung der Lernenden. Bei der Integration zwischen Spektren und der dazugehörigen chemischen Struktur sind ExpertInnen erfolgreicher und fokussieren zielgerichteter auf die richtige Antwort und auf spezifische Merkmale eines Spektrums, während die Fixationen der Lernenden fast gleichmäßig auf alle AOIs verteilt sind (Topczewski et al., 2017). Baluyut und Holme (2019) haben das Blickverhalten von Lernenden

mit unterschiedlichen Repräsentationen, z. B. Text, Symbol- und Teilchendarstellungen, untersucht und ebenfalls festgestellt, dass erfolgreiche Lernende sich von den nicht erfolgreichen Lernenden unterscheiden. Die erfolgreichen Lernenden hatten im Schnitt mehr Fixationen auf die Teilchendarstellungen und im Mittel häufigere Transitionen zwischen den AOIs. Die nicht erfolgreichen Lernenden hingegen zeigten über alle Aufgaben geringere Fixationen und Transitionen. Übergreifende Befunde im chemischen Kontext lassen sich dahingehend zusammenfassen, dass nicht erfolgreiche Lernende tendenziell eine Aufmerksamkeitsverteilung über alle AOIs hinweg zeigen, viele Wechsel zwischen AOIs vornehmen und weniger zielgerichtet fokussieren.

Modellieren von Blickbewegungen

Der Befund einer hohen Wirksamkeit von Modellbeispielen aus der Forschung zum Problemlösen (Renkl, 2014; van Gog & Rummel, 2010) wurde auch im Kontext der Verarbeitung komplexer Repräsentationen aufgegriffen. So haben beispielsweise van Gog et al. (2009) die Blickbewegungen von ExpertInnen mithilfe eines Eye-Trackers aufgenommen und diese als Hervorhebungen genutzt, z. B. als farbige Kreise oder durch Abschwächung der restlichen Abbildung, um so den Erschließungsprozess der jeweiligen Expertin oder des jeweiligen Experten zu modellieren und dadurch die Aufmerksamkeit der Lernenden zeitlich und räumlich auf relevante Merkmale der Repräsentation zu lenken. Dieser Instruktionsansatz zur Unterstützung einer authentischen und simultanen Blickführung hat sich in unterschiedlichen Domänen als lernförderlich herausgestellt (Jarodzka et al., 2012, 2013, 2017).

Die Modellierung von Handlungsabläufen durch Blickbewegungen beim Problemlösen ist jedoch nicht trivial. Der visuelle Erschließungsprozess beim Bearbeiten einer Aufgabe lässt sich über Eye-Tracking Methoden sichtbar machen, andere kognitive Prozesse sind jedoch von außen nicht beobachtbar. Zusätzlich ist es fraglich, ob Lernende nach dem Betrachten eines visuellen Modells in der Lage sind, das modellierte Blickverhalten inhaltlich zu deuten und ihr eigenes Blickverhalten dadurch zu reflektieren und bewusst an das modellierte Muster anzupassen (Kok et al., 2017).

Obwohl ExpertInnen optimierte Blickbewegungen beim Lösen von domänen-spezifischen Problemen zeigen (vgl. Reingold & Sheridan, 2011) und dies dafür spricht, dass es das „ideale“ Blickbewegungsmodell für visuell komplexe Aufgaben gibt, können ExpertInnen untereinander deutliche Unterschiede in ihrem Blickverhalten aufweisen (van Gog et al., 2005). Bei der Auswahl des geeigneten Modells zur Erstellung von Exe Movement Modeling Examples (EMMEs) wird in vielen Studien versucht, diesem Problem dahingehend zu begegnen, dass neben der durch den Berufsstand und die Arbeitstätigkeit zertifizierten Expertise als weiteres Kriterium häufig eine langjährige Lehrererfahrung in dem entsprechenden Fachbereich angeführt wird. Bei der technischen Erstellung der EMMEs werden ExpertInnen dann angewiesen, sich „didaktisch“ zu verhalten,

d. h. bezogen auf eine Zielgruppe zu erklären, welche relevanten Aspekte in der Repräsentation vorliegen, um neben den inhaltlichen Informationen auch eine zielführende Blickbewegung bei der Aufgabenbearbeitung zu vermitteln (Jarodzka et al., 2012; Richardson & Dale, 2005). Darüber hinaus gibt es Ansätze, in denen diese Vermittlung von Inhalt und Blickbewegung durch vorstrukturierte Skripte sichergestellt werden soll. Diese Skripte bilden dann das Drehbuch, auf dessen Grundlage die Blickbewegungen einer geschulten Person mithilfe eines Eye-Trackers aufgenommen werden (vgl. Krebs et al., 2019).

Hinsichtlich des Einflusses dieser verschiedenen Erstellungswege von EMMEs auf deren Wirksamkeit oder des Einflusses der Auswahlkriterien für die Blickbewegungsmodelle gibt es keine systematischen Untersuchungen. Der Rückgriff auf „ausgewiesene“ ExpertInnen unterliegt jedoch zumindest implizit der Annahme, dass es ein prototypisches und damit für eine Vermittlung relevantes Vorgehen der Aufgabenbearbeitung gibt.

Gibt es ein ideales Blickverhalten?

Im Idealfall lässt sich durch den Vergleich des Blickverhaltens des Modells und der Lernenden untersuchen, ob es für die Lösung eines domänenspezifischen Problems eine visuell „optimale“ Herangehensweise gibt und inwiefern sich diese von den Lernenden unterscheidet oder dieser ähnelt (Gegenfurtner et al., 2011). So zeigte sich im Kontext von FluglotsInnen, dass zum einen die Problemlösungen von ExpertInnen einander ähnlicher waren als die Lösungen von BerufsanfängerInnen und zum anderen die Problemlösungen von BerufsanfängerInnen wiederum einander ähnlicher waren als die Lösungen von Personen zu Beginn der Ausbildung (basierend auf Sequenzanalysen von AOI-Fixationen und sog. Levenshtein-Distanzen) (van Meeuwen et al., 2014). Mit Blick auf die Ähnlichkeit der Lösungen von ExpertInnen fügen die AutorInnen an, dass, wenn überhaupt Unterschiede auftreten, diese geringfügig sind und auf persönlichen Vorlieben basieren (van Meeuwen et al., 2014).

Demgegenüber stehen Befunde von Jarodzka et al. (2010), bei denen die ExpertInnen weniger ähnliche Problemlösestrategien verfolgten als AnfängerInnen. Die AutorInnen merken jedoch an, dass sich die Expertisegruppe sowohl aus fortgeschrittenen Promovierenden sowie ProfessorInnen zusammensetzte und diese Diversität womöglich zu einer höheren Vielfalt an Blickverhalten geführt haben könnte. Aus Sicht der AutorInnen führt dies jedoch auch zu der Schlussfolgerung, dass es sinnvoller sein könnte, die Wahrnehmungsprozesse eines sorgfältig ausgewählten Modells für die Erstellung von EMMEs zu verwenden, anstatt über mehrere ExpertInnen zu mitteln (Jarodzka et al., 2010). In ähnlicher Weise haben Litchfield et al. (2010) fünf ExpertInnen (RadiologInnen) gebeten, Lungenknoten auf Basis von Röntgenuntersuchungen zu identifizieren. Von diesen ExpertInnen wurde dann eine Person ausgewählt, die (auf Basis einer *free-response receiver operating characteristic curve*) die beste Performanz in der Aufgabenbearbeitung zeigte. Diese Person wurde dann angewiesen, während der

EMME-Aufnahme fiktiven Lernenden explizit die Herangehensweise bei der Aufgabenbearbeitung zu erläutern.

Während die Verwendung eines einzelnen (wenn auch nach bestimmten Kriterien ausgewählten) Modells zumindest das Auswahlproblem löst, stellt sich dennoch die Frage, inwieweit das entstandene EMME ein prototypisches oder ein durch individuelle Präferenzen gefärbtes Modell einer Problemlösung liefert. Während in klar strukturierten Problemstellungen (van Gog et al., 2009) eine optimale Strategie abstrakt angegeben werden kann, gibt es für Problemstellungen in visuell komplexen Domänen oftmals keine allgemeine Problemlösung. Vielmehr gibt es eine Palette unterschiedlicher Lösungen, die von suboptimal (oder falsch) bis optimal variieren können (Gronlund et al., 2005; Mumford et al., 2001). Bei AnfängerInnen ist anzunehmen, dass sie noch nicht ausreichend kognitive Schemata entwickelt haben, um komplexe Repräsentationen hinsichtlich spezifischer Problemlösungen zu klassifizieren. Bei der Bewältigung von derartigen Informationen müssen sie daher eine hohe Informationsdichte verarbeiten, was ihre Aufmerksamkeitssteuerung beeinträchtigt und zu unvollständigen oder fehlerhaften Problemlösungen führt, die situativ stark variieren können (Endsley, 1995). Umgekehrt steht ExpertInnen in der Regel ein Repertoire unterschiedlicher Schemata zur Verfügung, die bei konkreten Problemtypen zu spezifischen Problemlösestrategien führen (vgl. Reingold & Sheridan, 2011). Daher kann erwartet werden, dass diese „optimalen“ Problemlösungen personenübergreifend relativ ähnlich sind und bei ExpertInnen demnach eine geringere Varianz im Blickverhalten zu erwarten ist (van Meeuwen et al., 2014).

1.3 Forschungsfragen

Für die Auswahl eines geeigneten Modells zur Förderung von Lernenden im Umgang mit organisch-chemischen Aufgabenstellungen ist das Ziel der hier beschriebenen explorativen Studie zu untersuchen, inwiefern sich Expertisegruppen (ExpertInnen, fortgeschrittene Studierende, AnfängerInnen) beim visuellen Decodieren organisch-chemischer Repräsentationen, d. h. während des Problemlösens, unterscheiden und inwiefern sich das intuitive Blickverhalten beim mentalen Problemlösen (stilles Betrachten eines visuellen Stimulus und dessen Verarbeitung im Rahmen einer Problemstellung, jedoch ohne Erklären) von einem Blickverhalten während des Erklärens (im Sinne einer Laut-Denken-Verarbeitung desselben Stimulus im Rahmen derselben Problemstellung) unterscheidet.

Für die hier vorliegende Untersuchung haben wir Fallvergleiche als Aufgabenformat gewählt (Alfieri et al., 2013). Fallvergleichsaufgaben in der Organischen Chemie erfordern den Vergleich zweier chemischer Reaktionen anhand sowohl oberflächlich-visueller als auch implizit-chemischer Merkmale (Graulich & Schween, 2018). Ein ideales Blickbewegungsmodell könnte Lernende dabei unterstützen, derartige repräsentationsbasierte Fallvergleiche visuell zu erschließen, z. B. das Identifizieren der visuellen Ähnlichkeiten und Unterschiede.

Die explorative Studie gliedert sich in die folgenden drei Forschungsfragen:

1. Inwiefern unterscheiden sich die Blickbewegungen unterschiedlicher Expertisegruppen beim visuellen Decodieren organisch-chemischer Fallvergleiche?
2. Inwiefern sind die Blickbewegungen der unterschiedlichen Expertisegruppen konsistent beim Lösen mehrerer Fallvergleiche?
3. Inwiefern gibt es relative Ähnlichkeiten innerhalb der unterschiedlichen Expertisegruppen und über beide Bearbeitungsphasen hinweg?

Durch diese unterschiedlichen Vergleiche des Blickverhaltens von ExpertInnen, fortgeschrittenen Studierenden und AnfängerInnen lässt sich untersuchen, inwiefern sich die Blickbewegungen innerhalb und zwischen den Expertisegruppen ähneln oder voneinander abweichen, um Rückschlüsse zur Auswahl eines Modells zu ziehen.

1.4 Material und Methode

Stichprobe

An der Studie haben 46 ProbandInnen teilgenommen, davon 10 DozentInnen und ProfessorInnen (zwei weiblich) sowie 16 fortgeschrittene Studierende (neun weiblich) und 20 AnfängerInnen (13 weiblich) aus dem Bachelorstudiengang Chemie in ungefähr gleicher Verteilung von drei Universitätsstandorten. Die DozentInnen und ProfessorInnen hatten im Mittel 21 Jahre Erfahrung in der Lehre und wurden auf dieser Grundlage als ExpertInnen klassifiziert. Die Einteilung der Studierenden wurde anhand des Fachsemesters vorgenommen, wobei AnfängerInnen im 4. Semester aus der Veranstaltung Organische Chemie I und die fortgeschrittenen Studierenden im 6. Semester aus der Veranstaltung Organische Chemie II rekrutiert wurden.

Datenerhebung

Da die Daten zeitgleich an unterschiedlichen Standorten aufgenommen wurden, haben wir zwei unterschiedliche Eye-Tracker verwendet: einen mobilen Tobii Pro X3-120 (120 Hz, Accuracy bis 0.4°) mit der dazugehörigen Software Tobii Pro Studio und einem 25-Zoll-Monitor (Auflösung 1920×1080 Pixel) sowie einen SMI iView RED Eye-Tracker (120 Hz, Accuracy bis 0.4°) mit der Software iView X und einem 22-Zoll-Monitor (Auflösung 1680×1050 Pixel). Die Daten wurden mit beiden Geräten binokular aufgenommen mit einem mittleren Abstand von 65 cm zwischen Monitor und ProbandIn. Bei der weiteren Analyse wurden, um eine valide Trennung der Ereignisse zu erreichen, Fixationen auf der Basis eines I-VT-Klassifikators mit einer Geschwindigkeitsschwelle von $17^\circ/\text{s}$ detektiert.

Benachbarte Fixationen wurden zusammengeführt, wenn die Zeit zwischen den Fixationen nicht mehr als 75 ms betrug und der Winkel zwischen den Fixationen kleiner als 0.5° war. Die Ereignisdaten wurden von beiden Systemen exportiert, zu einem Datensatz zusammengeführt und gemeinsam analysiert.

Die Aufgaben wurden den ProbandInnen an den Standorten in gleicher Weise präsentiert, jeweils mit den chemischen Strukturrepräsentationen auf weißem Hintergrund. Zu Beginn der Datenerhebung wurde eine Kalibrierung durchgeführt, und bei Werten über 0.5° Abweichung wurde nachkalibriert. Vor der Bearbeitung der acht Items wurde den ProbandInnen eine Beispielaufgabe präsentiert, um sie mit dem Ablauf und der Aufgabenstellung vertraut zu machen. Danach wurden die ProbandInnen instruiert, die Aufgabe in der ersten Phase der Aufgabenbearbeitung still zu betrachten und die Reaktion zu identifizieren, die am schnellsten abläuft (Phase: Mentales Problemlösen). Wenn die ProbandInnen sich entschieden hatten, wurden sie gebeten, entweder A oder B zu benennen; in diesem Moment wurde die Zeit gestoppt. Anschließend wurde die gleiche Aufgabe abermals präsentiert mit der Bitte, die Lösung der Aufgabe zu erklären (Phase: Lautes Denken).

Aufgaben

Für die Untersuchung haben wir acht Fallvergleichsaufgaben zum Reaktionsmechanismus der Substitutionsreaktion gestaltet, die sich in verschiedenen Merkmalen unterscheiden (Abb. 1.1). Die Aufgabenstellung war zu entscheiden, welche der beiden Reaktionen schneller verläuft (in Abb. 1.1 ist dies Reaktion B) Die farbige Hervorhebung zeigt die AOIs, die für die Auswertung festgelegt

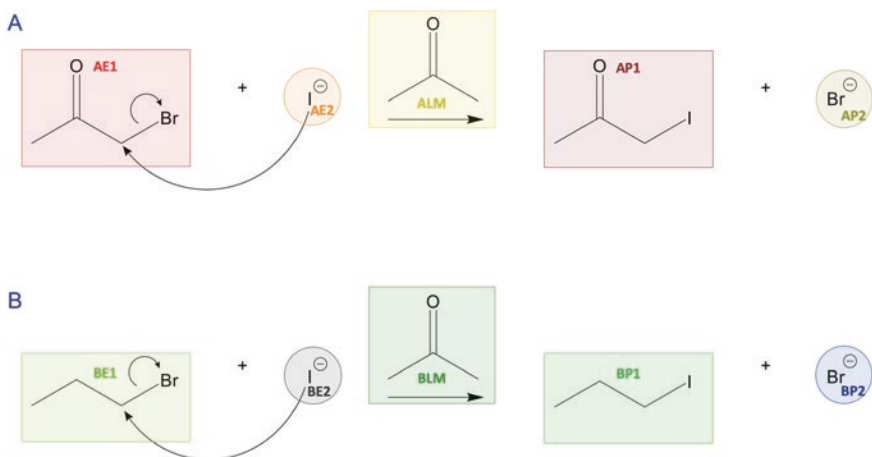


Abb. 1.1 Beispiel einer Fallvergleichsaufgabe zweier Substitutionsreaktionen mit den vorab festgelegten AOIs und deren Benennung (die AOIs wurden während der Aufgabenbearbeitung nicht dargestellt)

wurden (Abb. 1.1). Diese ergeben sich aus den beiden Reaktionen sowie aus der Unterteilung der jeweiligen Edukte, Lösungsmittel und Produkte.

Datenanalyse

Für die Datenanalyse wurde auf zwei Blickbewegungsmaße zurückgegriffen. Eine zentrale Rolle spielen dabei Fixationen, die je nach Position auf dem Bildschirm als Fixationen innerhalb spezifischer, vorab definierter AOIs (Abb. 1.1) oder als außerhalb dieser AOIs (sog. *white space*) klassifiziert wurden. Auf Basis der AOI-Klassifikation wurde die Dauer der einzelnen Fixationen pro AOI aufsummiert, um die gesamte Fixationszeit für jede AOI zu berechnen (Holmqvist et al., 2011). Diese Angaben geben die Zeit an, für die die Aufmerksamkeit auf die jeweiligen Bereiche der Abbildungen gerichtet wurde, bzw. die kumulierte Gesamtdauer aller AOI-Fixationen vom Beginn bis zum Ende der Aufgabe (vgl. Holmqvist et al., 2011). Alle Fixationszeiten werden in Sekunden angegeben. Für Gruppenvergleiche wurden die Maße auf Basis von Varianzanalysen (ANOVA) für Messwiederholungsdesigns ausgewertet.

Neben der Aufsummierung der Fixationsdauer wird auf die AOI-Sequenz als zweites Blickbewegungsmaß zurückgegriffen, d. h. auf die Reihenfolge, in der die ProbandInnen die unterschiedlichen AOIs nacheinander betrachten. Dieses Maß wird zum einen hinsichtlich der Transitionen untersucht, also der Übergänge von einer AOI zu einer anderen AOI. Dabei spielt weder die Dauer der einzelnen Fixationen noch die Anzahl der Refixationen innerhalb derselben AOI eine Rolle, sondern nur die Anzahl der Wechsel zwischen AOIs. Diese AOI- bzw. Transitionssequenz (exklusive Refixationen, also bspw. die Sequenz AE1-AE2-AE1-BE1-BE2; Abb. 1.1), wird zum anderen dafür herangezogen, die Ähnlichkeit der Blickbewegungen zweier ProbandInnen während der Aufgabenbearbeitung zu vergleichen. Dazu wird die sogenannte Levenshtein-Distanz bestimmt, die angibt, wie viele Umsortierungen, Ergänzungen und Löschungen in der AOI-Sequenz minimal vorgenommen werden müssen, um zwei Sequenzen ineinander zu überführen (bspw. sind zwei Umsortierungen und eine Ergänzung notwendig, um die obige Sequenz in die Sequenz AE1-BE1-BE2-AE2 zu überführen). Um zwei Sequenzen unterschiedlicher Länge zu vergleichen, wird die Distanz auf Basis der maximalen Sequenzlänge normalisiert. Dieses normalisierte Distanzmaß kann Werte im Intervall 0 bis 1 annehmen, wobei 1 identische Sequenzen und kleinere Werte zunehmend unähnliche Sequenzen charakterisieren (Holmqvist et al., 2011). Die Distanzwerte werden zur Erstellung einer Adjazenz- bzw. Nachbarschaftsmatrix berechnet, d. h. einer Aufstellung der normalisierten Levenshtein-Distanzen für alle paarweisen Vergleiche der AOI-Sequenzen aller ProbandInnen (gemittelt über alle Aufgaben). Diese Matrixdarstellung wird auf Basis einer Hauptkomponentenanalyse auf zwei Dimensionen reduziert, was in erster Linie der visuellen Projektion der relativen Ähnlichkeit bzw. Unähnlichkeit der Transitionssequenzen der ProbandInnen untereinander dient (Mardia, 1978). Die Dimensionen haben als Ursprung den Mittelwert über alle paarweisen

Levenshtein-Distanzen, aber keine unmittelbare inhaltliche Bedeutung (bspw. korrelieren sie weder mit Bearbeitungsdauer, Fixationsdauer noch mit dem Lösungserfolg), sondern dienen der Darstellung des „Abstands“ zwischen den ProbandInnen, sodass ProbandInnen, die sich über die Aufgaben hinweg in ihren Transitionssequenzen ähneln, einen geringen Abstand aufweisen, während ProbandInnen mit sehr unterschiedlichen Transitionssequenzen weit voneinander entfernt abgebildet werden. Entscheidend ist demnach der relative Abstand der ProbandInnen untereinander, nicht die einzelne Verortung in der zweidimensionalen Fläche.

1.5 Ergebnisse

Beim Vergleich der unterschiedlichen Gruppen zeigt sich, dass sich in vielen Fällen das Blickverhalten der ExpertInnen von dem der fortgeschrittenen Studierenden und AnfängerInnen unterscheidet. Mit Blick auf die Bearbeitungszeit sind die ExpertInnen ($M=37.7$ s, $SD=16.3$) in unserer Stichprobe signifikant schneller als die AnfängerInnen ($M=61.9$ s, $SD=34.9$) und die fortgeschrittenen Studierenden ($M=46.8$ s, $SD=39.4$), während sich die Studierenden untereinander sehr ähnlich sind und die Aufgaben ähnlich schnell lösen. Dieses Muster findet sich sowohl in der Bearbeitungsphase des mentalen Problemlösens als auch in der Phase des lauten Denkens (ExpertInnen: $M=101.0$ s, $SD=52.8$; Fortgeschrittene: $M=126.0$ s, $SD=70.4$; AnfängerInnen: $M=221.0$, $SD=113.0$). Die schnellste Bearbeitungszeit eines Experten in der Stichprobe liegt bei 6 s und zeigt exemplarisch, wie schnell hier ein Experte in der Lage ist, die Repräsentationen (Abb. 1.1) zu erfassen.

Die dokumentierte Schnelligkeit der ExpertInnen im Vergleich zu den anderen beiden Gruppen spiegelt sich auch in der geringeren Fixationsdauer der AOIs wider, wiederum sowohl beim mentalen Problemlösen ($F(2, 38)=4.21$, $p=.022$) als auch beim lauten Denken ($F(2, 38)=10.31$, $p<.001$; Abb. 1.2). Auch in diesem Fall unterscheiden sich die ExpertInnen signifikant von den fortgeschrittenen Studierenden und den AnfängerInnen. Die Fixationsdauer dieser beiden letzten Gruppen ist nur beim lauten Denken signifikant unterschiedlich.

Es stellt sich nun die Frage, ob die ExpertInnen relevante AOIs, wie Edukte und Produkte, in der gleichen Reihenfolge fixieren oder ob sich schon beim Erschließen der Aufgabenmerkmale Unterschiede zwischen den Gruppen zeigen. Betrachtet man die Transitionen zwischen den AOIs (Abb. 1.3), sind die Unterschiede zwischen den Gruppen marginal. Das Aufgabenformat scheint in allen Gruppen ähnliche Transitionsmuster zwischen den AOIs hervorzurufen. Beispielsweise zeigen alle Expertisegruppen häufige Transitionen zwischen den Edukten beider Reaktionen (AE1 und BE1) und ausgehend von diesen Edukten Transitionen zu den Produkten beider Reaktionen. Auch die relative Verteilung der Transitionen, an denen die verschiedenen AOIs beteiligt sind, scheint in allen drei Gruppen vergleichbar zu sein.

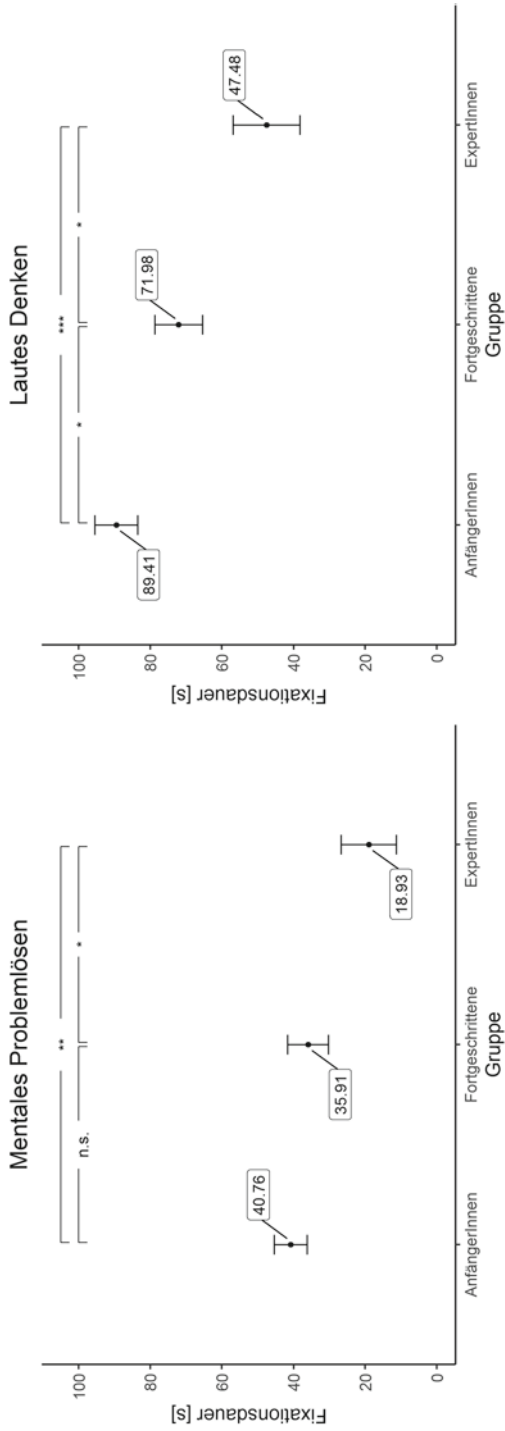


Abb. 1.2 Gesamte Fixationsdauer über alle AOIs, unterteilt nach Expertisegruppe und Bearbeitungsphasen, Angabe der jeweiligen Gruppenmittelwerte sowie Signifikanzniveau der paarweisen Gruppenvergleiche (ANOVA mit Messwiederholung und Post-hoc-Vergleiche mittels Games-Howell-Test und Bonferroni-Korrektur)

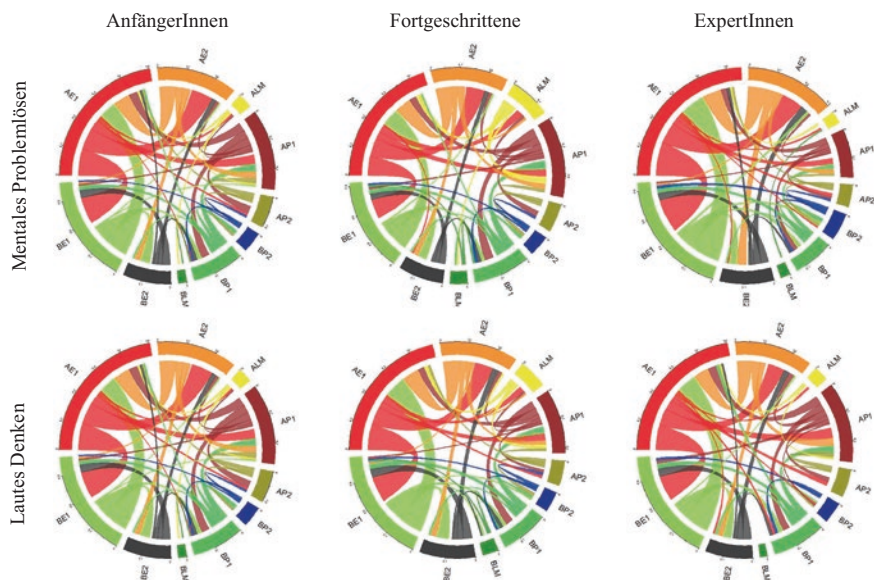


Abb. 1.3 Visualisierung der Transitionshäufigkeiten zwischen den AOIs, gemittelt über alle Aufgaben und die drei Gruppen, AnfängerInnen (links), Fortgeschrittene (Mitte) und ExpertInnen (rechts), beim mentalen Problemlösen (obere Reihe) und beim lauten Denken (untere Reihe). Transitionen zwischen AOIs sind immer in der Farbe der Ausgangs-AOI dargestellt, wobei die Dicke der Verbindungen zwischen zwei AOIs der relativen Häufigkeit der entsprechenden Transition entspricht

Während die Transitionshäufigkeit zwischen den AOIs ein gröberes Maß darstellt, lassen sich die Ähnlichkeiten zwischen den konkreten Transitionssequenzen der drei Gruppen pro Aufgabe durch die Levenshtein-Distanz ermitteln. Vergleicht man für jede Aufgabe die mittleren Levenshtein-Distanzen aller Transitionssequenzen innerhalb der drei Expertisegruppen (Intragruppenvergleich; Abb. 1.4), so zeigt sich, dass die mittleren Distanzen der AnfängerInnen und der fortgeschrittenen Studierenden über alle Aufgaben relativ konstant sind, sowohl beim mentalen Problemlösen als auch beim lauten Denken. Der Mittelwert der ExpertInnen liegt in beiden Phasen etwas höher, wonach die Transitionssequenzen der ExpertInnen im Mittel untereinander ähnlicher sind als in den beiden anderen Gruppen. Allerdings ist die Streuung der mittleren Levenshtein-Distanzen in beiden Bearbeitungsphasen der ExpertInnen deutlich größer als in den beiden anderen Expertisegruppen, sodass die relative Ähnlichkeit der Transitionssequenzen der ExpertInnen von Aufgabe zu Aufgabe stärker variiert als bei den Studierenden.

Betrachtet man die mittlere Levenshtein-Distanz aller Transitionssequenzen pro Aufgabe (Intergruppenvergleich), zeigen sich z. T. sehr unterschiedliche Verläufe über die acht Aufgaben (Abb. 1.5). So sind die Konfidenzintervalle der

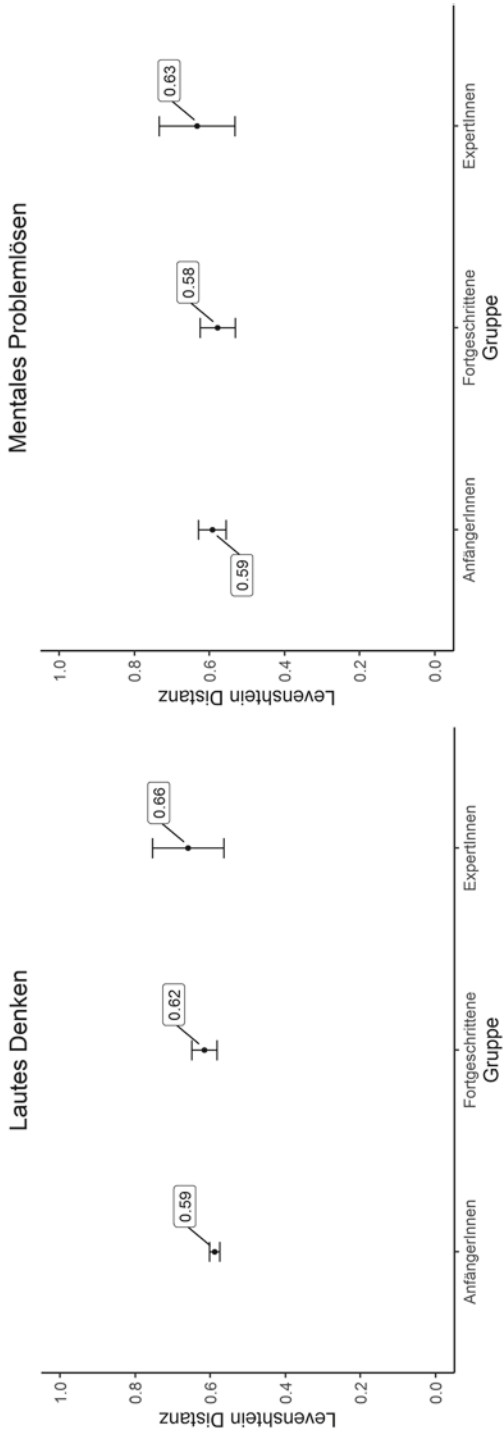


Abb. 1.4 Mittlere Levenshtein-Distanzen pro Expertisegruppe und Bearbeitungsphase. Bestimmung der Distanzwerte auf Basis paarweiser Vergleiche der Transitionsssequenzen innerhalb jeder Gruppe und Aufgabe (Intragruppenvergleich; s. unter „Datenanalyse“ in Abschn. 1.4)

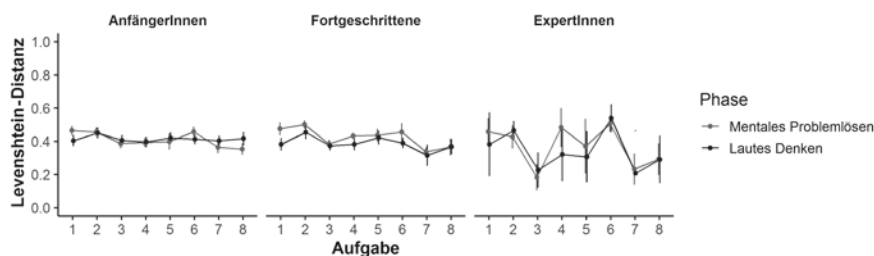


Abb. 1.5 Mittlere Levenshtein-Distanz (und 95 % CI) der Transitionssequenzen pro Aufgabe, unterteilt nach Expertisegruppen und Bearbeitungsphase. Bestimmung der Distanzwerte auf Basis paarweiser Vergleiche der Transitionssequenzen pro Aufgabe zwischen allen ProbandInnen (Intergruppenvergleich; s. unter „Datenanalyse“ in Abschn. 1.4)

Levenshtein-Distanzen bei den beiden Studierendengruppen sehr klein, und es zeigen sich fast konstante Mittelwerte der relativen Ähnlichkeit der Transitionssequenzen bei allen Aufgaben. Bei den ExpertInnen gibt es hingegen deutliche Abweichungen bei einzelnen Aufgaben, bei denen niedrige mittlere Levenshtein-Distanzen und insgesamt auch größere Konfidenzintervalle auf Transitionssequenzen hindeuten, die je nach Aufgabe mehr oder weniger stark von den Transitionssequenzen der anderen ProbandInnen abweichen. Der Verlauf ist dabei zwischen den beiden Phasen des mentalen Problemlösens und des lauten Denkens relativ ähnlich, nur bei einer Aufgabe (Aufgabe 6) zeigt sich in beiden Bearbeitungsphasen eine deutlich höhere Ähnlichkeit der Transitionssequenzen der ExpertInnen und der beiden Studierendengruppen.

Auf Basis der normalisierten Levenshtein-Distanzen für alle paarweisen Vergleiche der AOI-Sequenzen aller ProbandInnen (gemittelt über alle Aufgaben) wurde eine Adjazenz- bzw. Nachbarschaftsmatrix erstellt, die die relative Ähnlichkeit der AOI-Sequenzen aller ProbandInnen untereinander abbildet. Dabei zeigen sich mit Blick auf die ExpertInnen recht unterschiedliche Cluster, also unterschiedliche Konstellationen von ExpertInnen, deren Transitionssequenzen sich stärker ähneln, bzw. unterschiedliche „Ausreißer“ von Personen, deren Transitionssequenzen deutlich von den Sequenzen der anderen ExpertInnen abweichen. Zur besseren Visualisierung dieser relativen Ähnlichkeit der ProbandInnen wurde auf Basis der paarweisen Distanzen eine Hauptkomponentenanalyse durchgeführt, die diese Distanzen auf zwei Dimensionen reduziert (s. unter „Datenanalyse“ in Abschn. 1.4), wodurch ProbandInnen mit (über alle Aufgaben hinweg) sehr ähnlichen Transitionssequenzen einen geringen Abstand aufweisen und ProbandInnen mit sehr unterschiedlichen Transitionssequenzen weit voneinander entfernt abgebildet werden. Dadurch gruppieren sich die ExpertInnen beim mentalen Problemlösen und beim Lauten Denken in mehrere, jedoch unterschiedlich besetzte Cluster (Abb. 1.6). Einzelne ExpertInnen weichen in ihren Transitionssequenzen sowohl beim mentalen Problemlösen als auch beim lauten Denken deutlich von den weiteren ProbandInnen ab. Einige ExpertInnen sind sich untereinander, aber auch zu vielen Studierenden deutlich

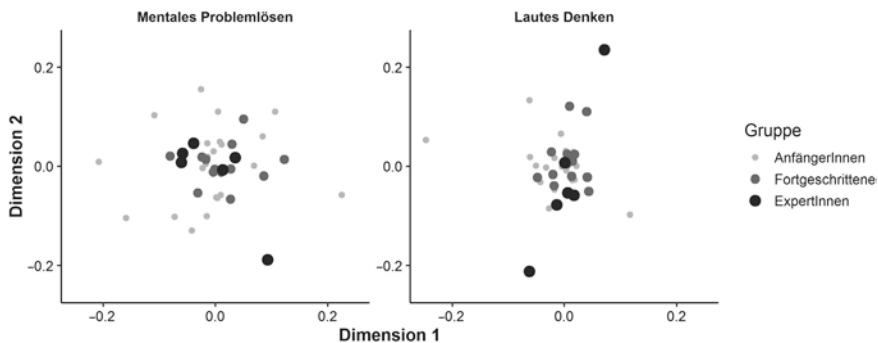


Abb. 1.6 Darstellung der relativen Ähnlichkeit der Transitionssequenzen aller ProbandInnen in beiden Bearbeitungsphasen

ähnlicher in ihren Transitionssequenzen, wobei die relative Ähnlichkeit zwischen den beiden Aufgabenbearbeitungsphasen variiert.

1.6 Diskussion

Basierend auf den Ergebnissen lässt sich festhalten, dass die drei Expertisegruppen sich nicht in allen Eye-Tracking-Maßen grundlegend unterscheiden. Die ExpertInnen unterscheiden sich etwas deutlicher von den AnfängerInnen, während die fortgeschrittenen Studierenden den AnfängerInnen ähnlicher sind. Betrachtet man die gesamte Fixationsdauer beim Lösen der Aufgaben, so ist festzustellen, dass in beiden Phasen die ExpertInnen deutlich schneller betrachten und erklären als die Studierendengruppen. Hier wären die fortgeschrittenen Studierenden in ihrem Blickverhalten, das im Mittel längere Fixationsdauern aufweist, unter Umständen geeigneter, um die AnfängerInnen als Blickbewegungsmodell durch die Aufgaben zu leiten.

Insgesamt zeigt sich, dass die ProbandInnen der Studierendengruppen ein ähnliches Blickverhalten zeigen. Dies lässt den Schluss zu, dass diese Art des Aufgabenformats ein eher einheitliches Blickverhalten erzeugt und eher wenige visuell unterschiedliche Blickbewegungen innerhalb der Studierendengruppe hervorruft. Der Vergleich der ExpertInnen untereinander zeigt jedoch auf, dass diese Gruppe in ihrem visuellen Blickverhalten weniger homogen ist als die Studierendengruppen. Hier sind die Blickbewegungsmuster sowohl im Vergleich der ExpertInnen untereinander als auch über den Verlauf der Bearbeitung der einzelnen Aufgaben heterogener. Die Hauptkomponentenanalyse zeigt ebenfalls auf, dass die relativen Ähnlichkeiten der einzelnen ProbandInnen in der Phase des mentalen Problemlösens wesentlich weiter streuen als in der Phase des lauten Denkens. Dieser Befund zeigt auf, dass sich die „Didaktisierung“, d. h. die Aufforderung zum Erklären, der ExpertInnen eignen könnte, um eine stärkere Ähnlichkeit zwischen den Studierenden und den ExpertInnen zu erzeugen. Inwiefern diese

stärkere Ähnlichkeit in der Phase des lauten Denkens mit anderen Variablen wie Löseerfolg und auch Qualität der parallel gegebenen Erklärung zusammenhängt, ist Bestandteil weiterer Untersuchungen.

Die Bildung unterschiedlicher Cluster von untereinander ähnlichen Transitionssequenzen im Rahmen der Hauptkomponentenanalyse wirft mit Blick auf die Identifizierung des idealen Modells die Frage auf, ob Lernende eher von einem ähnlichen oder einem unähnlichen Blickverhalten des Modells (im Vergleich zu dem eigenen Blickverhalten) profitieren würden bzw. wie groß letztendlich die Unähnlichkeit des modellierten Blickverhaltens sein darf, um noch lernförderlich zu sein. Die Varianz in der Ähnlichkeit der ExpertInnenlösungen wirft darüber hinaus die Frage auf, inwiefern eine ideale Kombination eines Modells und eines Lernenden auf andere Kontexte oder andere Repräsentationen innerhalb der Domäne übertragbar ist.

Eine neuere Studie weist daraufhin, dass Lernende mit geringem Vorwissen nicht per se von EMMEs profitieren, sondern nur in Fällen, in denen die *AOI entry time* ähnlich ist zum modellierten Blickverhalten (Chisari et al., 2020). Dies spricht dafür, dass eine Ähnlichkeit zwischen Lernenden und Modellierer gegeben sein muss, um die kommunizierte Information zu einem bestimmten Merkmal wahrzunehmen.

Neben der Diskussion um die Passung des Blickverhaltens schließt sich die Frage nach der kognitiven Verarbeitung der Repräsentationen an. Neben der reinen Blickführung spielt die konzeptuelle Verarbeitung ebenso eine große Rolle beim erfolgreichen Problemlösen wie das adäquate Entschlüsseln der Repräsentationen. Das explizite Strukturmerkmal der Repräsentation muss mit dem inhaltlichen Wissen verknüpft werden, was wiederum vom Vorwissen beeinflusst ist. In einer Studie haben Richardson et al. (2007) das Vorwissen von ProbandInnen vorab durch unterschiedliche Informationen variiert und dabei festgestellt, dass die Augenbewegungen in Dialogen eine größere Übereinstimmung mit dem Gegenüber zeigten, wenn sie zuvor die gleichen Informationen gehört hatten, im Vergleich zu abweichenden Informationen in der Vergleichsgruppe. Dies deutet darauf hin, dass Lernende mit geringerem Vorwissen möglicherweise nicht rechtzeitig die gleichen Merkmale wahrnehmen wie das Modell und damit die Passung zwischen Erklärung und Blickbewegungsmodell unter Umständen nicht förderlich ist. Auf der anderen Seite zeigt sich, dass mit höherem Vorwissen das Modellieren der Blickbewegung für die Lernenden redundant sein kann, da die Verknüpfung zwischen Begriff und Repräsentation für die Lernenden bereits gegeben ist (van Marlen et al., 2018). Trotz der positiven Befunde zum Einsatz von EMMEs bleiben viele Detailfragen noch offen. Besonders in visuell und konzeptuell komplexen Domänen ist der Einsatz noch unzureichend aufgeklärt.

Bezogen auf die Leitfrage zu diesem Kapitel lässt sich abschließend formulieren, dass sich anhand der herangezogenen Maße kein ideales Blickbewegungsmodell in unserer Stichprobe identifizieren lässt, das prototypisch die Blickbewegung bei der Aufgabenbearbeitung modelliert. Vielmehr zeigten sich starke Unterschiede in den Blickbewegungen, sowohl zwischen den ExpertInnen

als auch bei verschiedenen Aufgaben. Die Unterschiede zwischen den einzelnen Gruppen und insbesondere auch die vergleichsweise geringere Varianz in den Studierendengruppen unterstützen jedoch den Bedarf und auch das Potenzial in der Weiterentwicklung von visuellen Unterstützungsmöglichkeiten, die die visuelle Verarbeitung komplexer Repräsentationen unterstützen.

Literatur

- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychology*, 48(2), 87–113. <https://doi.org/10.1080/00461520.2013.775712>.
- Baluyut, J. Y., & Holme, T. A. (2019). Eye tracking student strategies for solving stoichiometry problems involving particulate nature of matter diagrams. *Chemistry Teacher International Best Practices Chemistry Education*, 1(1), 20180003. <https://doi.org/10.1515/cti-2018-0003>.
- Chisari, L. B., Mockevičiūtė, A., Ruitenburg, S. K., van Vemde, L., Kok, E. M., & van Gog, T. (2020). Effects of prior knowledge and joint attention on learning from eye movement modelling examples. *Journal of Computer Assisted Learning*, 1–11. <https://doi.org/10.1111/jcal.12428>.
- Cullipher, S., & Seavian, H. (2015). Atoms versus bonds: How students look at Spectra. *Journal of Chemical Education*, 92(12), 1996–2005. <https://doi.org/10.1021/acs.jchemed.5b00529>.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64. <https://doi.org/10.1518/001872095779049543>.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552. <https://doi.org/10.1007/s10648-011-9174-7>.
- Gegenfurtner, A., Siewiorek, A., Lehtinen, E., & Säljö, R. (2013). Assessing the quality of expertise differences in the comprehension of medical visualizations. *Vocations and Learning*, 6(1), 37–54. <https://doi.org/10.1007/s12186-012-9088-7>.
- Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, 14(5), 462–466. <https://doi.org/10.1111/1467-9280.02454>.
- Graulich, N., & Schween, M. (2018). Concept-oriented task design: Making purposeful case comparisons in organic chemistry. *Journal of Chemical Education*, 95(3), 376–383. <https://doi.org/10.1021/acs.jchemed.7b00672>.
- Gronlund, S. D., Dougherty, M. R. P., Durso, F. T., Canning, J. M., & Mills, S. H. (2005). Planning in air traffic control: Impact of problem type. *The International Journal of Aviation Psychology*, 15(3), 269–293. https://doi.org/10.1207/s15327108ijap1503_4.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154. <https://doi.org/10.1016/j.learninstruc.2009.02.019>.
- Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, 40(5), 813–827. <https://doi.org/10.1007/s11251-012-9218-5>.

- Jarodzka, H., van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70. <https://doi.org/10.1016/j.learninstruc.2012.11.004>.
- Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal Eye Movement Research*, 10(1), 1–18. <https://doi.org/10.16910/jemr.10.1.3>.
- Kok, E. M., Aizenman, A. M., Vö, M. L. H., & Wolfe, J. M. (2017). Even if I showed you where you looked, remembering where you just looked is hard. *Journal of Vision*, 17(12), 1–11. <https://doi.org/10.1167/17.12.2>.
- Krebs, M.-C., Schüler, A., & Scheiter, K. (2019). Just follow my eyes: The influence of model-observer similarity on eye movement modeling examples. *Learning and Instruction*, 61, 126–137. <https://doi.org/10.1016/j.learninstruc.2018.10.005>.
- Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology Applied*, 16(3), 251–262. <https://doi.org/10.1037/a0020082>.
- Mardia, K. V. (1978). Some properties of classical multi-dimensional scaling. *Communications in Statistics – Theory and Methods*, 7(13), 1233–1241. <https://doi.org/10.1080/03610927808827707>.
- Mumford, M. D., Schultz, R. A., & van Doorn, J. R. (2001). Performance in planning: Processes, requirements, and errors. *Review of General Psychology*, 5(3), 213–240. <https://doi.org/10.1037/1089-2680.5.3.213>.
- Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. *The Oxford handbook of eye movements* (S. 523–550). Oxford University Press.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38(1), 1–37. <https://doi.org/10.1111/cogs.12086>.
- Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29(6), 1045–1060. https://doi.org/10.1207/s15516709cog0000_29.
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, 18(5), 407–413. <https://doi.org/10.1111/j.1467-9280.2007.01914.x>.
- Tang, H., & Pienta, N. (2012). Eye-tracking study of complexity in gas law problems. *Journal of Chemical Education*, 89(8), 988–994. <https://doi.org/10.1021/ed200644k>.
- Tang, H., Day, E., Kendhammer, L., Moore, J., Brown, S., & Pienta, N. (2016). Eye movement patterns in solving science ordering problems. *Journal Eye Movement Research*, 9(3), 1–13. <https://doi.org/10.16910/jemr.9.3.6>.
- Topczewski, J. J., Topczewski, A. M., Tang, H., Kendhammer, L. K., & Pienta, N. J. (2017). NMR spectra through the eyes of a student: Eye tracking applied to NMR items. *Journal of Chemical Education*, 94(1), 29–37. <https://doi.org/10.1021/acs.jchemed.6b00528>.
- van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22(2), 155–174. <https://doi.org/10.1007/s10648-010-9134-7>.
- van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, 19(2), 205–221. <https://doi.org/10.1002/acp.1112>.
- van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25(3), 785–791. <https://doi.org/10.1016/j.chb.2009.02.007>.

- van Marlen, T., van Wermeskerken, M., Jarodzka, H., & van Gog, T. (2018). Effectiveness of eye movement modeling examples in problem solving: The role of verbal ambiguity and prior knowledge. *Learning and Instruction*, 58, 274–283. <https://doi.org/10.1016/j.learninstruc.2018.07.005>.
- van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., de Bock, J. J. P. R., & van Merriënboer, J. J. G. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, 32, 10–21. <https://doi.org/10.1016/j.learninstruc.2014.01.004>.

Eine qualitative Analyse von Blickdaten bei statischen und dynamischen Repräsentationen im naturwissenschaftlichen Erkenntnisprozess

Marit Kastaun und Monique Meier

Inhaltsverzeichnis

2.1	Einleitung	19
2.2	Theoretischer Hintergrund	20
2.3	Zielsetzung und Forschungsfragen	22
2.4	Material und Methode	22
2.5	Ergebnisse	27
2.6	Diskussion	33
	Literatur	36

2.1 Einleitung

Eine naturwissenschaftliche Fragestellung aufstellen, Variablen operationalisieren oder eine Fehleranalyse durchführen, können Lernende entlang des naturwissenschaftlichen Erkenntnisprozesses Schwierigkeiten bereiten (z. B. Meier, 2016). Lernunterstützungen hingegen können die fachmethodische und fachliche Komplexität im forschenden Lernprozess minimieren und die individuelle Wissensanwendung fördern (z. B. Arnold et al., 2014). Besonders effektiv sind jene Lernmaterialien/-unterstützungen, die hinsichtlich ihrer Ausgestaltung an das Niveau/an die Merkmale der Lernenden, wie bspw. das Vorwissen oder motivationale Einstellungen, angepasst sind (z. B. Schnotz & Rasch, 2005).

M. Kastaun (✉) · M. Meier

Didaktik der Biologie, Universität Kassel, Kassel, Deutschland

E-Mail: m.kastaun@uni-kassel.de

M. Meier

E-Mail: mmeier@uni-kassel.de

Schülerspezifische Präferenzen im Umgang mit unterschiedlichen Darstellungs- und Gestaltungsmodi des Lernmaterials (z. B. Bild-Text, Video), einhergehend mit der Ausprägung kognitiver Merkmale, wurden bislang wenig als Differenzierungsmöglichkeit in Betracht gezogen. Daher verfolgt das Projekt das Ziel, eine Individualisierung des experimentellen Lernprozesses mittels unterschiedlicher Repräsentationskombinationen und damit verbundener Gestaltungsmodi von Lernunterstützungen im naturwissenschaftlichen Erkenntnisprozess zu überprüfen. Einen zentralen Untersuchungsbereich bildet hierbei die Analyse des schüler-spezifischen Blickverhaltens während der Nutzung der Lernunterstützungen, um mögliche Strategien im Umgang mit den Repräsentationskombinationen zu identifizieren und diese mit den individuellen Lernermerkmalen sowie der Lernleistung beim Experimentieren in Beziehung setzen zu können.

2.2 Theoretischer Hintergrund

Die unterrichtliche Umsetzung forschenden Lernens bspw. in der Planung und Durchführung eines Experiments setzt fachmethodisches und fachliches Wissen voraus (Aepkers & Liebig, 2002). Obwohl dem forschenden Lernen als kognitiver Problemlöseprozess ein hohes Potenzial zum Erwerb naturwissenschaftlicher Kompetenzen zugesprochen wird, kann die Komplexität jedoch auch zu schüler*innenspezifischen Schwierigkeiten (z. B. Hammann, 2004) führen und sich negativ auf den individuellen Lernweg auswirken. Sowohl die methodischen Anforderungen wie bspw. die Umsetzung der Variablenkontrollstrategie (Baur, 2018; Meier, 2016) als auch der konstruktivistische Ansatz des forschenden Lernens können in einer kognitiven Überlastung des Arbeitsgedächtnisses münden und somit den langfristigen Wissenserwerb minimieren (Kirschner et al., 2006; Furtak et al., 2009). Besonders der Planungsprozess verlangt von den Lernenden eine hohe Vernetzung der deklarativen und prozeduralen Wissensfacetten, da hier das gesamte experimentelle Vorgehen bereits kognitiv durchlaufen werden muss. In einem ständigen Zusammenspiel zum fachlichen Kontext müssen die Variablen identifiziert, operationalisiert, konstant gehalten oder variiert werden. Darüber hinaus beinhaltet der Planungsprozess die gezielte Materialauswahl sowie das Aufstellen eines Messkonzepts (z. B. Arnold et al., 2017; Klahr & Dunbar, 1988; Lederman et al., 2014).

Lernunterstützungen in Form von *hard* oder *soft scaffolds* (Wu & Pedersen, 2011; Brush & Saye, 2002; Puntambekar & Hubscher, 2005), die entweder auf Basis typischer Schüler*innenhürden entwickelt wurden (*hard scaffolds*: z. B. Prompts oder Lösungsbeispiele) oder individuelle Schwierigkeiten adaptiv aufgreifen (*soft scaffolds*: z. B. Feedback), können Prozesse der Encodierung anregen und negative Lernbelastungen verringern (Arnold et al., 2014; Kaiser et al., 2018; Wichmann & Leutner, 2009). Der generelle Effekt der kognitiven Entlastung ist aber nicht eindeutig beschrieben. Er ist von der Angebotsgestaltung, der Integration der Lernunterstützungen in die Lern Gelegenheit sowie deren Nutzung und Wahrnehmung durch den Lernenden abhängig. Diese

zusammengefasste aktive Lernzeit wird beeinflusst durch die kognitiven (z. B. Repräsentationspräferenz, kognitiver Stil oder Vorwissen) und motivationalen Merkmale der Lernenden (Pashler et al., 2009; Höffler & Leutner, 2011; Koć-Januchta et al., 2019). Die Bereitstellung differenzierter Lernunterstützungen bspw. über die Variation der Repräsentationen stellt daher eine Möglichkeit dar, individualisierte *hard scaffolds* in den Erkenntnisprozess zu integrieren und somit die schüler*innenspezifischen Lernprozesse zu optimieren. Digitale Techniken eröffnen die Chance, vielfältige Repräsentationsformen zu konstruieren und einzusetzen (Irion & Scheiter, 2018). In digital gestützten Lernunterstützungen können multiple externe Repräsentationen nicht mehr nur aus analogen deskriptionalen und depiktionalen Repräsentationen (Ainsworth, 2006), sondern auch aus verbal-sprachlichen Audiospuren oder symbolischen Texten in Verbindung mit statischen und dynamischen Bildern (Nitz et al., 2014; Weidemann, 2002) zusammengesetzt werden. Ob analog oder digital konstruierte Repräsentationskombinationen, für eine zielgerichtete und effektive Nutzung sollte auf kognitionspsychologische Gestaltungsprinzipien (Mayer, 2014; Sweller, 2005) zurückgegriffen werden, um die materialbezogenen Belastungen (*extraneous loads*; Sweller, 2005) innerhalb des Lernprozesses zu minimieren und freie Kapazitäten im Arbeitsgedächtnis zum Aufbau von Schemata und damit zum Lernen zu schaffen (Schmidt-Weigand et al., 2008). Folgt man diesen empirisch geprüften Gestaltungshinweisen, so werden Informationen effektiver verarbeitet und miteinander vernetzt, wenn sie den Lernenden u. a. multimedial und in Beziehung zueinander dargeboten werden (z. B. Multimedia-Prinzip, Kohärenzprinzip: Mayer, 2001; Split-Attention-Effekt: Sweller, 1999).

Wie die einzelnen Informationsträger, z. B. eine Bild-Text-Kombination, von den Lernenden genutzt und ob unterschiedliche Strategien zur Erschließung von multimodalen oder monomodalen Repräsentationskombinationen angewendet werden, wurde noch nicht repräsentationsübergreifend für einen inhaltlichen Kontext beschrieben (Lesestrategien: z. B. Philipp, 2015). Eine Möglichkeit, das Blickverhalten im Umgang mit Repräsentationskombinationen sichtbar zu machen, stellt der gezielte Einsatz von Eye-Tracking-Systemen dar (Holmqvist et al., 2011). Über die Indikatoren Fixationen und Sakkaden lässt sich das Ausmaß der visuellen (offenen) Aufmerksamkeit messbar machen, das selbst als ein Indikator der bewussten Wahrnehmung und somit der weiteren Informationsverarbeitung angesehen werden kann (Ansorge & Leder, 2011; Krejtz et al., 2016). Dies unterliegt der Annahme, dass durch das Fixieren einzelner visueller Elemente die räumliche Auflösung des Sehens erhöht wird und somit kognitive Ressourcen zur Reiz-/ Informationsverarbeitung zur Verfügung stehen (z. B. Luck & Ford, 1998). Im Zusammenhang mit Repräsentationen werden Eye-Tracking-Systeme vor allem für das Erfassen von Informationsverarbeitungsprozessen eingesetzt und die gesammelten Blickdaten quantitativ über unterschiedliche Metriken ausgewertet (z. B. Park et al., 2015; Krebs et al., 2019; Brückner et al., 2020; Richter et al., 2019). Besonders aus der Leseforschung ist bekannt, dass vorrangig die Fixationen und Sakkaden in Bezug auf die textgebundene, visuelle Aufmerksamkeit untersucht werden (z. B. Smyrnakis et al., 2017; Solheim & Uppstad,

2011). Hier zeigt sich u. a., dass schlechtere Leser*innen eine höhere Fixationsanzahl und Fixationsdauer gegenüber geübten Leser*innen aufzeigen (Kim & Lombardino, 2016; Nilsson Benfatto et al., 2016). Inwieweit eine qualitative Beschreibung des Blickverhaltens zu ergänzenden Aussagen in der Nutzung von Lernunterstützungen über die Identifikation von Mustern im Umgang mit den Repräsentationskombinationen führt, ist noch unklar.

2.3 Zielsetzung und Forschungsfragen

Die vorliegende Studie verfolgt daher das Ziel, das repräsentationsspezifische Blickverhalten während der Nutzung multimedialer Lernunterstützungen zu analysieren, um mögliche Muster und Strategien im Umgang mit statischen und dynamischen Repräsentationskombinationen und der damit einhergehenden Informationsverarbeitung zu identifizieren und zu beschreiben. Diese Muster im Blickverhalten können zukünftig genutzt werden, um Zusammenhänge zwischen der Ausprägung unterschiedlicher kognitiver Merkmale von Lernenden und dem Umgang mit unterschiedlichen Repräsentationskombinationen in Form von Lernunterstützungen zu untersuchen. Mit Blick auf das langfristige Forschungsziel, mögliche (Daten-)Triangulationen vornehmen zu können, soll zunächst entlang der folgenden zwei Forschungsfragen untersucht werden, ob Blickdaten aus methodischer Perspektive qualitativ erfasst und beschrieben werden können:

1. Inwiefern lässt sich das schüler*innenspezifische Blickverhalten von statischen und dynamischen Repräsentationskombinationen in Form von Lernunterstützungen beim Experimentieren (Planungsphase) mittels induktiver Kategorien zuverlässig qualitativ beschreiben?
2. Inwieweit lassen sich Muster im Blickverhalten unter Einbezug quantifizierender Eye-Tracking-Maße über eine qualitative Analyse von statischen und dynamischen Repräsentationskombinationen beschreiben und kategorisieren?

2.4 Material und Methode

Design und Studienablauf

Zur Untersuchung der Forschungsfragen wurde eine Interventionsstudie durchgeführt (für eine ausführliche Beschreibung des Studiendesigns s. Kastaun et al., 2018). In einem dreistündigen Experimentiermodul in der Experimentier-Werkstatt Biologie FLOX, einem Lehr-Lern-Labor der Universität Kassel, untersuchten Schüler*innen des 9. und 11. Jahrgangs ($N=73$; ♀=66 %) den Einfluss der Temperatur auf die Aktivität von Hefezellen. Entlang des naturwissenschaftlichen Erkenntnisprozesses erarbeiteten sich die Lernenden in Kleingruppen das zuvor präsentierte Phänomen. Zu Beginn der Planungsphase wurden die Schüler*innen

angehalten, eigenständig in Einzelarbeit eine Lernunterstützung auszuwählen, diese zu betrachten und den Inhalt zu erfassen, um im Anschluss bspw. Variablen zu operationalisieren. Nach der Einzelintervention diskutierten die Lernenden ihre phasenbezogenen Ergebnisse und legten ihr weiteres gemeinsames, experimentelles Vorgehen fest.

Lernunterstützungen/Material

Die Lernunterstützungen wurden den Lernenden in vier unterschiedlichen Repräsentationskombinationen zur Verfügung gestellt. Die auf Basis kognitionspsychologischer Prinzipien (z. B. Mayer, 2014; Sweller, 2005) konstruierten Lernunterstützungen sind sowohl designspezifisch als auch inhaltlich identisch (Meier & Kastaun, 2021). Ohne fachlichen Bezug fassen sie das fachmethodische Wissen der jeweiligen Phase des Erkenntnisprozesses zusammen und stellen somit eine fach- und experimentübergreifende Unterstützung dar. Die Lernunterstützungen unterscheiden sich jedoch hinsichtlich ihrer Kombinationen an Einzelrepräsentationen. Die Bild-Text- sowie die Bild-Audio-Kombination umfassen die statischen Lernunterstützungen, die Animation, bewegtes Bild und Text, und das Video, bewegtes Bild und Audio, können als dynamische Repräsentationskombinationen zusammengefasst werden.

Methode

Zur Erfassung der Blickdaten wurden die Schüler*innen nach freier Auswahl einer Lernunterstützung gebeten, sich diese an einem Eye-Tracking-System, bestehend aus Bildschirm, Laptop und Eye-Tracker anzusehen (Kastaun et al., 2020). Zur Aufzeichnung wurde der monitorbasierte Eye-Tracker Tobii-Pro X3-120, verbunden mit einem 22-Zoll-Bildschirm (1920 × 1080), verwendet. Die Augenbewegungen (binokular) werden dabei in einer Frequenz von 120-mal pro Sekunde (120 Hz) durch die Reflexion des Infrarotlichtes im Auge aufgezeichnet. Das verwendete System hat zudem eine hohe Genauigkeit ($<0.4^\circ$) und erlaubt es den Probanden, sich während der Aufnahme in einem Abstand von etwa 60–80 cm frei vor dem Bildschirm zu bewegen. Zur Durchführung und Auswertung der Eye-Tracking-Daten wurde die Software Tobii Studio in der Version 3.4.8 eingesetzt (Tobii AB, 2016). Vor jeder Aufzeichnung der Blickdaten erfolgte eine individuelle 9-Punkt-Kalibrierung.

Auswertung

Entsprechend der dieser Studie zugrunde liegenden qualitativen Forschungsausrichtung fiel die Entscheidung in der Auswertungsmethodik gegen die Anwendung

einer rein quantitativen Analyse der standardisierten Eye-Tracking-Metriken (Holmqvist et al., 2011). Über diese werden meist Selektions-, Organisations- und Integrationsprozesse von repräsentationsbedingten Informationen erfasst und analysiert (z. B. Scheiter et al., 2018; Takacs & Bus, 2016). Da das Interventionsmaterial, vor allem aber die monomedialen Lernunterstützungen, eine Fülle an textlichen Informationen aufweist, wurde eine Auswertung der Sakkaden, deren Längen und Richtungen, als weniger aussagekräftig eingestuft. Ausschließlich die absoluten Fixationsanzahlen innerhalb der Repräsentationskombinationen sowie die qualitativ gemessene Fixationsdauer wurden mit in die Datenauswertung einbezogen.

Im Rahmen dieser Studie wurde eine inhaltsanalytische, studienangepasste Auswertungsmethodik entwickelt und angewendet. Die Zusammenführung von Eye-Tracking als Messinstrument mit der Inhaltsanalyse als Auswertungsmethode repräsentiert den explorativen Charakter dieser Studie. Auf Basis der qualitativen Inhaltsanalyse nach Mayring (2016) erfolgte eine systematisierte Bearbeitung der Blickdaten, welche mittels induktiver Kategorienbildung beschrieben und analysiert wurden. Diese methodische Vorgehensweise, so Mayring (2016, S. 86), „strebt nach einer möglichst naturalistischen, gegenstandsnahen Abbildung des Materials ohne Verzerrungen durch Vorannahmen des Forschers“ und ermöglicht im Anschluss eine Quantifizierung der fallbasierten, übergeordneten Oberkategorien (Kuckartz, 2018; Mayring, 2016). Des Weiteren wurden methodische Elemente der quantitativen Inhaltsanalyse (Lamneck & Krell, 2016; Neuen-dorf, 2017) genutzt, die eine „Forschungstechnik zur Erstellung replizierbarer und valider Inferenzen von Texten oder anderen aussagekräftigen Botschaften auf den Kontext ihres Gebrauchs“ (Hutter, 2019, S. 3) darstellen. Mittels der Frequenzanalyse, einer Form quantitativer Inhaltsanalyse, werden ausgewählte Materialelemente, wie in der vorliegenden Studie Blickpunkte, ausgezählt und als Indikatoren für die individuelle, visuelle Aufmerksamkeit analysiert (Lamneck & Krell, 2016).

Der Auswertung vorangestellt fand eine Aufbereitung des blickdatenbasierten Materials in Form von Transkripten statt (Abb. 2.1). Zu diesem Zweck wurden zunächst die gesammelten, individuellen Blickdaten mittels der *gaze plots*-Replays und der prozentualen Kalibrierungsgenauigkeit gesichtet (0). Während dieser ersten Materialsichtung wurden drei Fälle der Bild-Text-Kombination und zwei der Animation aufgrund schlechter Kalibrierung ausgeschlossen. Auf Basis verbindlicher Transkriptionsregeln (Codebuch) wurde der Blickweg jedes Probanden verschriftlicht (I). Bei der Transkription handelt es sich „um einen theoriegeleiteten selektiven Prozess“ (Dresing & Pehl, 2020, S. 7), der im Zuge einer regelbasierten Verschriftlichung von Blickdaten bereits zur ersten Kategorisierung im Datenmaterial führt. Ausschlaggebend hierbei ist das Format der Daten, die nicht, wie in der qualitativen (Sozial-)Forschung üblich, auf Gesprächen oder Handlungsbeschreibungen beruhen (z. B. Regelsystem für qualitative Forschung von Dresing & Pehl, 2015). Die Erstellung der Transkripte (I) der monomedialen Lernunterstützungen (Animation und Bild-Text-Kombination) erfolgte mithilfe der *gaze plot*-Replays, die den Blickverlauf in seiner zeitlichen Reihenfolge und der

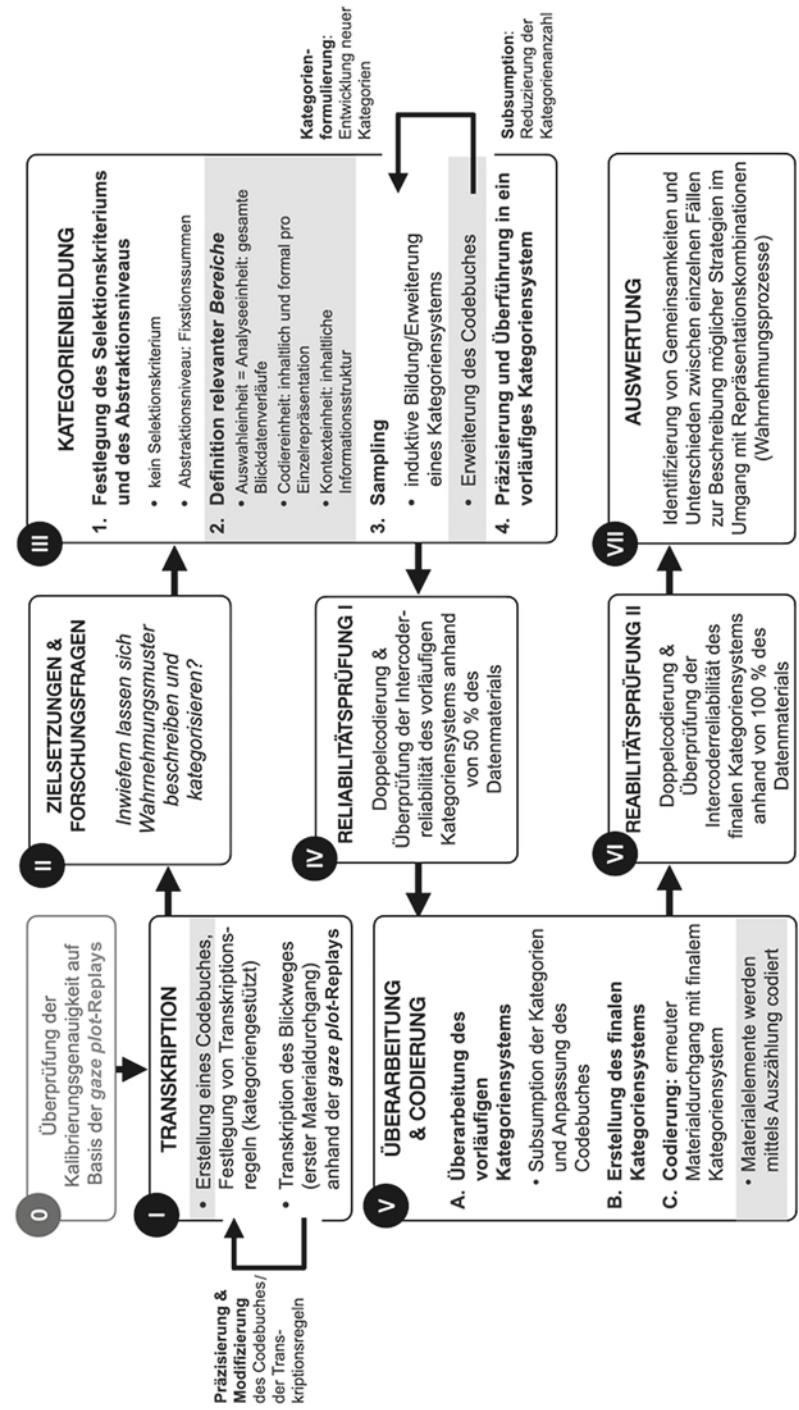


Abb. 2.1 Studienangepasstes Prozessmodell quantitativer (grau hinterlegt) und qualitativer Inhaltsanalyse von Blickdaten. (Angelehnt an Hutter, 2019; Lamneck & Krell, 2016; Mayring, 2016; Neundorff, 2017)

Fixationsdauer wiedergeben. Dabei wurden alle Fixationspunkte bis zum Ende der Aufzeichnung transkribiert. Die Transkription der Bild-Text-Kombination erfolgte timebasiert und satzweise. Die Blickverläufe in der Animation wurden hingegen eventbasiert, also pro inhaltliches Event, bestehend aus dem Erscheinen des Textes und ggf. der dazugehörigen Visualisierung, vollzogen.

Im Anschluss der Transkription erfolgten eine inhaltlich-strukturierende, qualitative Inhaltsanalyse und eine Erweiterung des Codebuches mit dem dazugehörigen Sampling, Training und Codierung (Frequenzanalyse) (Abb. 2.1). Da der inhaltsanalytische Umgang mit Blickdaten ein exploratives Vorgehen darstellt, wurde das Datenmaterial in Form der transkribierten Blickverläufe in Gänze (Auswahleinheit) analysiert (kein Selektionskriterium, Auswahleinheit = Analyseinheit). Die Codiereinheiten ergeben sich aus den Transkriptionselementen, die die jeweiligen Einzelrepräsentationen der Repräsentationskombination darstellen. So wurden das statische bzw. bewegte Bild und der gesamte Text bzw. die Textelemente zum einen nach formalen Kategorien, wie bspw. der Anzahl der Bild-Text-Verknüpfungen (Transitionen), und zum anderen nach inhaltlichen Kategorien, wie bspw. der Anzahl der Wortfixierungen auf Fachbegriffen, ausgewertet. Zu diesen wurde ein mittelhohes Abstraktionsniveau repräsentations-spezifisch festgelegt, d. h., es wurden nicht die Einzelfixationspunkte, sondern die Fixationssummen pro Einzelrepräsentation pro zeitliches oder eventbasiertes Ereignis betrachtet. Als Kontexteinheiten wurden die inhaltlichen Abschnitte der textlichen Informationen definiert. Demnach wurden je nach Repräsentationskombination entweder mehrere Satzabschnitte oder Events den übergeordneten inhaltlichen Kontexten, der Einführung, der Variablenoperationalisierung, dem Messkonzept und der Zusammenfassung zugeordnet, die als Hilfskonstrukte in die Inhaltsanalyse mit einbezogen wurden (Rössler, 2017). Im Anschluss daran erfolgte das Sampling (III 3.), indem anhand der Transkripte induktiv Kategorien gebildet wurden. Anhand der in den gewonnenen Hauptkategorien vorstrukturierten Blickdateninformationen wurden am Material Subkategorien in zwei weiteren kumulativen Materialdurchgängen mit dem kompletten Datensatz über Kategorienformulierung, Subsumption oder neuer Kategorienbildung sowie weitere Subsumption und Reduktion der Kategorienanzahl gebildet (III 4.). Nach der ersten Kategorienbildung wurden 50 % des Materials doppelt codiert und die Intercooderreliabilität ermittelt (IV). Mit dem überarbeiteten Kategoriensystem erfolgte ein finaler Materialdurchgang (V). Im letzten Schritt, der Kategoriensystementwicklung, wurde das gesamte Datenmaterial (100 %) von zwei Ratern doppelt codiert, um die Reliabilität zu prüfen, mittels interpersoneller, diskursiver Konsensbildung hinsichtlich der Kategorienbeschreibungen zu verbessern und das Kategoriensystem zu finalisieren (VI). Nach Abschluss der Kategorienbildung und Fallzuordnung wurde das Blickverhalten fallspezifisch beschrieben und es wurden die Häufigkeiten der Kategorien aufgeführt, um mögliche fallbezogene Gemeinsamkeiten und Unterschiede hinsichtlich der individuellen Subkategorien zu analysieren (VII) (Mayring, 2016).

2.5 Ergebnisse

Ergebnisse zu Forschungsfrage 1

Im Vordergrund der Entwicklung und Anwendung von induktiv gewonnenen Kategorien steht die inhaltsgetreue Beschreibung des Blickverhaltens von statischen und dynamischen Repräsentationen als Unterstützungsquelle beim Experimentieren. Auf Basis der Codiereinheiten können sowohl formale als auch inhaltliche text- und bildbezogene Kategorien aus dem Material abgeleitet und beschrieben werden, die übergreifend für beide Repräsentationskombinationen (Bild-Text und Animation) gültig sind (Tab. 2.1). Die kategoriale Zuordnung der Fälle ($n_{\text{Bild-Text}} = 21$; $n_{\text{Animation}} = 13$) zeigt eine sehr gute Zuverlässigkeit des zugrunde gelegten Systems (Bild-Text: $0.89 \leq \kappa \leq 1$, $p = .000$; Animation: $0.91 \leq \kappa \leq 1$, $p = .000$) (Bewertung nach Landis & Koch, 1977; Gwet, 2012).

Insgesamt konnten 14 repräsentationsübergreifende Kategorien identifiziert und über 39 Subkategorien näher beschrieben werden (Tab. 2.1). Darüber hinaus konnten weitere inhalts- und formalbezogene Kategorien spezifisch für jede Repräsentationskombination identifiziert werden, die sich durch die Eigenschaften der jeweiligen Einzelrepräsentation ergeben.

Weitere Kategorien zur Bild-Text-Kombination: Da die Bild-Text-Kombination eigenständig und selbstreguliert durch den Lernenden genutzt werden konnte, können die Kategorien in zwei zeitliche Abschnitte eingeteilt werden – den ersten und zweiten Leseprozess. Es werden sieben weitere Kategorien innerhalb der Bild-Text-Kombination identifiziert. Dazu zählen für den zweiten Leseprozess sechs formale und eine inhaltliche Kategorie in der Codiereinheit Text sowie die in Tab. 2.1 aufgeführten bildbasierten Kategorien zum zweiten Lesen (Tab. 2.2).

Weitere Kategorien zur Animation: Die Animation konnte im Gegensatz zur Bild-Text-Kombination nur bedingt selbstreguliert genutzt werden, da die Informationsdarstellung in ihrer zeitlichen Abfolge festgeschrieben ist. Durch das Erscheinen und Verschwinden einzelner Textelemente sowie durch das bewegte, aufbauende Bild wird die visuelle Aufmerksamkeit der Lernenden auf beide Einzelrepräsentationen gesteuert. Da durchschnittlich wesentlich mehr *Bild-Text-Verknüpfungen* (Tansitionen) und *Bilderschließungen* vorliegen, werden hier die Fixationsanzahlen für jede Kategorie hinzugezogen. In der Folge werden zwei weitere Kategorien – *Fixationsanzahl Bild-Text-Verknüpfung* ($\kappa = 1$, $p = .000$) und *Fixationsanzahl Bilderschließung* ($\kappa = 1$, $p = .000$) induktiv aus dem Material gebildet.

Ergebnisse zu Forschungsfrage 2

Nach der kategorienbasierenden Beschreibung des Blickverhaltens erfolgt für jede monomodale Repräsentationskombination die Auswertung formatgestützter Muster im Blickverhalten mittels einer Fallzuordnung. Dazu werden zunächst

Tab. 2.1 Repräsentationsübergreifende formale und inhaltliche Kategorien und deren kategorienbezogene Beschreibung. SK = Subkategorie

Codiereinheit		Oberkategorie	Codieranweisung	SK
Text	Formal	<i>Formale Lesewiederholungen</i>	Ein Satz wird als wiederholt gewertet, wenn mindestens 1/3 der Wörter des jeweiligen Satzes erneut gelesen wurden.	3
		Anzahl der Satzwiederholungen über den gesamten Leseprozess		
		<i>Formale Lesegenauigkeit</i>	Es werden satzweise die Fixationsanzahlen ausgezählt und mit der absoluten Wortanzahl pro Satz ins Verhältnis gestellt.	2 (B-T) 3 (A)
		Intensität des Lesens bezogen auf die Anzahl der Wörter pro Satz im Verhältnis zur tatsächlichen Fixationsanzahl pro Satz		
		<i>Lesefluss</i>	Es wird die formale Lesegenauigkeit über den gesamten Text hin überprüft (konstante Lesegenauigkeit, wechselnde Lesegenauigkeit, max. in zwei Sätzen erhöhte Lesegenauigkeit).	3 (B-T) 2 (A)
		Durchschnittliche Lesegenauigkeit über den gesamten Leseprozess (Text oder Events)		
		<i>Wortfixierungen nach Fixationsanzahl</i>	Eine Wortfixierung nach Fixationsanzahl wird angesehen, wenn mehr als zwei Fixationen pro Silbenanzahl eines Wortes gezählt werden konnten.	3
		Wortfixierungen nach Fixationsanzahl bezieht sich auf die Fixationshäufigkeit eines Wortes		
		<i>Wortfixierungen nach Fixationsdauer</i>	Ist ein Fixationspunkt im Verhältnis zu dem durchschnittlichen Punktdurchmesser pro Satz mindestens doppelt so groß, so wird dies als Wortfixierung angesehen.	3
		Wortfixierungen nach Fixationsdauer bezieht sich auf die Größe bzw. den Durchmesser einzelner Fixationspunkte		
		<i>Gesamte Wortfixierungen</i>	Es werden die Häufigkeiten der Wortfixierungen nach Fixationsanzahl und Fixationsdauer gezählt.	3
		Die Summe aus der Anzahl der Wortfixierungen nach Fixationsanzahl und Fixationsdauer		
		<i>Fixationsanzahl Text</i>	Es werden alle Fixationen innerhalb des textlichen Elementes gezählt.	2
		Alle Fixationspunkte pro textliches Element		
	Inhaltlich	<i>Inhaltliche Lesewiederholungen</i>	Es wird die Anzahl der Sätze in Bezug auf den jeweiligen Textabschnitt bzw. der jeweiligen Events inhaltsbezogen ausgewertet, welche mehrfach gelesen wurden (formale Lesewiederholung).	4
		Die Sätze (bezogen auf die Kontexteinheiten), die mehrfach im gesamten Text oder in einzelnen Events gelesen werden (Lesewiederholungen pro Kontexteinheit)		

(Fortsetzung)

Tab. 2.1 (Fortsetzung)

Codiereinheit		Oberkategorie	Codieranweisung	SK
		<i>Inhaltliche Lesegenauigkeit</i>	Es wird die Anzahl der Sätze in Bezug auf den jeweiligen Textabschnitt bzw. der jeweiligen Events inhaltsbezogen ausgewertet, welche eine erhöhte Fixationsanzahl pro Satz als Wörter pro Satz aufweisen (formale Lesegenauigkeit).	4
		Die Sätze (bezogen auf die Kontexteinheiten), die eine erhöhte formale Lesegenauigkeit über den gesamten Text oder in den einzelnen Events aufweisen (Lesegenauigkeit pro Kontexteinheit)		
		<i>Wortfixierungen nach Inhalt</i>	Es werden die fixierten Wörter auf Basis der Fixationsanzahl und der Fixationsdauer in Gruppe „Fachbegriffe“ oder „keine Fachbegriffe“ zugeordnet.	4
		Auf Basis der Wortfixierungen nach Fixationsanzahl und Fixationsdauer werden die fixierten Wörter inhaltsanalytisch beschrieben		
Bild	Formal	<i>Anzahl Bild-Text-Verknüpfungen</i>	Eine Bild-Text-Verknüpfung wird dann gezählt, wenn innerhalb des Lesens Blicksprünge zur bildlichen Repräsentation und im Anschluss erneut auf den Text (auf die letzte oder nächstgelegene Textstelle) vorgenommen werden. Dabei wird der Blickweg wie folgt beschrieben: letzte Textstelle, bildliche Teilelemente, neue Textstelle.	3
		Summe der absoluten Verknüpfungen zwischen beiden Einzelrepräsentationen (Transitionen)		
		<i>Anzahl Bilderschließungen</i>	Eine Bilderschließung wird dann gezählt, wenn keine Bild-Text-Verknüpfungen (Transitionen) vorgenommen werden. Es wird der Blickweg über das Aufzählen der bildlichen Teilelemente dokumentiert.	2
		Summe textungebundener Bildbetrachtung (alleiniges Betrachten der Elemente innerhalb des Bildes, keine Transitionen zum Text)		
		<i>Fixationsanzahl Bild</i>	Es werden alle Fixationen innerhalb des bildlichen Elementes gezählt.	2
		Alle Fixationspunkte pro bildliches Element		
	Inhaltlich	<i>Anzahl inhaltlich korrekter Verknüpfungen</i>	Es werden die korrekten Verknüpfungen zwischen den Wörtern und/oder Sätzen und der bildlichen Elemente innerhalb der Einzelrepräsentation analysiert, ausgezählt sowie prozentual ins Verhältnis zu den Bild-Text Verknüpfungen gestellt.	3
		Prozentuale Auswertung der fachlich korrekten Bild-Text Verknüpfungen		

Tab. 2.2 Kategorien für das zweite Lesen bei der Bild-Text-Kombination ($0.95 \leq \kappa \leq 1$, $p = .000$). SK = Subkategorie

Codiereinheit		Oberkategorie	SK
Text	Formal	Zweites Lesen	2
		Beginn zweites Lesen	2
		Reihenfolge beim zweiten Lesen	2
		Satzanzahl beim zweiten Lesen	3
		Lesefluss zweites Lesen	3
		Fixationsanzahl Text beim zweiten Lesen	3
	Inhaltlich	Inhaltsbezogene Lesewiederholung	3
Bild	Formal	Anzahl Bild-Text-Verknüpfungen beim zweiten Lesen	3
		Anzahl der Bilderschließungen beim zweiten Lesen	2
		Fixationsanzahl Bild zweites Lesen	2
	Inhaltlich	Anzahl inhaltlich korrekter Verknüpfungen beim zweiten Lesen	3

die formalen Kategorien genutzt, um repräsentationsbedingte Muster zu analysieren und Triangulationen mit repräsentationsbezogenen, schüler*innen-spezifischen Merkmalsausprägungen (z. B. kognitiver Stil) zu ermöglichen (Kastaun et al., 2021). Basierend auf einer Auswahl formaler Kategorien werden vier Merkmalsräume für die einzelnen Repräsentationskombinationen gebildet (Tab. 2.3). Es erfolgt eine Zuordnung der Fälle (A-UBT) zu den einzelnen Zellen des Merkmalsraumes mittels der kategorienbasierten Beschreibungen. Über eine Fallkontrastierung, also den Vergleich aller Fälle innerhalb einer Zelle, wird die interne Homogenität geprüft, um die Abgrenzung zu den anderen gruppierten Fällen zu gewährleisten (externe Heterogenität) (Kelle & Kluge, 2010).

Tab. 2.3 Ein vierdimensionaler Merkmalsraum am Beispiel der formal-textbezogenen Kategorien der Bild-Text-Kombination (BT)

				LESEFLUSS						
				Wechsel innerhalb des Leseflusses (1/2/0)		max. 2 Wechsel im Lesefluss		durchweg mehr Fixationen als Wörter pro Satz (2)		
				FORMALE LESEGENAUIGKEIT						
				kein oder max. 1 Satz hat 20 Fixationen / Satz als Wörter/ Satz mehr	mind. 2 Sätze haben 20 Fixationen/ Satz als Wörter/ Satz mehr	kein oder max. 1 Satz hat 20 Fixationen / Satz als Wörter/ Satz mehr	mind. 2 Sätze haben 20 Fixationen/ Satz als Wörter/ Satz mehr	kein oder max. 1 Satz hat 20 Fixationen / Satz als Wörter/ Satz mehr	mind. 2 Sätze haben 20 Fixationen/ Satz als Wörter/ Satz mehr	
FIXATIONSANZAHL TEXT (Erstes Lesen)	200-350 Fixation	ZWEITES LESEN	ja	JBT / KBT / PBT/ TBT / UBT		HBT				
			nein	LBT		IBT				
	351-500 Fixationen		Ja		RBT	GBT	BBT	DBT / FBT		
			nein			NBT	EBT / QBT			CBT / MBT
	> 500 Fixationen		Ja							OBT / SBT
			nein							ABT

Unter Hinzunahme theoretischer und empirisch geprüfter Konzepte zum Umgang mit Texten und Bildern (z. B. Philipp, 2015; Schnotz & Lowe, 2003) wird eine Reduktion der Merkmalsräume vorgenommen, um die Gemeinsamkeiten und Unterschiede der Fallzuordnungen zu spezifizieren und größere Fallgruppen zu bilden. Dabei können zwölf unterschiedliche Strategien im Umgang mit den bildlichen und textlichen Informationen identifiziert werden (Tab. 2.4). Exemplarisch werden die formal-textbezogenen Strategien der Bild-Text-Kombination und die formal-bildbezogenen Strategien der Animation im Folgenden näher beschrieben.

Formal-textbezogene Strategien der Bild-Text-Kombination: Mittels der formal-textbezogenen Kategorien können drei unterschiedliche Strategien (Tab. 2.4) im Umgang mit der textlichen Repräsentation innerhalb der Bild-Text-Kombination herausgearbeitet werden. Anhand der Fälle JBT/KBT/PBT/TBT/UBT/LBT und RBT (Tab. 2.3) und deren gemeinsamer subkategorialer Ausprägungen in den Oberkategorien *Lesefluss*, *Formale Lesegenauigkeit*, der *Fixationsanzahl Text beim ersten Lesen* sowie *Zweites Lesen* wird die Strategie des selektiv-gezielten Lesens abgeleitet. Wie an den Beispielfällen JBT/KBT/TBT und UBT ersichtlich (Abb. 2.2), haben sie einen ständigen Wechsel innerhalb ihres Leseflusses beim ersten Leseprozess gemein, d. h., sie lesen die einzelnen Sätze des Textes mit einer stetig variierenden Lesegenauigkeit. Darüber hinaus haben sie mit einer Ausnahme (Fall RBT) keinen bzw. nur einen Satz mit einer wesentlich höheren Fixationsanzahl (mindestens 20 mehr) pro Satz als Wörter pro Satz gelesen. Bezieht man die Kategorien *Fixationsanzahl Text* beim ersten und zweiten Lesen mit ein, so weisen beide eine sehr geringe Fixationsanzahl (200–350 Fixationen/ Ausnahmefall RBT) und ein erneutes Lesen nach dem ersten Erfassen des Textes auf. Kontrastierend dazu stehen die Fälle DBT/FBT/CBT/MBT/OBT/ SBT und ABT (Abb. 2.2). Sie vereint ein intensiver *Lesefluss*, da die Anzahl der Fixationen beim ersten Lesen durchweg höher ist als die Anzahl der Wörter pro Satz. Mit Ausnahme der Fälle DBT und FBT können fallübergreifend mindestens zwei Sätze mit mehr als 20 Fixationen mehr als Wörter pro Satz in der Kategorie *Lesegenauigkeit* festgestellt werden. Im Gegensatz zu den selektiv-gezielten

Tab. 2.4 Formal-text- und bildbezogene Strategien im Umgang mit der Bild-Text-Kombination und der Animation ($\kappa = 1, p = .000$)

		Strategien im Umgang mit Bild-Text-Kombination	Strategien im Umgang mit Animation
Text	Formal	Selektiv-gezieltes Lesen	Extensiv-suchendes Lesen
		Intensiv-umfassendes Lesen	Wiederholend-intensives Lesen
		Partiell-detailliertes Lesen	Selektiv-wiederholendes Lesen
Bild	Formal	Bild-Text-Erschließung	Bild-Text-Erschließung
		Selektive Bilderschließung und Text-Bild-Integration	Selektive Bilderschließung und Text-Bild-Integration
		Text-Bild-Integration	Text-Bild-Integration

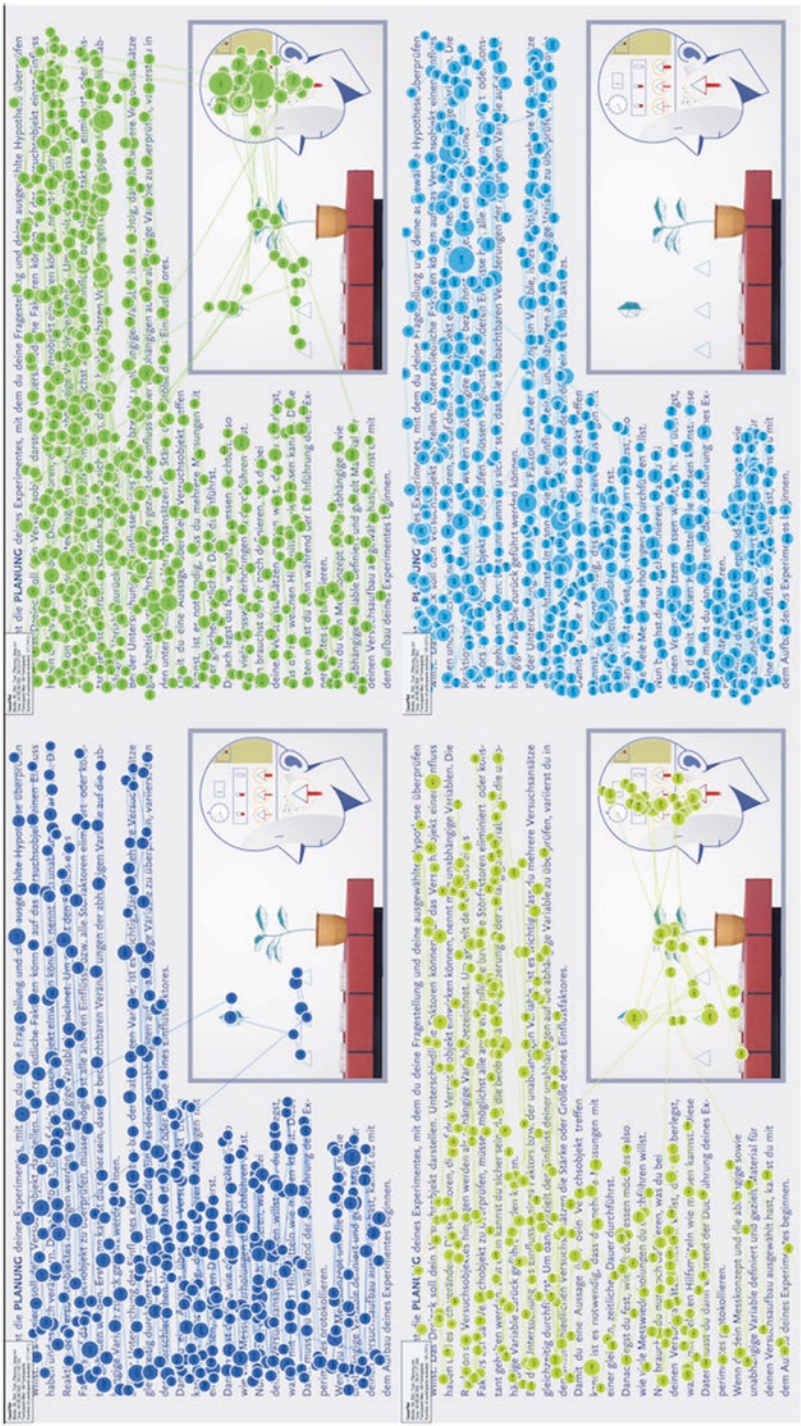


Abb. 2.2 Strategien: Selektiv-gezieltes Lesen (Fälle: JBT, UBT; links) und intensiv-umfassendes Lesen (Fälle: OBT, SBT; rechts) anhand der *gaze plot*-Visualisierungen der Bild-Text-Kombination

Leserinnen und Lesern variieren diese Fälle in der Durchführung eines zweiten Leseprozesses und weisen im Durchschnitt eine wesentlich höhere Fixationsanzahl auf den Text beim ersten Lesen auf. Fußend auf diesen subkategorialen Ausprägungen werden diese Fälle als intensiv-umfassende Leserinnen und Leser zusammengefasst. Als weitere textbezogene Strategie kann das partiell-detaillierte Lesen identifiziert werden. Die zugeordneten Fälle eint ein *Lesefluss* mit maximal zwei Wechseln in der satzbezogenen Lesegenauigkeit und eine mittlere *Fixationsanzahl Text beim ersten Lesen*, sie variieren aber in ihren Ausprägungen innerhalb der Kategorien *Formale Lesegenauigkeit* und *Zweites Lesen*.

Formal-bildliche Strategien der Animation: Auf Basis der formal-bildlichen Kategorien können über Fallzuordnungen drei unterschiedliche Strategien im Umgang mit der dynamischen Bildrepräsentation ermittelt werden (Tab. 2.4). Über die Zuordnung der Fälle BA/DA/MA und NA in den Kategorien der *Anzahl Bild-Text-Verknüpfungen*, der *Anzahl Bilderschließungen* sowie der Einbeziehung der *Fixationszahl Bild* werden fallübergreifende Gemeinsamkeiten hinsichtlich der Ausprägung in den Subkategorien identifiziert. Zusammengefasst als die Strategie der *selektiven Bilderschließung und Text-Bild-Integration* weisen diese Fälle eine geringe Anzahl an Bilderschließungen sowie Bild-Text-Verknüpfungen auf und somit eine geringe Gesamtfixationsanzahl auf dem bewegten Bild. Im Gegensatz dazu weisen die Fälle AA/GA/HA/IA eine geringe Anzahl an Bild-Text-Verknüpfungen auf. Sie eint eine hohe Einbeziehung des Bildes innerhalb des Leseprozesses, teilen aber eine geringe Anzahl an Bilderschließungen. Die Anzahl der Fixationen auf dem Bild variiert jedoch fallspezifisch. Zusammengefasst lassen sich diese Fälle als Bild-Text-Integrierer beschreiben. Die Strategie der *Bild-Text-Erschließung* lässt sich über die subkategorialen Ausprägungen bezüglich einer hohen Anzahl an Bilderschließungen (Ausnahmefall: KA) und einer mittleren Anzahl an Bild-Text-Verknüpfungen beschreiben. Die Fälle EA/JA/KA/FA und LA variieren jedoch ebenfalls hinsichtlich ihrer Gesamtfixationsanzahlen auf dem Bild (Abb. 2.3).

2.6 Diskussion

In der vorliegenden Studie wurde eine methodische Vorgehensweise (weiter-) entwickelt, Blickdaten qualitativ über kategoriale Beschreibungen des Blickverhaltens von Schüler*innen zu analysieren. Ein zentraler Befund stellt zunächst die Transkription der individuellen Blickdaten über das kategoriengestützte Auszählen und Summieren von Fixationspunkten auf bildlichen, symbolischen und textlichen Elementen dar. Auf Basis dieser Transkripte und unter Hinzunahme der methodischen Schritte der quantitativen und qualitativen Inhaltsanalyse (Hutter, 2019; Mayring, 2016) konnten (weitere) Kategorien induktiv aus dem Material übergreifend und spezifisch für die Bild-Text-Kombination und Animation abgeleitet und definiert werden. Über 14 repräsentationsübergreifende und elf (Bild-Text) bzw. zwei (Animation) spezifische Kategorien lässt

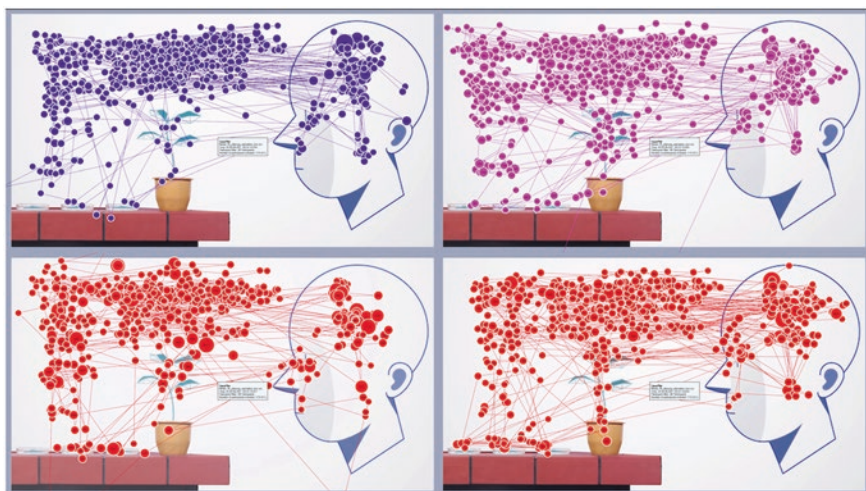


Abb. 2.3 Strategien: Selektive Bilderschließung und Text-Bild-Integration (Fälle: BA, MA; links) und Bild-Text-Erschließung (Fälle: JA, BA; rechts) anhand der *gaze plot*-Visualisierungen der Animation

sich das Blickverhalten der einzelnen Probanden sowohl inhaltsbezogen als auch formal deskriptiv beschreiben (F1). Unter Hinzunahme theoretischer Konzepte, wie der Lesearten und -strategien (Leisen, 2010; Mayer & Marks, 2019; Philipp, 2015), empirischer (Eye-Tracking) Befunde aus der Leseforschung (z. B. Kim & Lombardino, 2016) sowie des indikatorischen Bildverstehens (z. B. Niegemann et al., 2008; Schnotz & Lowe, 2003) und der kategorienbezogenen Zuordnung der Fälle, konnten zwölf unterschiedliche Strategien im Umgang mit den textlichen und bildlichen Einzelrepräsentationen beschrieben werden (F2). Offen bleibt jedoch, ob das schülerspezifische Blickverhalten hinsichtlich der untersuchten monomedialen Lernunterstützungen auf unterschiedliche Ausprägungen kognitiver und motivationaler Merkmale basiert. Mögliche Einflüsse könnten auf Unterschiede in der individuellen Varianz des Vorwissens (z. B. Schnotz & Lowe, 2003; Weidenmann, 1993), der sprachlichen Fähigkeiten, der Repräsentationskompetenz (Mayer & Massa, 2003) oder des kognitiven Stils (Koć-Januchta et al., 2019) zurückgeführt werden. Dies gilt es, zukünftig über (Daten-)Triangulationen inhaltlich zu untersuchen, um auch die (methodische) Aussagekraft der qualitativen Ergebnisse der hier vorliegenden Studie zu überprüfen. Neben den individuellen Merkmalen der Lernenden kann aber auch die Gestaltung der Lernunterstützungen selbst eine Ursache für das aufgezeichnete Blickverhalten sein. So liegt der visuellen Darstellungsweise der statischen und bewegten Bilder in den Lernunterstützungen keine fachliche Logik zugrunde, wie es bei der Erfassung von Diagrammen oder Schemata der Fall ist (z. B. Schnotz, 2002). Hier greifen attentive Regeln, welche u. a. eine Systematik in der Bildernutzung und eine größere mentale Anstrengung seitens der Lernenden verlangen (Niegemann et al., 2008).

Die textlichen Elemente der Lernunterstützungen sind dahingehend konzipiert, dass teilweise sprachlich ein direkter Bezug auf die Abbildungen gegeben und so das Bildverstehen unterstützt wird (Weidenmann, 1993). Ebenso ist die visuelle Aufmerksamkeitssteuerung der bewegten Bilder in die Interpretation der Ergebnisse mit einzubeziehen. So zeigt sich ein erheblicher Unterschied in der Nutzungshäufigkeit der bildlichen Darstellungen (z. B. gemessen über die Kategorie *Anzahl Bild-Text-Verknüpfung*, *Anzahl Bilderschließung*) zwischen der Bild-Text-Kombination und der Animation, welche sich auf die Dynamik zurückführen lässt.

Aufgrund der hohen Informationsdichte in der Planungsphase ist vor allem der Text in den Lernunterstützungen umfangreich und komprimiert. In der Zusammenführung der in dieser Studie eingesetzten Lernunterstützungen mit der Eye-Tracking-Mess- und Datenerfassungsmethodik haben sich insbesondere bei den textreichen Formaten methodisch-technische Grenzen gezeigt. Die Auswertung der Blickpunkte auf den textlichen Elementen stellte eine besondere Herausforderung dar, da teilweise an den Rändern des Bildschirms bzw. der jeweiligen Lernunterstützungen Kalibrierungsfehler sichtbar wurden, d. h., die Blickpunkte verrutschten in den Satzzeilen oder lagen außerhalb des aufgenommenen Bereichs und erschwerten somit die eindeutige Transkription und demzufolge die qualitative Auswertung. Dies führte bei starker Ausprägung zum Fallausschluss. Zukünftig sollte in der Gestaltung von Lernmaterialien, die mittels einer Blickdatenanalyse untersucht werden sollen, darauf geachtet werden, dass diese auch an das genutzte Werkzeug, d. h. die technischen Voraussetzungen des Eye-Tracking-Systems, angepasst sind.

Zusammenfassend lässt sich feststellen, dass die qualitative Analyse der Blickdaten eine neue Forschungsmöglichkeit liefert, das Verhalten von Lernenden während der Nutzung von Lernmaterialien zu analysieren. Gemessen an dem hohen zeitlichen Aufwand kann auch die Tragweite der hier dargelegten Auswertungsmethodik als umfassend beschrieben werden. Das hier gewählte methodische Vorgehen liefert Aufschlüsse über das individuelle Blickverhalten und bildet in einem hypothesengenerierenden Forschungsansatz Ausgangspunkt für sich vielfältig anschließende quantitative Forschungsfelder. Zukünftig sollten die gewonnenen methodischen Erkenntnisse auch in die (Weiter-)Entwicklung digitaler Möglichkeiten der qualitativen Analyse von Blickdaten einfließen bzw. Wege identifiziert werden, um über ökonomischere Auswertungen einen breiteren Zugang in die Eye-Tracking-Forschung zu realisieren. Aus Sicht der fachdidaktischen Forschung bieten die Ergebnisse der qualitativen Analyse besonders unter dem inhaltlichen Aspekt neue und erweiterte Förderperspektiven. Die über Fallzuordnungen identifizierten Muster und beschriebenen Strategien im Umgang mit den Repräsentationskombinationen zeigen individuelle Unterschiede, die sich formatspezifisch auf die jeweilige Einzelrepräsentation oder, inhaltsbezogen, auf den fachdidaktischen Kontext – die Inhalte der experimentellen Planungsphase – beziehen. Besonders Letzteres eröffnet Anknüpfungspunkte, individuelle Fördermaßnahmen zu erstellen und die experimentellen Kompetenzen der Schüler*innen sowie den Umgang mit Fachtermini und dem dazugehörigen konzeptionellen Verständnis zu stärken bzw. zu unterstützen.

Danksagung und Förderhinweis Hiermit möchten wir uns hinsichtlich der technischen Unterstützung ganz herzlich bei der AG von Prof. Dr. Jochen Kuhn der TU Kaiserslautern (bis 2021) bedanken, ohne die eine Umsetzung dieser Studie nicht möglich gewesen wäre. Das diesem Beitrag zugrunde liegende Vorhaben wurde von der Deutschen Telekom Stiftung im Rahmen des Programms Fellowship Fachdidaktik MINT gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen.

Literatur

- Ansorge, U., & Leder, H. (2011). *Wahrnehmung und Aufmerksamkeit*. Springer Fachmedien.
- Aepkers, M., & Liebig, S. (2002). *Entdeckendes, forschendes und genetisches Lernen*. Schneider-Verlag Hohengehren.
- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16, 183–198.
- Arnold, J., Kremer, K., & Mayer, J. (2014). Understanding students' experiments – What kind of support do they need in inquiry tasks? *International Journal of Science Education*, 36, 2719–2749.
- Arnold, J., Kremer, K., & Mayer, J. (2017). Scaffolding beim Forschenden Lernen – Eine empirische Untersuchung zur Wirkung von Lernunterstützungen. *Zeitschrift für Didaktik der Naturwissenschaften*, 23, 21–37.
- Baur, A. (2018). Fehler, Fehlkonzepte und spezifische Vorgehensweisen von Schüler*innen und Schülern beim Experimentieren: Ergebnisse einer videogestützten Beobachtung. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1), 115–129.
- Brush, T. A., & Saye, J. W. (2002). A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment. *The Journal of Interactive Online Learning*, 1(2), 1–12.
- Brückner, S., Zlatkin-Troitschanskaia, O., Küchemann, S., Klein, P., & Kuhn, J. (2020). Changes in students' understanding of and visual attention on digitally represented graphs across two domains in higher education: A postreplication study. *Frontiers in Psychology*, 11, 2090.
- Dresing, T., & Pehl, T. (2015). *Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende* (6. Aufl.). Marburg. www.audiotranskription.de/praxisbuch.
- Dresing, T., & Pehl, T. (2020). Transkription. Implikationen, Auswahlkriterien und Systeme für psychologische Studien. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 1–20). Springer Fachmedien.
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2009). *Recent experimental studies of inquiry-based teaching: A meta-analysis and review*. Paper presented at the European Association for Research on Learning and Instruction, August 25–29, Amsterdam, Netherlands.
- Gwet, K. L. (2012). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters*. Advanced Analytics, LLC.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle. Merkmale und ihre Bedeutung – dargestellt anhand von Kompetenzen beim Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht*, 57(4), 196–203.
- Höfler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations. *Computers in Human Behavior*, 27, 209–216.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Hutter, S. (2019). Quantitative Inhaltsanalyse. In C. Wagemann, A. Goerres, & M. M. Siewert (Hrsg.), *Handbuch Methoden der Politikwissenschaft*. Springer VS.

- Irion, T., & Scheiter, K. (2018). Didaktische Potenziale digitaler Medien für den Grundschulunterricht. Der Einsatz digitaler Technologien aus grundschuldidaktischer und medien-didaktischer Sicht. *Grundschule aktuell*, 142, 8–11.
- Kaiser, I., Malai, D., & Mayer, J. (2018). Self-generation in the context of inquiry-based learning. *Frontiers in Psychology*, 9, 2440.
- Kastaun, M., & Meier, M. (2018). Kognitive Lernunterstützungen im naturwissenschaftlichen Erkenntnisprozess (Projektskizze). In D. Krüger (Hrsg.), *Erkenntnisweg Biologiedidaktik 17* (S. 123–136). Köln.
- Kastaun, M., Hunold, C., & Meier, M. (2020). Eye-Tracking – Visuelle Wahrnehmung sichtbar machen. *Biologie in unserer Zeit*, 3, 172–173.
- Kastaun, M., Meier, M., Küchemann, S., & Kuhn, J. (2021). Validation of cognitive load during inquiry-based learning with multimedia scaffolds using subjective measurement and eye movements. *Frontiers in Psychology*, 12, 703857.
- Kelle, U., & Kluge, S. (2010). *Vom Einzelfall zum Typus – Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung*. Springer Fachmedien.
- Kim, S., & Lombardino, L. J. (2016). Simple sentence reading and specific cognitive functions in college students with Dyslexia: An eye-tracking study. *Clinical Archives of Communication Disorders*, 1(1), 48–61.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1–48.
- Krebs, M.-C., Schüler, A., & Scheiter, K. (2019). Just follow my eyes: The influence of model-observer similarity on eye movement modeling examples. *Learning and Instruction*, 61, 126–137.
- Krejtz, K., Duchowski, A. T., Krejtz, I., Kopacz, A., & Chrzastowski-Wachtel, P. (2016). Gaze transitions when learning with multimedia. *Journal of Eye Movement Research*, 9, 1.
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Beltz Juventa.
- Koć-Januchta, M. M., Höffler, T. N., Eckhardt, M., & Leutner, D. (2019). Does modality play a role? Visual-verbal cognitive style and multimedia learning. *Journal of Computer Assisted Learning Psychology*, 35(6), 1–11.
- Lamneck, S., & Krell, C. (2016). *Qualitative Sozialforschung*. Beltz.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 1.
- Lederman, J. S., Lederman, N. G., Bartos, S. A., Bartels, S. L., Meyer, A. A., & Schwartz, R. S. (2014). Meaningful assessment of learners' understandings about scientific inquiry – The Views About Scientific Inquiry (VASI) questionnaire. *Journal of Research in Science Teaching*, 51(1), 65–83.
- Leisen, J. (2010). *Handbuch Sprachförderung im Fach. Sprachsensibler Fachunterricht in der Praxis*. Varus.
- Luck, S. J., & Ford, M. A. (1998). On the role of selective attention in visual perception. *PNAS*, 95(4), 825–830.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.
- Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning*. Cambridge University Press.
- Mayer, A., & Marks, D.-K. (2019). Förderung des Textverständnisses durch die Vermittlung von Verstehensstrategien – Eine Metaanalyse zur Effektivität. *Forschung Sprache*, 1, 4–33.
- Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *JoEP*, 95(4), 833–846.
- Mayring, P. (2016). *Qualitative Inhaltsanalyse*. Beltz.

- Meier, M. (2016). *Entwicklung und Prüfung eines Instrumentes zur Diagnose der Experimentierkompetenz von Schüler*innen und Schülern*. Logos.
- Meier, M., & Kastaun, M. (2021). Lernunterstützungen als Werkzeug individualisierter Förderung im naturwissenschaftlichen Erkenntnisprozess. In M. Meier, C. Wulff, & K. Ziepprecht (Hrsg.), *Vielfältige Wege biologiedidaktischer Forschung – Vom Lernort Natur zur Naturwissenschaftlichen Erkenntnisgewinnung in die Lehrerverprofessionalisierung*. Waxmann.
- Neuendorf, K. A. (2017). *The content analysis guidebook*. Sage.
- Niegemann, H. M., Domagk, S., Hessel, S., Hein, A., Hupfer, M., & Zobel, A. (2008). *Kompodium multimediales Lernen*. Springer.
- Nilsson Benfatto, M., Öqvist Seimyr, G., Ygge, J., Pansell, T., Rydberg, A., & Jacobson, C. (2016). Screening for dyslexia using eye tracking during reading. *PLoS ONE*, 11, 12.
- Nitz, S., Ainsworth, S., Nerdel, C., & Precht, H. (2014). Do student perceptions of teaching predict the development of representational competence and biological knowledge? *Learning and Instruction*, 31, 13–22.
- Park, B., Korbach, A., & Brünken, R. (2015). Do learner characteristics moderate the seductive-details-effect? A cognitive-load-study using eye-tracking. *Journal of Educational Technology & Society*, 18(4), 24–36.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning styles – Concepts and evidence. *Psychological Science in the Public Interest*, 9(3), 105–119.
- Philipp, M. (2015). *Lesestrategien. 2015 Bedeutung, Formen und Vermittlung*. Beltz Juventa.
- Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(2), 1–12.
- Richter, J., Scheiter, K., & Eitel, A. (2018). Signaling text-picture relations in multimedia learning: The influence of prior knowledge. *Journal of Educational Psychology*, 110(4), 544–560.
- Rössler, P. (2017). *Inhaltsanalyse*. UTB.
- Scheiter, K., Schubert, C., & Schüler, A. (2018). Self-regulated learning from illustrated text: Eye movement modeling to support use and regulation of cognitive processes during learning from multimedia. *British Journal of Educational Psychology*, 88, 80–94.
- Schmidt-Weigand, F., Franke-Braun, G., & Hänze, M. (2008). Erhöhen gestufte Lernhilfen die Effektivität von Lösungsbeispielen? Eine Studie zur kooperativen Bearbeitung von Aufgaben in den Naturwissenschaften. *Unterrichtswissenschaft*, 36(4), 365–384.
- Schnotz, W. (2002). Wissenserwerb mit Texten, Bildern und Diagrammen. In L. J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia und Internet* (3. Aufl., S. 65–81). Psychologie Verlags Union.
- Schnotz, W., & Lowe, R. K. (2003). External and internal representations in multimedia learning. *Learning and Instruction*, 13, 117–123.
- Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research and Development*, 53(3), 47–58.
- Smyrnakis, I., Andreadakis, V., Selimis, V., Kalaitzakis, M., Bachourou, T., Kaloutsakis, G., et al. (2017). RADAR: A novel fast-screening method for reading difficulties with special focus on dyslexia. *PLoS ONE*, 12, 8.
- Solheim, O. J., & Upstad, P. H. (2011). Eye-tracking as a tool in process-oriented reading test validation. *International Electronic Journal of Elementary Education*, 4(1), 153–168.
- Sweller, J. (1999). *Instructional design in technical areas*. ACER Press.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (S. 19–30). Cambridge University Press.
- Takacs, Z. K., & Bus, A. G. (2016). Benefits of motion in animated storybooks for children's visual attention and story comprehension. An eye-tracking study. *Frontiers in Psychology*, 7, 1591.

- Tobii AB. (2016). Tobii Studio User's Manual version 3.4.8 Stockholm: Tobii AB. <https://www.tobii.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf>. Zugegriffen: 6. März 2020.
- Weidenmann, B. (1993). Informierende Bilder. In B. Weidenmann (Hrsg.), *Wissenserwerb mit Bildern* (S. 9–58). Huber.
- Weidenmann, B. (2002). Abbilder in Multimediaanwendungen. In L. J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia und Internet: Lehrbuch für Studium und Praxis* (3. Aufl., S. 83–96). Psychologie Verlags Union.
- Wichmann, A., & Leutner, D. (2009). Inquiry learning: Multilevel support with respect to inquiry, explanations and regulation during an inquiry cycle. *Zeitschrift für Pädagogische Psychologie*, 23(2), 117–127.
- Wu, H. L., & Pedersen, S. (2011). Integrating computer- and teacher-based scaffolds in science inquiry. *Computers & Education*, 57(4), 2352–2363.

Muster- und Strukturerkennungsprozesse bei potenziell mathematisch begabten Grundschulkindern untersuchen – Das Potenzial von Eye-Tracking und retrospektivem Interview im Vergleich

Dinah Reuter und Julia Bruns

Inhaltsverzeichnis

3.1	Einleitung	41
3.2	Muster- und Strukturerkennung im Kontext von Begabung	42
3.3	Erfassung der Muster- und Strukturerkennung	43
3.4	Forschungsfragen	44
3.5	Material und Methode	44
3.6	Ergebnisse	47
3.7	Diskussion	54
	Literatur	56

3.1 Einleitung

Muster- und Strukturerkennung beschreibt eine individuelle, konstruktive Tätigkeit aufgrund der Verarbeitung visueller Wahrnehmung (Lüken, 2012). Studienergebnisse stützen die angenommene hohe Bedeutung der Muster- und Strukturerkennung

D. Reuter (✉)

Institut für Mathematische Bildung Freiburg,
Pädagogische Hochschule Freiburg, Freiburg im Breisgau, Deutschland
E-Mail: dinah.reuter@ph-freiburg.de
URL: <https://home.ph-freiburg.de/reuter>

J. Bruns

Institut für Mathematik (Mathematikdidaktik),
Universität Paderborn, Paderborn, Deutschland
E-Mail: julia.bruns@uni-paderborn.de

für die mathematische Entwicklung (Rittle-Johnson et al., 2019). Zusätzlich wird die Fähigkeit zur Muster- und Strukturerkennung als Merkmal mathematischer Begabung angesehen (z. B. Heinze, 2005; Käpnick, 1998, 2008; Singer et al., 2016). Dabei wird der *Prozess* des Erkennens von Mustern bzw. deren Struktur fokussiert (ebd.). Im Widerspruch dazu werden in vorliegenden Studien zur Muster- und Strukturerkennung größtenteils Produkte der Kinder, jedoch nicht der Erkennungsprozess betrachtet (Lüken & Sauzet, 2020). Studien zum Muster- und Strukturerkennungsprozess basieren häufig auf (Material-)Handlungen der Kinder zu Musterfolgen (z. B. Lüken & Sauzet, 2020). Im Kontext mathematischer Begabung werden mehrheitlich aufgabenbegleitende *verbale Erklärungen* genutzt. Dabei kommen auch Muster mit räumlicher Struktur zum Einsatz (z. B. Aßmus, 2017). Beide Zugänge zum Prozess der Muster- und Strukturerkennung – Handlungen und verbale Erklärungen – unterliegen Beschränkungen: Während der Zugang über Materialhandlungen nur bedingt auf Muster mit räumlicher Struktur anwendbar ist, ist der verbale Zugang durch metakognitive Kompetenzen sowie Verbalisierungsfähigkeiten der Kinder beeinflusst (Peter-Koop, 2002; Schindler & Lilienthal, 2019). Inwieweit das letztgenannte Argument auch die Validität der verbalen Aussagen mathematisch begabter Grundschulkinder gefährdet, die häufig über hohe metakognitive Kompetenzen verfügen, ist bislang unklar. Um sich dieser Frage zu nähern, kann Eye-Tracking als weiterer Ansatz herangezogen werden, da Eye-Tracking zusätzlich zu Interviews auch Prozessdaten zur Muster- und Strukturerkennung liefert (z. B. Schindler et al., 2019; Sprenger & Benz, 2020). Zentrales Erkenntnisinteresse der vorliegenden Studie ist daher die Untersuchung des Potenzials des Eye-Trackings zur Erfassung des Muster- und Strukturerkennungsprozesses mathematisch begabter Grundschulkinder.

3.2 Muster- und Strukturerkennung im Kontext von Begabung

Versteht man Mathematik als Wissenschaft der Muster (z. B. Devlin, 1998; Sawyer, 1955), da sie das Ziel verfolgt, Regelmäßigkeiten, Beziehungen und Zusammenhänge aufzuspüren und zu beschreiben, wird die Bedeutung der Muster- und Strukturerkennung als grundlegende mathematische Tätigkeit deutlich (Käpnick, 1998, 2008). Bei erkannten Strukturen handelt es sich um Deutungen (Lüken, 2012), die aufgrund der individuellen Interpretation unterschiedlich sein können (Steinweg, 2013).

Quantitative Studien zur Muster- und Strukturerkennung von jungen Kindern stellen Zusammenhänge zwischen der Muster- und Strukturerkennung und der kognitiven Flexibilität (Schmerold et al., 2017), dem Arbeitsgedächtnis (Fyfe et al., 2017; Schmerold et al., 2017) sowie den allgemeinen mathematischen Fähigkeiten (Lüken, 2012; Schmerold et al., 2017) heraus. Dabei erklärt die Fähigkeit der Muster- und Strukturerkennung über die kognitiven Fähigkeiten hinaus Varianz in der mathematischen Leistung (Schmerold et al., 2017). Qualitative Studien beschreiben die Ausprägung der Muster- und Struktur-

erkennung von Kindern sowie Unterschiede in der Muster- und Strukturerkennung in Abhängigkeit von der mathematischen Leistung (Lüken, 2012; Mulligan & Mitchelmore, 2009). Diese Ergebnisse deuten im Einklang mit den Ergebnissen quantitativer Studien darauf hin, dass ausgeprägte Fähigkeiten der Muster- und Strukturerkennung eng mit hohen mathematischen Leistungen zusammenhängen. Ergänzend zeigt sich in Studien, dass mathematisch begabte Kinder im Vergleich zu anderen Kindern in einem außergewöhnlichen Maß über die Fähigkeit der Muster- und Strukturerkennung verfügen, weswegen diese Fähigkeit als mathematisches Begabungsmerkmal eingestuft wird (z. B. Aßmus, 2017; Heinze, 2005; Käpnick, 1998, 2008; Kruteskii, 1976; Nolte & Kießwetter, 1996).

Die Fähigkeit der Muster- und Strukturerkennung umfasst das „Erkennen und das Bilden von Mustern bzw. Anordnungs- und Gliederungsprinzipien [...] in gegebenen oder zu konstruierenden mathematischen Sachverhalten (z. B. geometrischen Figuren, Zahlen, Gleichungen)“ (Käpnick, 1998, S. 111). Dies beinhaltet die Fähigkeit, wichtige von unwichtigen Informationen trennen und zwischen den relevanten Informationen Beziehungen herstellen zu können. Kießwetter (1993) zufolge werden erkannte Strukturen zu größeren Einheiten, sog. *Superzeichen*, zusammengefasst. Die *Fähigkeit zum Transfer und Verallgemeinern* impliziert, dass erkannte mathematische Muster und Strukturen auf ähnliche Sachverhalte übertragen werden, um neue Probleme zu lösen. Ein entsprechender Rückgriff auf die Muster- und Strukturerkennung findet sich beim Merkmal *Gedächtnisfähigkeit für mathematische Sachverhalte*, bei dem erkannte Muster und Strukturen zum Speichern mathematischer Sachverhalte genutzt werden (Käpnick, 1998, 2008).

3.3 Erfassung der Muster- und Strukturerkennung

Obwohl der Forschungsstand auf eine zentrale Bedeutung der Muster- und Strukturerkennung insbesondere für den Kontext der mathematischen Begabung hinweist, zeigt sich die Schwierigkeit, den *mental*en Prozess der Muster- und Strukturerkennung zu operationalisieren (z. B. Lüken & Sauzet, 2020). Bei diesem Wahrnehmungsprozess muss zwischen *Wahrnehmung*, der sensorischen Erfahrung, und *Erkennen*, der verliehenen Bedeutung, unterschieden werden (Goldstein, 2008; Guski, 2000). Verbale, schriftliche oder handelnde Zugänge zur Muster- und Strukturerkennung adressieren vornehmlich den Prozess des Erkennens. Die zugehörigen Wahrnehmungsprozesse können als weitere Informationsquelle bezüglich der Muster- und Strukturerkennung betrachtet werden. Dem Eye-Tracking wird das Potenzial zugeschrieben, diese Wahrnehmungsprozesse zu erfassen (Holmqvist et al., 2011; Just & Carpenter, 1976; Strohmaier et al., 2020). Bezüglich der Muster- und Strukturerkennung kann vermutet werden, dass einzelne Elemente eines Musters, zwischen denen Beziehungen hergestellt werden, gezielt und ggf. auch hintereinander wahrgenommen werden. Eye-Tracking wird daher in verschiedenen (mathematikdidaktischen) Kontexten genutzt, in denen die Wahrnehmung von Strukturen

bedeutsam ist (z. B. Schindler et al., 2019; Sprenger & Benz, 2020). Im Sinne der *eye-mind assumption* (Just & Carpenter, 1980), die davon ausgeht, dass Fixationen kognitive Verarbeitungsprozesse spiegeln, werden die Blickbewegungen als Indikator für kognitive Wahrnehmungsprozesse interpretiert. Verschiedene Autor*innen weisen auf die Grenzen der *eye-mind assumption* hin, indem sie betonen, dass nicht jede Fixation auf eine kognitive Verarbeitung der entsprechenden Information hindeutet (z. B. Schindler & Lilienthal, 2019; Strohmaier et al., 2020).

Während sich die Nutzung von Eye-Tracking zur Erfassung anderer Wahrnehmungsprozesse von Kindern mit Schwierigkeiten im Bereich Mathematik bereits als geeignet erwiesen hat (Simon & Schindler, 2020), wurde das Potenzial des Eye-Trackings zur Erfassung der Muster- und Strukturerkennungsprozesse mathematisch begabter Grundschul Kinder bislang nicht betrachtet.

3.4 Forschungsfragen

Den *mental*en Prozess der Muster- und Strukturerkennung zu operationalisieren, ist eine Herausforderung empirischer Forschung zur Muster- und Strukturerkennung (Lüken & Sauzet, 2020). Der verbale Zugang zu mentalen Prozessen ist – nicht nur im Kontext der Muster- und Strukturerkennung – begrenzt (s. oben). Eye-Tracking kann als weiterer Zugang zur Erfassung der Muster- und Strukturerkennungsprozesse herangezogen werden, bei dem der Fokus auf den Wahrnehmungsprozess gelegt wird. Dabei interessiert, inwieweit Eye-Tracking Einblicke in die kognitiven Wahrnehmungsprozesse der Muster- und Strukturerkennung ermöglicht und inwiefern sich die auf Grundlage der Eye-Tracking-Daten angenommenen Muster- und Strukturerkennungsprozesse mit den angenommenen Muster- und Strukturerkennungsprozessen auf Grundlage retrospektiver Beschreibungen mathematisch begabter Grundschul Kinder decken. Daraus ergeben sich folgende Forschungsfragen:

1. Welche Annahmen zur Muster- und Strukturerkennung mathematisch begabter Grundschul Kinder ermöglicht die Methode des Eye-Trackings?
2. Welche Unterschiede bzw. Gemeinsamkeiten zeigen sich zwischen den Annahmen zur Muster- und Strukturerkennung, die aufgrund der Blickbewegungen mathematisch begabter Grundschul Kinder getroffen werden, und den Annahmen zur Muster- und Strukturerkennung, die aufgrund ihrer verbalen Erklärungen getroffen werden?

3.5 Material und Methode

Zur Untersuchung der Forschungsfragen wurde eine qualitative *multiple case study* im Laborsetting durchgeführt (Yin, 2018). Dieser Ansatz ist gekennzeichnet durch exploratives Vorgehen mit einem kleinen Sample (Döring & Bortz, 2016).

Stichprobe

Die Gelegenheitsstichprobe¹ besteht aus $N=10$ ($n=6$ männlich, $n=4$ weiblich) Dritt- und Viertklässler*innen. Alle Schüler*innen wurden von ihren Lehrkräften als mathematisch begabt eingeschätzt und nahmen an einem Enrichment-Mathematikangebot an der PH Freiburg teil, wo sie durch ihre Lösungsprozesse auffielen. Da für keines der Kinder ein Begabungsnachweis vorliegt, wird von potenziell mathematisch begabten Kindern gesprochen.

Methode

Zur Erfassung der Muster- und Strukturerkennung wurden die Aufzeichnung der Blickbewegungen mittels Eye-Tracking und ein teilstandardisiertes Interview kombiniert.

Die Aufzeichnung der Blickbewegungen erfolgte durch einen stationären Eye-Tracker des Modells RED Dual Screen Setup von SensoMotoric Instruments monokular (rechtes Auge) mit der Software SMI Experiment Center™ 2.0 und einer Frequenz von 120 Hz. Die maximale Abweichung für die Kalibrierung wurde mit 0.5° festgelegt. Vor jeder Aufgabe erfolgte eine Neukalibrierung. Die Aufgaben wurden auf einem 24-Zoll-Bildschirm mit einer Auflösung von 1920×1080 Pixeln dargestellt. Die Schüler*innen saßen im Abstand von 65–70 cm vor dem Bildschirm und wurden vor jeder Aufgabe darauf hingewiesen, sich an dem Stuhl anzulehnen, um diesen Abstand konstant zu halten.

Direkt nach Bearbeitung jeder einzelnen Aufgabe wurden die Kinder im Sinne eines teilstandardisierten Interviews (Helfferich, 2014) retrospektiv zu ihrem Bearbeitungsvorgehen befragt, ohne die aufgezeichneten Blickbewegungen zu visualisieren. Die Interviewfragen richteten sich dabei auf den Prozess der Muster- und Strukturerkennung (z. B. „Wie hast du dir das so schnell gemerkt?“, „Was würdest du einem Kind in der ersten Klasse für einen Tipp geben, wenn es dieselbe Aufgabe bearbeiten muss?“).

Stimulus: Indikatoraufgaben nach Käpnick (1998)

Als Stimulus für die Datenerhebung dienen vier ausgewählte Indikatoraufgaben zur Diagnose potenziell mathematisch begabter Grundschulkinder nach Käpnick (1998). Die Aufgaben wurden so ausgewählt, dass sowohl geometrische (Aufgabe 1, 2) als auch arithmetische (Aufgabe 3, 4) Muster und Strukturen jeweils mit hohem (Aufgabe 1, 3) und niedrigerem Strukturgehalt (Aufgabe 2, 4) vertreten sind. Durch diese Auswahl ist die Grundanforderung für alle Aufgaben

¹Die Daten wurden in Zusammenarbeit mit Masterstudierenden im Rahmen von Abschlussarbeiten erhoben.

gleich, und es können Muster- und Strukturerkennungsprozesse in unterschiedlichen Inhaltsbereichen bzw. bei Aufgaben mit unterschiedlichem Strukturgehalt verglichen werden. In allen Aufgaben soll ein Stimulus nach einer Phase der Präsentation (10 bzw. 60 s) reproduziert werden. Der Auftrag lautete: „Du hast gleich XX Sekunden Zeit, dir zu merken, was du siehst, und dann sollst du genau aufzeichnen, was du gesehen hast.“

In der ersten Aufgabe (*Zeichenfolge*) werden drei geometrische Muster mit unterschiedlicher Struktur dargestellt (Abb. 3.1a). In der zweiten Aufgabe (*Haus*) ist eine aus geometrischen Figuren zusammengesetzte Figur dargestellt (Abb. 3.1b). Beide Aufgaben werden von Käpnick (1998) zur Untersuchung der Fähigkeit des Speicherns anschaulich-geometrischer Sachverhalte im Kurzzeitgedächtnis vorgeschlagen. In der dritten, *Zahlenfeld* (Abb. 3.1c), und vierten Aufgabe, *Zahlenmauer* (Abb. 3.1d), wird jeweils eine Zahlenanordnung präsentiert. Käpnick (1998) ordnet diese Aufgaben der Fähigkeit des Speicherns gesehener Zahlen im Kurzzeitgedächtnis ggf. unter Nutzung erkannter mathematischer Strukturen zu.

Datengrundlage und Auswertung

Für diese Studie ist von Interesse, inwiefern Schüler*innen Muster und Strukturen in den Stimuli wahrnehmen und erkennen. Unter Einsatz des Eye-Trackings soll beantwortet werden, welche Elemente des Stimulus in welcher Reihenfolge fokussiert werden. Diese Aspekte lassen sich mittels der Visualisierungen Bee-Swarm und Scan-Path (Strohmaier et al., 2020) darstellen, die mithilfe der Software SMI BeGaze™ gewonnen wurden. In beide Visualisierungen gehen alle Blickbewegungen der Schüler*innen ein und nicht nur *real fixations* zwischen 200 und 300 ms (Holmqvist et al., 2011), da insbesondere die Reihenfolge der Fixationen interessiert. Dabei werden die Blickbewegungen – im Sinne der *eye-mind assumption* (Just & Carpenter, 1980) – als Indikator für das Erkennen von

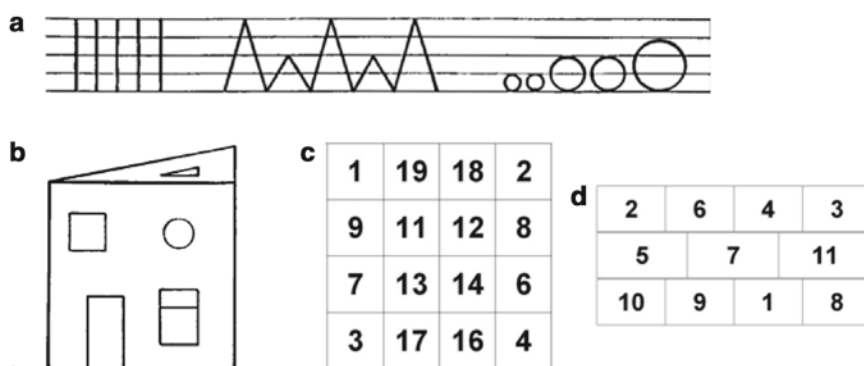


Abb. 3.1 Ausgewählte Aufgaben. (Nach Käpnick 1998, S. 144–147)

Muster und Strukturen interpretiert. Da kognitive Prozesse des Erkennens von Mustern und Strukturen untersucht werden, die eng mit dem Wahrnehmen des Stimulus zusammenhängen, scheint eine Interpretation der Daten nach der *eye-mind assumption* vertretbar.

Die Analyse der Blickbewegungen der Schüler*innen beim Einprägen des Stimulus sowie die verbale Beschreibung des Vorgehens aus den leitfadengestützten, retrospektiven Interviews erfolgte mit der qualitativen Inhaltsanalyse (Mayring, 2015) und einem gemeinsamen Kategoriensystem. Die Codierung der Eye-Tracking-Daten erfolgte am Bee-Swarm-Video; die Daten aus den leitfadengestützten Interviews wurden ebenfalls am Video codiert. Die Codierung der Daten erfolgte anhand deduktiver, aus theoretischen Vorüberlegungen entwickelter Kategorien, die (mögliche) Vorgehensweisen bei der Einprägung des Stimulus beschreiben (z. B. für die Aufgabe *Zahlenmauer*: Fokussierung auf Zahlengruppen (mit operationalen Beziehungen); s. auch Käpnick, 1998, 2008). Als Analyseeinheit wurde für jede Aufgabe das Bee-Swarm-Video bzw. der Videoausschnitt des jeweiligen Interviewausschnitts festgelegt, als Codiereinheit die (angenommene) strukturelle Beziehung zwischen einzelnen Elementen eines Musters. Die Festlegung der Codiereinheit erfolgt demnach nicht a priori, sondern inhaltlich-semantic durch die Codierenden. Die Abgrenzung der Codiereinheit kann in den Bee-Swarm-Daten bspw. durch mehrmaliges Fokussieren der gleichen Elemente des Stimulus in gleicher Reihenfolge erfolgen. In den Interviewdaten werden sie durch verbale Äußerungen der Probanden markiert (z. B. „Als Erstes habe ich ..., und dann habe ich noch ...“). Die Codierung der Daten erfolgte durch zwei unabhängige Personen. Dabei erwiesen sich insbesondere der Detailgrad der Codiereinheit sowie die Kürze der codierten Segmente als Gefahr für die Reliabilität. Dies führte zu einer Diskussion und Überarbeitung der Codierregeln. Abschließend wurde die Übereinstimmung der beiden Codierer*innen für 50 % der Daten geprüft. Dabei zeigte sich für die Aufgabe *Haus* bzw. die Aufgabe *Zahlenmauer* mit einer prozentualen Übereinstimmung von 72 % bzw. 79 % eine ausreichende Übereinstimmung für diese Art der Codierung (Wirtz & Caspar, 2002).

Nach der Codierung aller Daten erfolgte ein Vergleich der Codierungen der Eye-Tracking-Daten und der Interviewdaten.

3.6 Ergebnisse

Im Folgenden werden die Ergebnisse zweier Fälle zu den Aufgaben *Zahlenfeld* und *Zahlenmauer* dargestellt. Anschließend erfolgt eine zusammenfassende Darstellung der Ergebnisse aller Fälle.

Der Fall Tilda

Tilda durchläuft mit ihren Blickbewegungen nach einer Phase der Orientierung (Abb. 3.2), in der sich bereits erste Anzeichen für ein Erkennen von Abschnitten

Abb. 3.2 Überblick
verschaffen



aus der Reihe der natürlichen Zahlen zeigen, die Zahlenreihe 6 bis 9 (Abb. 3.3). In den nächsten Sekunden betrachtet Tilda die erste und letzte Spalte (Abb. 3.4). Hier fokussiert sie nacheinander die Zahlen von 1 bis 4 in den Ecken des Quadrats. Im Anschluss erfolgt eine Fokussierung der Zahlenreihenabschnitte 1 bis 4 und 6 bis 9 (Abb. 3.5). Dabei springt die Blickbewegung direkt von der 4 auf die 6.

Nach etwa einer halben Minute wendet sich Tilda den beiden mittleren Spalten zu (Abb. 3.6). Auch hier fokussiert sie nacheinander Zahlen in der Reihenfolge der natürlichen Zahlen, jetzt allerdings rückwärts, beginnend mit der Zahl 19. Im Bee-Swarm fällt auf, dass Tilda beim Durchlaufen der Zahlenreihe nach der 16 unmittelbar die 14 fixiert. Zudem betrachtet Tilda jeweils die sowohl in der Tabelle

Abb. 3.3 Zahlenreihe 6 bis 9



Abb. 3.4 Spalten



Abb. 3.5 Zahlenreihe 1 bis 9



Abb. 3.6 Mittlere Spalten



als auch in der Zahlenreihe nebeneinanderliegenden Zahlen (z. B. 16 und 17 sowie 13 und 14) immer wieder abwechselnd. Es zeigt sich also ein strukturiertes Blickverhalten mit Bezug zu den Zahlenreihen 1 bis 9 und 19 bis 11. Sie fokussiert jeweils die Zahlen 4 und 6 sowie 16 und 14 unmittelbar nacheinander, was darauf hindeuten könnte, dass sie erkannt hat, dass sich in dem Zahlenquadrat keine Zahlen aus der Fünferreihe befinden.

In ihrer anschließenden Erklärung beschreibt Tilda ihre Entdeckungen über die Zahlenreihen 1 bis 4 und 6 bis 9 zunächst isoliert und anschließend in Verbindung zueinander, indem sie auch auf das Fehlen der Zahlen 5 und 10 hinweist.

36	T:	[...] ich habe halt erstmal nach der Zahlenreihenfolge eigentlich geschaut. Also eins, zwei, drei, vier (<i>zeigt auf die linke obere, dann auf die rechte obere, anschließend auf die linke untere, dann auf die rechte untere Ecke</i>), und das war schon gut mit den Ecken und dann mit den sechs, sieben, acht, neun (<i>zeigt auf die entsprechenden Zahlen in der Zeichnung</i>). So hier immer so in den Schlangenreihen (<i>zeichnet mit den Fingern nochmal die Reihenfolge 6 bis 8 nach</i>) [...]. Und das Gleiche gab es bei mir auch hier (<i>zeigt auf die zweite und dritte Spalte</i>) mit den so (<i>zeichnet mit den Fingern die Reihenfolge 11 bis 17 nach</i>) und dann wieder hier hoch (<i>geht von der 17 hoch zur 18</i>), wie hier (<i>zeigt mit den Fingern nochmal auf die Reihenfolge 6 bis 8</i>).
37	I:	Also du hast nach der Zahlenreihenfolge geguckt [...] eine Besonderheit bei der Zahlenreihenfolge gibt es aber da, oder?
38	T:	(<i>Geht mit den Fingern nochmal die Zahlen 1 bis 4 nach</i>) Die Fünf und die Zehn ist nicht da.

Anschließend geht Tilda auf die Zahlenreihe 19 bis 11 sowie die unterschiedliche Anordnung der Nachbarzahlen 11 und 12 und 13 und 14 ein.

43	T:	Also wie bei dem (<i>zeigt auf 19, 18, 17, 16</i>), nur andersrum und dass es halt richtig aneinander ist.
----	----	--

Dabei nutzt sie das Wort „richtig“ für die beiden Nachbarzahlen, die in der Tabelle wie in der Reihe der natürlichen Zahlen angeordnet sind.

Zusammenfassend lässt sich für diese Aufgabe festhalten, dass sich die aufgrund der Blickbewegungen aufgestellten Annahmen zu den Muster- und Strukturerkennungsprozessen von Tilda mit ihren anschließenden Erläuterungen zu ihren Entdeckungen decken.

In der Aufgabe *Zahlenmauer* lässt sich beobachten, dass Tilda ausschließlich die beiden oberen Reihen fokussiert (Abb. 3.7 bis Abb. 3.9).

In der ersten Hälfte der Bearbeitungszeit zeigt sich eine Orientierung auf die erste Reihe mit einer starken Fokussierung auf die Dreiergruppe 2–6–4 (Abb. 3.7). Die häufigen Sakkaden zwischen diesen drei Zahlen können darauf hindeuten, dass Tilda (operationale) Beziehungen zwischen diesen Zahlen herstellt. Gleichzeitig finden auch Blickbewegungen zu den Zahlen 3 und 5 und später auch zur 7 statt (Abb. 3.8 und 3.9); möglicherweise orientiert sie sich an der Reihenfolge der natürlichen Zahlen. In der zweiten Hälfte der Bearbeitungszeit fokussiert Tilda ausschließlich die Zahlen 3, 4, 5, 6 und 7. Dabei steuert ihr Blick die Zahlen in der Reihenfolge der natürlichen Zahlen an – wobei sie jeweils mit unterschiedlichen Zahlen beginnt. Diese Interpretation wird durch Tildas Erklärung gestützt, in der sie die Positionen der Zahlen mit Bezug zur Reihe der natürlichen Zahlen erläutert.

38	T:	(lacht) Also ich habe mir irgendwie die beiden Reihen gemerkt (<i>zeigt auf die zweite und erste Reihe</i>). Also habe ich zwei, drei (<i>zeigt auf das linke und rechte obere Kästchen</i>), wie beim letzten Mal, und dann bin ich hierhin gegangen (<i>zeigt auf die 4 im dritten oberen Kästchen</i>), dann wusste ich, wenn ich hierhin (<i>zeigt auf die 5 im ersten Kästchen der zweiten Reihe</i>), und das konnte ich mir auch anders merken, dass die Fünf hier war. Wenn ich fünf, sieben (<i>zeigt auf die 7 neben der 5</i>), und dann dazwischen ist die Sechs (<i>zeigt auf die 6 in der ersten Reihe</i>), und dann habe ich hier (<i>zeigt auf die 8 im letzten Kästchen in der zweiten Reihe</i>) halt noch hingeschaut, ist die Acht, und dann weiß ich halt noch, ob da oder da (<i>zeigt auf das dritte und vierte leere Kästchen der untersten Reihe</i>) die Elf ist. Also ich habe, unten habe ich eigentlich überhaupt nicht geschaut.
----	----	---

Abb. 3.7 Scan-Path-Ausschnitt 1

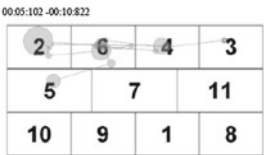


Abb. 3.8 Scan-Path-Ausschnitt 2

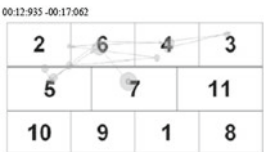
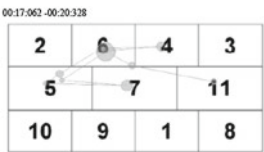


Abb. 3.9 Scan-Path-Ausschnitt 3



Darüber hinaus verweist Tilda auf eine Gemeinsamkeit zur vorherigen Aufgabe: Sie beschreibt, dass die Zahlen 2 und 3, wie auch in Aufgabe 3, in den Ecken notiert sind. Zusammenfassend zeigt sich, dass Tildas Beschreibung sich zu großen Teilen mit den aufgrund der Blickbewegungen getroffenen Annahmen über die erkannten Muster und Strukturen decken. Zusätzlich lassen Tildas Blickbewegungen jedoch vermuten, dass sie weitere Zahlbeziehungen in der obersten Zeile erkannt haben könnte.

Der Fall Aylin

Aylins Blickbewegungen zeigen, dass sie zunächst die erste Spalte und anschließend die erste und zweite Spalte betrachtet (Abb. 3.10). Danach steht die dritte Spalte zunächst isoliert und anschließend in Verbindung mit der vierten Spalte im Fokus (Abb. 3.11). Dies deutet darauf hin, dass Aylin die erste Spalte mit der zweiten sowie die dritte Spalte mit der vierten verknüpft. Im Bee-Swarm zeigt sich, dass bei dieser zweisepaltigen Fokussierung jeweils ausgewählte diagonale Zahlenpaare betrachtet werden, die die Gemeinsamkeit der Zehnerdifferenz aufweisen (9/19, 4/14, 6/16). In der zweiten Hälfte der Betrachtungszeit fixiert sie nur noch die Zahlen der ersten und dritten Spalte, indem diese Spalten immer wieder nacheinander von oben nach unten durchlaufen werden (Abb. 3.12). Dies kann darauf hindeuten, dass Aylin versucht, sich die Zahlen dieser Spalten einzuprägen.

Diese Vermutung wird in ihrer Erklärung bestätigt: Hier gibt Aylin an, dass sie sich „nur“ die erste und dritte Spalte gemerkt hat:

33	A:	Also ich habe nur das hier und das hier mir gemerkt (<i>zeigt auf die erste und dritte Spalte</i>) [...], und dann habe ich gemerkt, dass immer neun und dann hier neunzehn und hier die Eins und hier die Elf und hier die Siebzehn und hier die Sieben (<i>zeigt immer auf die gesagten Zahlen in der Zeichnung</i>). Also das hier und das hier habe ich mir überhaupt nicht gemerkt (<i>zeigt auf die zweite und vierte Spalte</i>). Ich habe einfach herausgefunden, dass ich hier dann eine Zehn dazu mache (<i>zeigt auf die 1 und die 11</i>) und hier auch (<i>zeigt auf die 13</i>), hier auch (<i>zeigt auf die 17</i>), hier auch (<i>zeigt auf die 19</i>). Und hier mache ich die Zehn weg (<i>zeigt zwischen der 12 und 2 hin und her</i>), hier auch (<i>zeigt auf die 18 und die 8</i>), hier auch (<i>zeigt auf die 14 und die 4</i>) und hier auch (<i>zeigt auf die 16 und die 6</i>).	
----	----	--	--

Abb. 3.10 Erste und zweite Spalte



Abb. 3.11 Dritte und vierte Spalte



Abb. 3.12 Fokussierung der ersten und dritten Spalte



Außerdem verweist sie auf die von ihr erkannte Struktur der diagonalen Zahlbeziehung und erläutert dabei auch, dass sie sich die Zahlen in den beiden anderen Spalten über die erkannte Zahlbeziehung ± 10 herleiten konnte. Sowohl die Erklärungen zu den erkannten Strukturen als auch zum „Auswendiglernen“ decken sich mit den aufgrund der Blickbewegungen aufgestellten Annahmen zu den Muster- und Strukturerkennungsprozessen von Aylin.

Bei der vierten Aufgabe lassen sich in Aylins Blickbewegungen insbesondere ein zeilenweises Fokussieren sowie ein Fokussieren der ersten Spalte erkennen (Abb. 3.13 bis 3.15).

In der ersten Sequenz werden die Zahlen der ersten Reihe, ergänzt durch die 5 aus der zweiten Reihe, in den Blick genommen (Abb. 3.13). Im folgenden Verlauf zeigt sich die Fokussierung mehrerer Dreiergruppen 5–7–11, 2–5–10 sowie 11–10–9 (ggf. ergänzt durch eine abschließende kurze Fokussierung der 1)

Abb. 3.13 Scan-Path-Ausschnitt 1



Abb. 3.14 Scan-Path-Ausschnitt 2

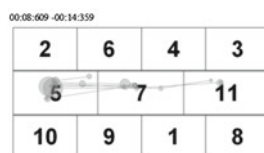


Abb. 3.15 Scan-Path-
Ausschnitt 3



(Abb. 3.14 und 3.15). Dadurch lässt sich annehmen, dass sie vorwiegend zwischen den Zahlen einer Reihe sowie den Zahlen der ersten Spalte Beziehungen herstellt.

45	I:	Aber warum hast du jetzt hier angefangen (<i>zeigt auf die 5</i>)? Wenn ich ein Buch lese, fange ich immer da oben an.
46	A:	Ja, weil hier habe ich das zusammengemacht, also fünf plus sieben ist gleich zwölf, und dann habe ich mir die Zwölf gemerkt, und dann war da halt nur die Elf (<i>zeigt auf die 11 im letzten Kästchen der zweiten Spalte</i>).
47	I:	Okay, und wie konntest du dir das unten merken?
48	A	Weil zehn, neun, und dann kommt die Eins und dann erst eine Acht.
49	I:	Okay, und was hat das miteinander zu tun?
50	A:	Also ich meine, zehn, neun, acht, sieben, und so nur nach der Neun kommt eine Eins

In Aylins Begründung wird die Vermutung der Fokussierung auf die Dreiergruppe 5–7–11 aufgrund eines Rückgriffs auf eine operationale Beziehung deutlich. Aylin hat diese drei Zahlen über die Rechnung $5 + 7 = 12$ miteinander verknüpft und sich ergänzend gemerkt, dass statt der 12 eine 11 eingetragen ist. In der letzten Reihe gibt sie an, dass sie sich die Zahlen über eine Verknüpfung mit der rückwärts-gewandten Zahlenreihe gemerkt hat, und verweist hier nicht auf eine Verknüpfung mit der 11, sondern stattdessen auf die 8, die sich in den Eye-Tracking-Daten nicht zeigt. Auf die in den Eye-Tracking-Daten codierten Dreiergruppen 2–5–10 sowie 11–10–9 geht Aylin nicht ein. Zu den hier vermuteten Strukturierungen lassen sich durch die Verbalisierung keine Bezüge herstellen.

Einordnung der Ergebnisse aller Fälle

Am Beispiel der beiden Fälle Aylin und Tilda lässt sich eine hohe Übereinstimmung zwischen den auf Grundlage der Eye-Tracking-Daten (Bee-Swarm) getroffenen Annahmen und den verbalen Beschreibungen der Kinder erkennen – dies bestätigt sich in den übrigen Aufgaben und Fällen. Zusätzlich zeigen sich Abweichungen zwischen den aufgrund der unterschiedlichen Methoden gebildeten Annahmen zum Prozess der Muster- und Strukturerkennung. Vereinzelt wurden die Muster und Strukturen, die von den Kindern in die Darstellungen hineinge-deutet wurden, nicht in den Eye-Tracking-Daten erkannt – wie bei Aylin und der Aufgabe *Zahlenmauer* (Verknüpfung der Zahlen 5, 7 und 11) dargestellt.

Dies zeigt sich insbesondere bei Aufgaben mit vergleichsweise wenig Struktur (*Zahlenmauer*, *Haus*). So berichten alle Kinder – bis auf Fall 6 und 8 –, in der Aufgabe *Haus* zunächst die Figur als Ganzes und anschließend Fenster und eine Tür gesehen zu haben. In den Blickbewegungen lassen sich jedoch kaum Fixationen auf den Außenlinien erkennen, sondern überwiegend auf den inneren Teilfiguren. In den strukturstärkeren Aufgaben zeigen sich dagegen höhere Übereinstimmungen. So beziehen sich in der Aufgabe *Zahlenfeld* neun der zehn Kinder im Interview auf ausgewählte Spalten; in den Eye-Tracking-Daten spiegelt sich diese spaltenweise Betrachtung außer bei den Fällen 5 und 10 wider. In der Aufgabe *Geometrische Musterfolge* argumentieren alle Kinder im Interview über markante Musterelemente. Hier finden sich – bis auf Fall 6 und 7 – die entsprechenden Fixationen auch in den Blickbewegungen.

Zusätzlich lassen sich sowohl bei den Aufgaben mit starker als auch bei Aufgaben mit weniger Struktur aufgrund der Blickbewegungen weitere Muster- und Strukturerkennungsprozesse vermuten, auf die in den verbalen Beschreibungen kein Bezug genommen wird – wie beispielsweise bei Tildas Bearbeitung der zweiten Aufgabe. Weitere Beispiele dieser Art finden sich bei dieser Aufgabe für die Fälle 1, 8, 9 und 10. Insgesamt lassen diese Beobachtungen annehmen, dass die verbalen Beschreibungen der Kinder eher spiegeln, welche Muster und Strukturen final in die Darstellungen hineingedeutet wurden, und sich weniger am Wahrnehmungsprozess orientieren.

3.7 Diskussion

Limitationen

Mit dem gewählten explorativen Forschungsansatz gehen verschiedene Limitationen einher, da potenziell beeinflussende Variablen nicht kontrolliert werden. Beispielsweise sind die Beschreibungen der Kinder insgesamt durch hohes sprachliches Niveau gekennzeichnet. Unabhängig von der mathematischen Begabung der Kinder könnte auch die hohe Verbalisierungsfähigkeit ursächlich für die Übereinstimmung zwischen den Annahmen über die Muster- und Strukturerkennung auf Grundlage der Eye-Tracking-Daten und den verbalen Beschreibungen der Kinder sein. Zusätzlich liegen für die Kinder zwar Indizien einer mathematischen Begabung, jedoch keine formalen Nachweise vor. Die Leistungen der Kinder könnten ebenso durch bspw. hohe Raumvorstellungsfähigkeiten bedingt sein. Mit Blick auf die vorrangig methodischen Fragestellungen scheinen diese Einschränkungen jedoch weniger zentral.

Eine weitere, zentrale Limitation ergibt sich aus dem Interviewleitfaden. Als Hauptimpuls zur verbalen Erklärung wurde die Frage „Wie hast du dir das so schnell gemerkt?“ gewählt. Dieser Impuls hat die Kinder zwar angeregt, ihr Vorgehen detailliert zu beschreiben, möglicherweise legt dieser Frageimpuls jedoch nahe, ausschließlich Muster und Strukturen zu beschreiben, die sich im Merkvorgang als hilfreich erwiesen haben. Da jedoch auch anschließende Aufforderungen,

weitere bemerkenswerte Aspekte der Aufgabe zu benennen, nicht genutzt wurden, um weitere Muster und Strukturen zu benennen, bleibt an dieser Stelle offen, welcher Impuls geeignet wäre, um detaillierte Beschreibungen aller erkannten Muster und Strukturen hervorzurufen.

Ergebnisdiskussion

Unter Berücksichtigung der Limitationen zeigt sich in dieser Studie das Potenzial des Eye-Trackings zur Ergänzung bisheriger Methoden zur Erfassung der Muster- und Strukturerkennung (mathematisch begabter) Schüler*innen – wie bereits in ähnlichen Studien zu anderen Kontexten (z. B. Schindler et al., 2019; Sprenger & Benz, 2020) und mit Kindern mit mathematischen Schwierigkeiten (Simon & Schindler, 2020). Beispielsweise konnte bei allen potenziell mathematisch begabten Grundschüler*innen aufgezeigt werden, welche Teile der Stimuli in welcher Reihenfolge betrachtet wurden, sodass Annahmen über erkannte Muster und Strukturen aufgestellt werden konnten. Dabei ergaben sich Unterschiede mit Bezug auf die Aufgabenstellung: Während sich bei Aufgaben mit starker Strukturierung – trotz der hohen metakognitiven und sprachlichen Ansprüche (s. auch Peter-Koop, 2002) – hohe Übereinstimmungen zwischen den Annahmen zur Muster- und Strukturerkennung auf Grundlage der verbalen Beschreibungen und den diesbezüglich aus den Eye-Tracking-Daten abgeleiteten Annahmen zeigten, ergaben sich bei Aufgaben mit weniger Strukturgehalt Unterschiede. Möglich wäre, dass die Schüler*innen nur hilfreiche Muster und Strukturen verbal erklären oder dass nicht alle erkannten Muster und Strukturen bewusst zur Rekonstruktion der Darstellung genutzt werden. Zusätzlich ist nicht auszuschließen, dass Blickbewegungen, die als Muster- und Strukturerkennung gedeutet werden, eher Suchprozesse und keine Muster- und Strukturerkennung darstellen – also, dass die *eye-mind assumption* hier nicht zutrifft (s. auch Schindler & Lilienthal, 2019).

Gleichzeitig zeigen sich in dieser Studie Muster- und Strukturerkennungsprozesse, die nicht allein aufgrund der Blickbewegungen erkannt werden (können). So gibt es verbale Erklärungen, die keinen Bezug zu den Blickbewegungen aufweisen (z. B. Superzeichen Haus). Zusätzlich geben die verbalen Erklärungen Einblicke, wie die Kinder wichtige Informationen der Muster verbinden (Abschn. 3.3), und zu gleichen Blickbewegungen liefern Kinder unterschiedliche verbale Erklärungen. Diese deuten auf die starke Individualität der Muster- und Strukturerkennungsprozesse hin (s. auch Steinweg, 2013) und sind ein Hinweis auf die Validität und den Nutzen der retrospektiven Interviews.

Insgesamt stützt diese Studie die Annahmen verschiedener Autor*innen (Abschn. 3.3), dass mathematisch begabte Schüler*innen besondere Fähigkeiten zur Muster- und Strukturerkennung mitbringen: Alle Schüler*innen erkennen Muster. Auf Grundlage des Eye-Trackings konnten dabei insbesondere bei Aufgaben mit weniger Strukturgehalt Blickbewegungen ausgemacht werden, die als Suche nach Mustern und Strukturen gedeutet werden können (s. auch Käpnick, 1998, 2008). Unklar ist jedoch, inwieweit diese Suchprozesse Merkmal

mathematischer Begabung sind oder auch in Blickbewegungen von Kindern ohne mathematische Begabung erkennbar werden.

Zusammenfassend zeigt diese Studie das hohe Potenzial der Methode des Eye-Trackings zur Erfassung von Muster- und Strukturerkennungsprozessen – insbesondere in Kombination mit verbalen Erklärungen (s. auch Simon & Schindler, 2020). Zusätzlich stützt diese Studie für die hier untersuchte Stichprobe potenziell mathematisch begabter Grundschüler*innen die Validität von verbalen Erklärungen im Kontext der Muster- und Strukturerkennung – zumindest für Aufgaben mit starker Strukturierung. Weiterer Forschungsbedarf besteht für andere Stichproben sowie für Aufgaben mit weniger Strukturgehalt.

Literatur

- Aßmus, D. (2017). *Mathematische Begabung im frühen Grundschulalter unter besonderer Berücksichtigung kognitiver Merkmale*. WTM.
- Devlin, K. (1998). *Muster der Mathematik: Ordnungsgesetze des Geistes und der Natur*. Spektrum, Akademischer Verlag.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer.
- Fyfe, E. R., Evans, J. L., Matz, L. E., Hunt, K. M., & Alibali, M. W. (2017). Relations between patterning skill and differing aspects of early mathematics knowledge. *Cognitive Development*, 44, 1–11. <https://doi.org/10.1016/j.cogdev.2017.07.003>
- Goldstein, E. B. (2008). *Wahrnehmungspsychologie*. Spektrum.
- Guski, R. (2000). *Wahrnehmung: eine Einführung in die Psychologie der menschlichen Informationsaufnahme*. Kohlhammer.
- Heinze, A. (2005). *Lösungsverhalten mathematisch begabter Grundschulkinder – aufgezeigt an ausgewählten Problemstellungen*. Lit.
- Helfferich, C. (2014). Leitfaden- und Experteninterviews. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 559–574). Springer VS. https://doi.org/10.1007/978-3-531-18939-0_39
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford Univ. Press.
- Just, M. & Carpenter, P. (1976). Eye fixation and cognitive processes. *Cognitive Psychology*, 8(4), 441–480. [https://doi.org/10.1016/0010-0285\(76\)90015-3](https://doi.org/10.1016/0010-0285(76)90015-3).
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>
- Käpnick, F. (1998). *Mathematisch begabte Kinder. Modelle, empirische Studien und Förderungsprojekte für das Grundschulalter*. Lang.
- Käpnick, F. (2008). Diagnose und Förderung mathematisch begabter Kinder im Spannungsfeld zwischen interdisziplinärer Komplexität und Bereichsspezifität. In C. Fischer, F. J. Mönks, & U. Westphal (Hrsg.), *Individuelle Förderung: Begabungen entfalten – Persönlichkeit entwickeln* (S. 3–23). LIT.
- Kießwetter, K. (1993). Vernetzung als unverzichtbare Leitidee für den Mathematikunterricht. *mathematik lehren*, 58, 5–7.
- Krutetskii, V. A. (1976). *The psychology of mathematical abilities in schoolchildren*. University of Chicago Press.
- Lüken, M. M. (2012). Young children's structure sense. *Journal Für Mathematik-Didaktik*, 33(2), 263–285. <https://doi.org/10.1007/s13138-012-0036-8>

- Lüken, M. M., & Sauzet, O. (2020). Patterning strategies in early childhood: A mixed methods study examining 3- to 5-year-old children's patterning competencies. *Mathematical Thinking and Learning*, 22(3), 1–21. <https://doi.org/10.1080/10986065.2020.1719452>
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz.
- Mulligan, J., & Mitchelmore, M. (2009). Awareness of pattern and structure in early mathematical development. *Mathematics Education Research Journal*, 21(2), 33–49.
- Nolte, M., & Kießwetter, K. (1996). Können und sollen mathematisch besonders befähigte Schüler schon in der Grundschule identifiziert und gefördert werden? *ZDM*, 5, 143–157.
- Peter-Koop, A. (2002). Kommunikation von Denk- und Lösungsstrategien: Aufgaben für Forschung und Förderung. In A. Peter-Koop & P. Sorger (Hrsg.), *Mathematisch besonders begabte Grundschulkinder als schulische Herausforderung* (S. 142–149). Mildenerger.
- Rittle-Johnson, B., Zippert, E. L., & Boice, K. L. (2019). The roles of patterning and spatial skills in early mathematics development. *Early Childhood Research Quarterly*, 46, 166–178. <https://doi.org/10.1016/j.ecresq.2018.03.006>
- Sawyer, W. W. (1955). *Prelude to mathematics*. Penguin.
- Schindler, M., & Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: Towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics*, 101(1), 123–139. <https://doi.org/10.1007/s10649-019-9878-z>.
- Schindler, M., Lilienthal, A. J., Schindler, F., & Schabmann, A. (2019). Quantity recognition in structured whole number representations of students with mathematical difficulties: An eye tracking study. *Learning Disabilities: A Contemporary Journal*, 17(1), 5–28.
- Schmerold, K., Bock, A. M., Peterson, M., Leaf, B., Vennergrund, K., & Pasnak, R. (2017). The relations between patterning, executive function, and mathematics. *The Journal of Psychology*, 151(2), 207–228. <https://doi.org/10.1080/00223980.2016.1252708>
- Simon, A. L., & Schindler, M. (2020). A comparative analysis of eye tracking and thinking aloud in number line estimation tasks: A study on students with and without mathematical difficulties. In M. Inprasitha, N. Changsri, & N. Boonsena (Hrsg.), *Interim proceedings of the 44th conference of the international group for the psychology of mathematics education* (S. 537–545). PME.
- Singer, F., Sheffield, L., Freiman, V., & Brandl, M. (2016). *Research on and activities for mathematically gifted students*. Springer. https://doi.org/10.1007/978-3-319-39450-3_1
- Sprenger, P., & Benz, C. (2020). Children's perception of structures when determining cardinality of sets – Results of an eye-tracking study with 5-year-old children. *ZDM*, 52, 753–765. <https://doi.org/10.1007/s11858-020-01137-x>
- Steinweg, A. S. (2013). *Algebra in der Grundschule. Muster und Strukturen – Gleichungen – funktionale Beziehungen*. Springer.
- Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, 104(2), 147–200. <https://doi.org/10.1007/s10649-020-09948-1>
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Hogrefe.
- Yin, R. K. (2018). *Case study research and applications: Design and methods*. SAGE.



Eye-Tracking als Methode zur Analyse der Wirkung unterstützender Hinweise in digitalen Lernumgebungen

Christoph Hoyer, Bianca Watzka, Maleen Hurzlmeier, Bernhard Ertl und Raimund Girwidz

Inhaltsverzeichnis

4.1	Einleitung und Forschungslage	60
4.2	Theoretischer Hintergrund	60
4.3	Beschreibung der Experimente	62
4.4	Diskussion	70
4.5	Resümee	72
	Literatur	72

C. Hoyer (✉) · R. Girwidz

Lehrstuhl für Didaktik der Physik, Ludwig-Maximilians-Universität München, München, Deutschland

E-Mail: christoph.hoyer@lmu.de

URL: https://www.didaktik.physik.uni-muenchen.de/die_arbeitsgruppe/personen/hoyer/index.html

R. Girwidz

E-Mail: girwidz@lmu.de

M. Hurzlmeier · B. Ertl

Erziehungswissenschaft mit dem Schwerpunkt Lernen und Lehren mit Medien, Universität der Bundeswehr München, München, Deutschland

E-Mail: maleen.hurzlmeier@unibw.de

B. Ertl

E-Mail: bernhard.ertl@unibw.de

B. Watzka

Didaktik der Physik, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Deutschland

E-Mail: bianca.watzka@ovgu.de

4.1 Einleitung und Forschungslage

Der Einsatz von Hinweisen (engl. *cueings*, *signalings*) in Multimedia-Anwendungen kann die Aufmerksamkeit lenken oder inhaltliche Aspekte akzentuieren. Eye-Tracking bietet Möglichkeiten, die Wirkung von Hinweisen auf die visuelle Aufmerksamkeit zu untersuchen. Dies wird anhand von drei unterschiedlichen Beispielen aus der Physik aufgezeigt. In diesem Kapitel wird deutlich gemacht, wie verschiedene Eye-Tracking-Metriken dabei helfen können, die Wirkung von unterschiedlichen unterstützenden Hinweisen zu analysieren.

- Im ersten Experiment wird analysiert, inwieweit sich Unterschiede im Vorwissen beim Lernen mit visuellen Hinweisen im Blickverhalten widerspiegeln. Hierzu werden die Blickpfade qualitativ ausgewertet.
- Im zweiten Experiment wird untersucht, welchen Einfluss Hinweise auf das Lernen mit Experimentiervideos und die visuelle Aufmerksamkeit der Lernenden haben. Dabei werden auch Modalitätseffekte berücksichtigt. Hier werden die Anzahl der Fixationen und die Blickdauer näher betrachtet.
- Im dritten Experiment werden verschiedene Hinweise zur Dokumentation von Messwerten in Diagrammen miteinander verglichen. Zur Interpretation der Ergebnisse werden Blinksakkaden in Abhängigkeit von ihrer Richtung, Länge und Häufigkeit analysiert.

4.2 Theoretischer Hintergrund

Bekannt ist, dass die kognitiven Verarbeitungsprozesse *Selektieren*, *Organisieren* und *Integrieren* durch aufmerksamkeitslenkende bzw. informierende Hinweise unterstützt werden können (vgl. van Gog, 2014; Richter et al., 2015; Schneider et al., 2018). Dabei spielen u. a. folgende Faktoren eine Rolle für die Wirkung von Hinweisen: Vorwissen, Präsentationsform des Lerninhalts (dynamisch vs. statisch) sowie Art und fachlicher Inhalt der Hinweise. Diese Faktoren sollen im Folgenden genauer betrachtet werden.

Vorwissen

Die Metastudie von Richter et al. (2016) untersuchte den Einfluss von relationalen Hinweisen zur Förderung der Integration von verbal und bildhaft präsentierten Informationen. Die Ergebnisse zeigen, dass diese Hinweise zum Verständnis beitragen, indem sie die Integration der Informationen anregen, vor allem bei Lernenden mit niedrigem Vorwissen.

Nach der Metaanalyse von Schneider et al. (2018) lassen sich jedoch diese Ergebnisse nicht verallgemeinern. Die Autoren zeigen einen positiven Effekt hohen Vorwissens auf die Wirkung der Signalisierung. Im Einzelfall sollte daher

überprüft werden, inwieweit sich die gegebenen Hinweise unter Berücksichtigung des Vorwissens auf den Lernerfolg auswirken. Um Unterschiede in der visuellen Wahrnehmung aufzudecken, können Eye-Tracking-Daten genutzt werden. Experiment 1 beschreibt ein mögliches qualitatives Vorgehen. Unterschiede im Blickverhalten zwischen Personen mit hohem und niedrigem Vorwissen beim Umgang mit visuellen Hinweisen zu elektrischen Schaltungen werden aufgezeigt.

Präsentationsform des Lerninhalts (dynamisch vs. statisch)

Lerninhalte können durch statische oder dynamische Repräsentationen dargeboten werden. Nach Höffler und Leutner (2007) sind dynamische Repräsentationen vor allem dann von Vorteil, wenn die Dynamik explizit zum Lerninhalt gehört. Berney und Bétrancourt (2016) konnten hingegen einen solchen Zusammenhang nicht nachweisen.

Untersuchungen zeigen, dass sich die Wirkung von Hinweisen in statischen Repräsentationen von der in dynamischen Repräsentationen unterscheidet (vgl. de Koning et al., 2009). So konnte durch Hinweise in Illustrationen und Texten beispielsweise die Selektion von Inhalten unterstützt werden, während gleichzeitig die Erinnerungsleistung an hervorgehobene Elemente sowie die Fähigkeit der Lernenden zum Problemlösen erhöht wurden. Für Animationen ließ sich eine lernförderliche Wirkung hingegen nicht immer nachweisen (z. B. de Koning et al., 2010; Kriz & Hegarty, 2007). Es ist daher nicht zwangsläufig zu erwarten, dass Hinweise, die sich in statischen Lernmaterialien als effektiv erwiesen, dieselbe positive Wirkung in dynamischen Lernmaterialien zeigen.

Bei der Beurteilung der Wirksamkeit von Hinweisen sollten weiterhin nicht nur die Lernergebnisse im Blickpunkt stehen, sondern auch die Lernprozesse, die zu diesen Ergebnissen führten. Durch Eye-Tracking können Aussagen über die visuelle Aufmerksamkeit der Lernenden bei der hinweisgestützten Bearbeitung von statischen und dynamischen Lernmaterialien getroffen werden.

In den Experimenten 2 und 3 werden entsprechende Vorgehensweisen zur Beurteilung der Wirkungen von Hinweisen in physikalischen dynamischen bzw. statischen Repräsentationen aufgezeigt.

Art und fachlicher Inhalt der Hinweise

Hinweise können genutzt werden, um auf Elemente des Lernmaterials aufmerksam zu machen. Ein Lenken der visuellen Aufmerksamkeit gelingt z. B. über Spotlights, d. h. durch hell hervorgehobene Bildelemente (z. B. de Koning et al., 2010; Jarodzka et al., 2013), oder Pfeile (z. B. Kriz & Hegarty, 2007).

Neben dieser Akzentuierung von Elementen der visuellen Oberflächenstruktur können durch Hinweise auch Elemente der Tiefenstruktur des Lerninhalts hervorgehoben werden. So kann auch die Verknüpfung von Elementen innerhalb einer Darstellung oder repräsentationsübergreifend über mehrere Darstellungen hinweg

unterstützt werden. Zur Analyse, wie integrative Hinweise verarbeitet werden, kann eine Betrachtung von Blicksakkaden hilfreich sein. Während in manchen Untersuchungen eine erhöhte Anzahl an Sakkaden in Verbindung mit besseren Ergebnissen standen (Johnson & Mayer, 2012; Klein et al., 2019; Mason et al., 2013; O’Keefe et al., 2014), konnten andere Untersuchungen einen solchen Zusammenhang nicht nachweisen (Ott et al., 2018; Scheiter & Eitel, 2015; Schüler, 2017).

Hinweise können auch fachliche Informationen enthalten (z. B. durch textuelle oder gesprochene Erläuterungen) oder Hilfen zur kognitiven Verarbeitung geben, beispielsweise durch inhaltliche Anmerkungen zu einer Konvention oder zur Symbolik.

Bei einem überwiegend bildbasierten Lernmaterial haben vor allem Hinweise in Form gesprochener Texte positive Effekte auf den Lernerfolg (z. B. Glaser & Schwan, 2015). Eine Erklärung für die bessere Eignung gesprochener Texte im Vergleich zu geschriebenen gibt die Kognitive Theorie des Multimedialen Lernens (CTML: Mayer, 2001). Aufgrund der begrenzten Kapazität der Verarbeitungskanäle ist es demnach sinnvoll, Informationen auf den visuellen und den auditiven Verarbeitungskanal aufzuteilen.

Ergebnisse von Xie et al. (2019) zeigen, dass Kombinationen aus visuellen und auditiven Hinweisen den Lernerfolg erhöhen können. Die Autoren weisen dabei einen Zusammenhang moderater Effektstärke zwischen der Informationssuche und der Lernleistung nach. Eine kürzere Suchzeit nach relevanten Informationen ist verknüpft mit einer besseren Behaltens- und Transferleistung. Diese beiden Größen korrelierten weiterhin positiv mit der Dauer der Fixationen und der Anzahl der Fixationen auf lernrelevante Inhalte.

Dies zeigt, dass Eye-Tracking-Metriken die Wirksamkeit von Hinweisen indizieren können. In den Experimenten 2 und 3 wird die Wirksamkeit von Hinweisen in dynamischen und statischen physikalischen Repräsentationen untersucht. Dabei wird beschrieben, wie Fixationen (Experiment 2) und Sakkaden (Experiment 3) im Blickverhalten der Probandinnen und Probanden dazu genutzt werden können, die Lernergebnisse zu interpretieren.

4.3 Beschreibung der Experimente

Experiment 1: Unterschiede im Blickverhalten bei hohem bzw. niedrigem Vorwissen

In Experiment 1 wurde eine qualitative Analyse des Blickverhaltens beim Lernen mit visuellen Hinweisen zu elektrischen Schaltskizzen in Abhängigkeit vom Vorwissen durchgeführt. Dabei dienten die Hinweise dazu, die Integration zwischen den textuellen Erklärungen und den Elementen der Schaltskizzen zu unterstützen. Das resultierende Blickverhalten von Lernenden mit hohem Vorwissen wurde dem Blickverhalten von Lernenden mit niedrigem Vorwissen gegenübergestellt. Der Mehrwert qualitativer Analysen von Blickbewegungen soll dadurch aufgezeigt werden.

Forschungsfrage Inwieweit spiegeln sich Unterschiede im Vorwissen beim Lernen mit visuellen Hinweisen in den Blickpfaden wider?

Stichprobe In der betrachteten Stichprobe lernten 23 Probandinnen und Probanden ($n = 6$ weiblich, $n = 17$ männlich) mit einer Lernumgebung zu elektrischen Schaltungen. Das durchschnittliche Alter lag bei $M = 22.34$ Jahren ($SD = 2.55$). Die Teilnehmerinnen und Teilnehmer stammten fast alle aus dem bildungswissenschaftlichen Bereich, lediglich eine Person war einem technischen Bereich zuzuordnen.

Material und Methode Nach der Erhebung des Vorwissens folgten zwei Lernphasen zu elektrischen Schaltungen. In der ersten Lernphase wurden Bauteile und Schaltsymbole mit den entsprechenden Fachtermini (z. B. elektrischer Widerstand oder Spannungsmessgerät) bildgestützt präsentiert. In der zweiten Lernphase wurden acht Lernvideos mit jeweils unterschiedlichen Lernaufgaben gezeigt. Diese bezogen sich auf das Identifizieren einer physikalisch identischen Schaltung, auf das „Übersetzen“ einer Schaltskizze in eine andere Darstellungsform oder auf das Erkennen physikalisch relevanter Strukturen wie Knotenpunkte und Maschen (Abb. 4.1). In dieser Lernphase waren visuelle Hinweise (Pfeile der Textfelder) im Lernmaterial implementiert. Diese führten durch die einzelnen Schaltungen und verknüpften die in Textform beschriebenen Elemente mit den in den Skizzen enthaltenen Bauteilen, Schaltungstypen und etwaigen Verzweigungen. Die Hinweise hatten somit zweierlei Aufgaben: Zum einen sollten sie die Aufmerksamkeit der Lernenden auf relevante Elemente lenken, zum anderen sollten sie inhaltlich wichtige Elemente hervorheben.

Erhebungsinstrumente Erhebungsinstrumente waren ein Eye-Tracker der Marke SMI-RED (Teltow, Germany) sowie ein Vorwissenstest. Die hinweisbegleitete Lernumgebung (zweite Lernphase) wurde durch das SMI Experiment CenterTM präsentiert. Der verwendete Bildschirm hatte eine Größe von 15 Zoll und eine Auflösung von 1920×1080 Pixeln. Die Distanz zwischen Lernenden und Bildschirm betrug in etwa 60cm. Das Blickverhalten wurde mit einer Abtastfrequenz von 60 Hz und einer Genauigkeit von $< 0.4^\circ$ gemessen.

Das Vorwissen der Teilnehmenden wurde durch die Abfrage von grundlegenden Symbolen elektrischer Schaltskizzen (z. B. Schalter oder elektrischer Widerstand) erfasst. Insgesamt waren 6 Vorwissenspunkte möglich, von denen durchschnittlich $M = 3.87$ ($SD = 2.18$) Punkte erreicht wurden.

Datenauswertung Die aufgezeichneten Blickbewegungsdaten wurden mit der Software BeGaze Version 3.7 ausgewertet. Für die qualitative Analyse der Blickbewegungsdaten wurden für jede Person die Blickpfade untersucht. Die Datenauswertung erfolgte anhand einer deskriptiven Analyse von erkennbaren Besonderheiten im Blickverhalten. Dabei wurden ausschließlich Lernende der Extremgruppen im Vorwissen untersucht. Personen mit 0–1 Vorwissenspunkten wurden der Kategorie „niedriges Vorwissen“ ($n = 5$) und Personen

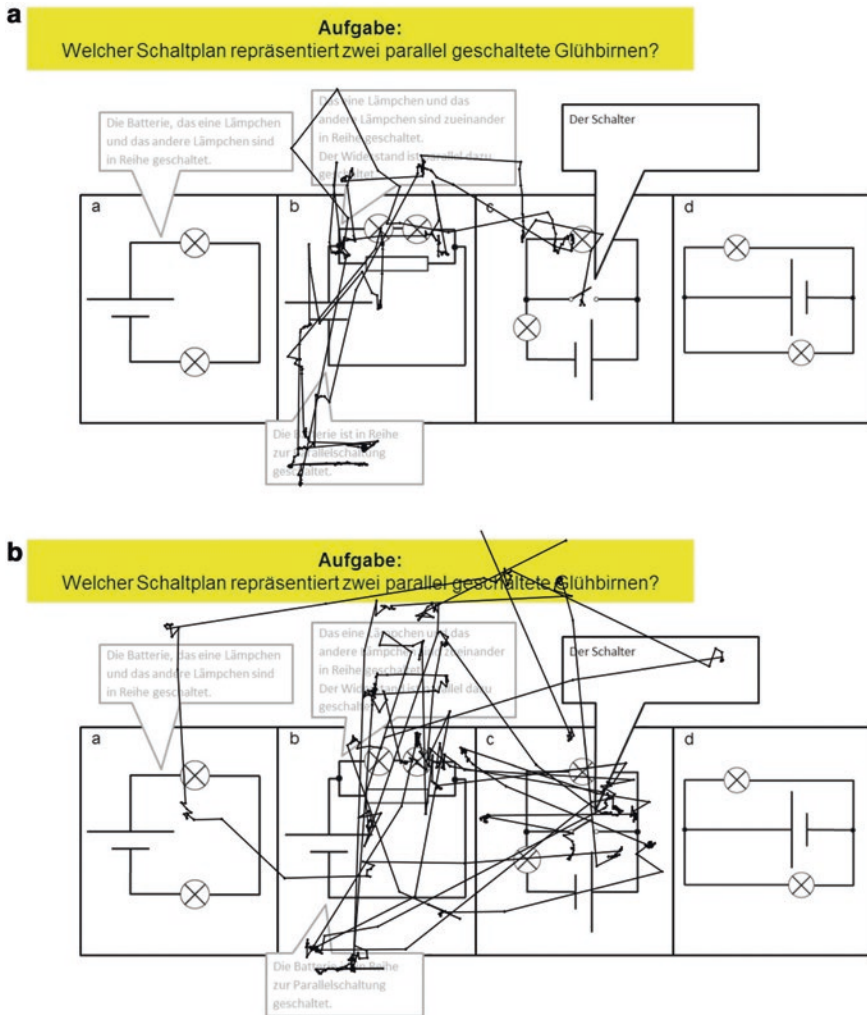


Abb. 4.1 Darstellung von Blickpfaden (Zeitspanne: 10 s) während der Gabe von Hinweisen. **a** Detail- bzw. elementorientiertes Blickverhalten bei niedrigem Vorwissen. **b** Ganzheitliches bzw. schaltungsorientiertes Blickverhalten bei hohem Vorwissen

mit 6 Vorwissenspunkten der Kategorie „hohes Vorwissen“ zugeordnet ($n = 9$). Eine derartige Unterteilung maximiert die Unterschiede und kann dadurch charakteristische Muster der Blickpfade besonders deutlich herausstellen. Die aktuellen Auswertungen beziehen sich auf 10-s-Intervalle im zweiten Drittel der Aufgabenbearbeitung. Ein 10-s-Ausschnitt wurde gewählt, um die Blickpfade nachvollziehbar und übersichtlich halten zu können; der gewählte Zeitpunkt umfasst Hinweisgaben zu Schaltung b und c (Abb. 4.1), sodass anfängliche

Orientierungsprozesse nicht mehr erfasst werden und primär die Wirkung der Hinweise betrachtet werden kann.

Ergebnisse von Experiment 1 Abb. 4.1a stellt das Blickverhalten einer Person mit niedrigem Vorwissen dar. Es traten überwiegend Blickpfade zwischen den textuellen Erklärungen und den dadurch adressierten Elementen der Schaltungen auf. Die Blickpfade orientieren sich deutlich an den Hinweisen. Weiterhin zeigen sich vergleichende Blickbewegungen zwischen den einzelnen Elementen innerhalb von Schaltung b, bevor der Blickpfad dem erscheinenden Hinweis auf Schaltung c folgt. Abb. 4.1b stellt die Blickpfade einer Person mit hohem Vorwissen dar und zeigt ein deutlich anderes Blickverhalten. Es traten ebenfalls Blickpfade zwischen den textuellen Erklärungen und den Elementen der Schaltung auf – allerdings ist die Wirkung der Hinweise weitaus weniger offensichtlich. So zeigt Abb. 4.1b verschiedene Blickpfade, die auch auf Vergleiche zwischen den Schaltungen a, b und c sowie auf Vergleiche zwischen der Aufgabenstellung (gelber Block) und den Schaltungen hinweisen. Personen mit niedrigem Vorwissen zeigten also ein detail- bzw. elementorientiertes Blickverhalten, das auf differenzierte Übersetzungsprozesse zwischen den textuellen Erklärungen und der grafisch repräsentierten Schaltung schließen lässt. Bei Personen mit hohem Vorwissen sind solche Übersetzungsprozesse in deutlich geringerem Umfang nachvollziehbar. Die Vergleiche zwischen den einzelnen Schaltungen sowie mit der Aufgabenstellung wiesen darüber hinaus auf ein eher ganzheitliches bzw. schaltungsorientiertes Blickverhalten hin (weitere Ergebnisse siehe Hurzlmeier et al., 2021).

Experiment 2: Wirkung von Hinweisen in dynamischen Repräsentationen

Neben dem Vorwissen ist die Dynamik der zu verarbeitenden Repräsentation ein Faktor, der die Wirkung von Hinweisen auf das Lernen beeinflussen kann. Eine Analyse einzelner Blickpfade wird vor allem bei dynamischen Lernmaterialien und hohen Teilnehmendenzahlen sehr aufwendig. Es bietet sich daher an, Eye-Tracking-Metriken, wie die Blickdauer etc., zu nutzen. Experiment 2 untersuchte für das Lernen mit Experimentiervideos, in welchem Ausmaß Eye-Tracking-Metriken Unterschiede in den Lernergebnissen aufklären, wenn die Aufmerksamkeit durch visuelle bzw. auditive Hinweise gelenkt wird.

Forschungsfrage Inwieweit erklären Eye-Tracking-Metriken und die Hinweismodalität Unterschiede im Lernergebnis beim Lernen mit Experimentiervideos?

Stichprobe 44 Studierende für das Lehramt Physik nahmen am Experiment teil. Das Durchschnittsalter der Teilnehmenden lag zwischen 21 und 30 Jahren. Die Studie war als Laborstudie umgesetzt und weist die beiden Bedingungen visuelle Hinweise bzw. auditive Hinweise auf. Die Teilnehmenden wurden zufällig auf die beiden Bedingungen verteilt.

Material und Methode Die Lernmaterialien bestanden aus Videos, also dynamischen Repräsentationen. Diese zeigten den Aufbau und die Durchführung von zwei Standardexperimenten zum Thema *Wirbelströme* (Scheibenbremse und Waltenhofen-Pendel). Mit dem Lernmaterial wurde der Erwerb von deklarativem Wissen über den Aufbau und die Durchführung der gezeigten Experimente adressiert.

Die Videos variierten hinsichtlich der Hinweismodalität. In der Variante mit visuellen Hinweisen erschienen in den Videos synchron zur Handlung erklärende Textfelder und Hell-Dunkel-Kontraste zum jeweils relevanten Bildbereich. In den Videos mit auditiven Hinweisen wurden dieselben inhaltlichen Informationen und aufmerksamkeitssteuernden Anweisungen angeboten wie in den Videos der visuellen Bedingung. Allerdings wurden die auditiven Hinweise ausschließlich in gesprochener Form präsentiert. Fest vorgegebene Präsentationszeiten stellten eine konstante Lernzeit bei beiden Varianten sicher.

Erhebungsinstrumente Erhebungsinstrumente waren ein Eye-Tracker, ein Pre- und ein Posttest. Bei dem Eye-Tracker handelte es sich um einen EyeFollower von LC Technology mit einer Abtastrate von 120 Hz. Der verwendete Monitor hatte eine Größe von 24 Zoll und eine Auflösung von 1920×1200 Pixeln. Die Distanz zwischen Display und Kopf lag zwischen 46 und 97 cm. Die Blickbewegungen wurden mit einer Genauigkeit von $< 0.4^\circ$ erfasst. Zum Einsatz kamen vier Kameras, die Augenbewegungen und Kopfbewegungen erfassten. Durch die Berücksichtigung der Kopfbewegungen können Blickbewegungen auch dann präzise erfasst werden, wenn die Teilnehmenden ihren Kopf bewegen. So kann eine möglichst natürliche Bearbeitung der Übungseinheit ohne Einschränkungen durch das Eye-Tracking-System gewährleistet werden. Die Kalibrierung erfolgte mittels 9-Punkten. Die Blickbewegungen wurden durch die Software Nyan 2.0 ausgewertet. Diese nutzt den LC Fixation Detector zur Unterscheidung zwischen Fixationen und Sakkaden.

Der schriftliche Pretest fragte demografische Daten, Kontrollvariablen und das Vorwissen bezogen auf die gezeigten Versuchsmaterialien ab. Der schriftliche Posttest prüfte die Lernergebnisse durch offene Fragen zu folgenden Bereichen ab: a) Angabe der Geräte, b) fachsprachlich korrekte Bezeichnung der Geräte, c) beschriftete Skizze und verbale Beschreibung des Aufbaus sowie d) Beschreibung der Durchführung.

Datenauswertung Die Auswertung der Daten erfolgte auf der Basis hierarchischer Regressionsmodelle. Dazu gingen die Ergebnisse des Posttests als Gesamtmaß (über alle Subskalen hinweg) als abhängige Variable ein. Prädiktoren wurden schrittweise eingebunden. Die Variable „Hinweismodalität“ wurde dabei zuerst in das Modell aufgenommen, da anzunehmen ist, dass das Blickverhalten sehr stark von dieser Variable beeinflusst ist. Anschließend folgte die gemeinsame Aufnahme der Prädiktoren Blickdauer und Anzahl der Fixationen auf experimentelle Details.

Ergebnisse von Experiment 2 Die Eye-Tracking-Daten von vier Probandinnen und Probanden mussten aufgrund von Aussetzern des Eye-Tracking-Systems von der weiteren Analyse ausgenommen werden. So wurden die Eye-Tracking-Daten von 40 Studierenden ausgewertet.

Der Mittelwert der Posttestergebnisse über beide Bedingungen hinweg lag bei $M = 64.34\%$ der maximal erreichbaren Punkte ($SD = 12.89\%$). Die Hinweismodalität erklärte fast 60 % der Varianzen im Posttestergebnis ($p < .001$, korrigiertes $R^2 = 0.574$, $95\%CI = [13.829, 24.743]$). Beim gemeinsamen Hinzufügen der Blickdauer und der Anzahl der Fixationen auf experimentelle Details als Prädiktoren verbesserte sich das Modell signifikant ($p < .001$), und die Varianzaufklärung stieg auf insgesamt, $R^2 = 0.877$ (korrigiert). Die standardisierten Regressionskoeffizienten des letzten Modells, in das alle Prädiktoren aufgenommen wurden, hoben die Rolle von folgenden zwei Prädiktoren hervor. Sowohl die Hinweismodalität ($p < .001$, $stand.\beta = -20.035$, $95\%CI = [-28.816, -11.794]$) als auch die Anzahl der Fixationen ($p < .001$, $stand.\beta = 1.545$, $95\%CI = [1.077, 1.965]$) erklärten signifikant Unterschiede im Lernergebnis. Für die Blickdauer ($p > .10$) konnte hingegen im Modell keine Signifikanz festgestellt werden.

Zusammengefasst zeigen die Ergebnisse, dass durch Eye-Tracking-Metriken (hier die Anzahl der Fixationen) ein beachtlicher Anteil der Varianzen in den Posttestergebnissen erklärt werden kann.

Experiment 3: Unterschiede in der Verteilung von Blicksakkaden bei der Bearbeitung verschiedener Hinweistypen

Nicht nur Eye-Tracking-Metriken, die mit Fixationen in Zusammenhang stehen, können für die Interpretation der Wirksamkeit von Hinweisen genutzt werden. Wie Experiment 3 zeigt, kann auch eine quantitative Analyse von Blicksakkaden Aufschlüsse geben. Im Experiment 3 wurde untersucht, mit welchen Hinweisen eine Übertragung von Messwerten vom Laborsystem des Experiments in das Koordinatensystem eines Diagramms am effektivsten geübt werden kann. Die Wirkungen dynamischer, interaktiver und statischer Hinweise auf eine spätere Dokumentationsaufgabe wurden miteinander verglichen (s. auch Hoyer & Girwidz, 2020). Zur Analyse der Blickbewegungen wurden Polardiagramme angefertigt, die eine kombinierte Sicht auf die Längen- und Richtungsverteilung der Sakkaden während des Trainings ermöglichen. In Ergänzung zur Arbeit von Hoyer und Girwidz (2020) werden die Unterschiede im Leistungszuwachs von Vor- zu Nachtest zwischen den Gruppen analysiert und unter Berücksichtigung der aufgezeichneten sakkadischen Blickbewegungen interpretiert.

Forschungsfragen

1. Ergeben sich unterschiedliche Leistungszuwächse bei der Dokumentation von Messwerten in Diagrammen, nachdem die Lernenden eines der folgenden drei Trainings durchgeführt haben: a) Training mit dynamischen Hinweisen, b) Training mit interaktiven Hinweisen oder c) Training mit statischen Hinweisen?
2. Gibt es für die verschiedenen Trainings charakteristische Blickbewegungen, in denen sie sich unterscheiden?

Stichprobe Am Experiment nahmen 119 Schüler*innen und Schüler aus acht verschiedenen elften Klassen teil. Die Teilnehmenden wurden zufällig auf die drei Versuchsbedingungen verteilt. Die Gruppe mit dynamischen Hinweisen umfasste $N_{\text{dyn}} = 38$, die Gruppe mit interaktiven Hinweisen $N_{\text{int}} = 41$ und die Gruppe mit statischen Hinweisen $N_{\text{stat}} = 40$ Schüler*innen und Schüler.

Material und Methode Die Anwendung, mit der die Probandinnen und Probanden während des Experiments arbeiteten, basierte auf einem virtuellen Labor, mit dem das Feld eines Permanentmagneten vermessen und die aufgenommenen Messwerte auf unterschiedliche Weise visualisiert werden können (vgl. Hoyer & Girwidz, 2018). Zur Übung der Dokumentation von Messwerten wurden in diese Anwendung Hinweise integriert. Sie sollten dabei helfen, Verknüpfungen zwischen dem Laborsystem und dem Koordinatensystem des Diagramms herzustellen, sind also den relationalen Hinweisen (de Koning et al., 2009) zuzuordnen. Die Wirkungen folgender Übungsphasen wurden miteinander verglichen:

- Dynamische Hinweise: Probandinnen und Probanden in dieser Übungsgruppe übten mit animierten Überblendungen, die zeigten, wie das Laborsystem in das Koordinatensystem des Diagramms überführt werden kann.
- Interaktive Hinweise: In der zweiten Übungsgruppe sollten die Probandinnen und Probanden den zur Messung passenden Ort im Diagramm finden. Nachdem sie den entsprechenden Ort mit einem Klick markierten, wurde sofort der korrekte Punkt in Rot im Diagramm gekennzeichnet.
- Statische Hinweise: In der dritten Übungsgruppe wurden den Probandinnen und Probanden Diagramme präsentiert, bei denen bereits der korrekte Punkt im Diagramm markiert war. Viele computergestützte Experimente zeigen ein ähnliches Verhalten. Dort werden aufgenommene Messwerte automatisiert korrekt in ein Diagramm dokumentiert. Daher wurde diese Gruppe als Referenzgruppe verwendet. In Bezug dazu wurden die Effekte der anderen beiden Trainingsgruppen verglichen.

Jede Teilnehmerin und jeder Teilnehmer absolvierten einzeln und nacheinander einen Pretest gefolgt von einer Übungsphase und einem abschließenden Posttest.

Erhebungsinstrumente Die Erhebungsinstrumente waren ein Pre- und ein Posttest, sowie der Eye-Tracker EyeFollower (zu Details s. Erhebungsinstrumente von

Experiment 2). Im Pre- und Posttest mussten die Teilnehmerinnen und Teilnehmer mit einem Klick möglichst genau den Punkt im Diagramm markieren, der zu einer vorgegebenen Messung im Labor gehörte. Die Präzision in der Bearbeitung dieser Aufgabe wurde mit einem Punktesystem bewertet. In Abhängigkeit vom Abstand zum korrekten Ort im Diagramm wurden jeweils zwischen 0 und 5 Punkten vergeben. Für die statistischen Analysen wurde die Gesamtpunktzahl von Pretest und Posttest herangezogen. Weiterhin wurden die Daten zu Blickbewegungen der Probandinnen und Probanden während der Übungsphase untersucht.

Datenauswertung Der Leistungszuwachs von Pre- zu Posttest wurde durch eine ANOVA ausgewertet. Die unabhängige Variable war die Hinweisgruppe.

Für die Analyse der Verteilung sakkadischer Blickbewegungen während der Intervention wurden für alle Teilnehmenden über ein Python-Skript-Polardiagramm angefertigt. Hierfür wurden die Bibliotheken NumPy, SciPy und Matplotlib genutzt. Das Skript ging wie folgt vor:

Zunächst wurden die Sakkaden entsprechend ihrer Richtung in Winkelfelder zu je 30° gruppiert. Für jede dieser Gruppen wurde eine Kerndichteschätzung durchgeführt, um die Wahrscheinlichkeitsdichte der Längenverteilung der Sakkaden im jeweiligen Winkelfeld abzuschätzen. Hierfür wurde ein Gaußkern genutzt, die Bandbreite wurde mithilfe der Regel von Scott berechnet. Im Anschluss wurden für jedes Winkelfeld die Sakkaden in Polardiagrammen dokumentiert; dabei wird die Länge der Sakkaden in Zentimeter durch den Abstand der Markierung zur Mitte des Polardiagramms dargestellt. Jede der Markierungen wurde farblich gekennzeichnet, wobei die Farbe den jeweiligen Wert der berechneten Wahrscheinlichkeitsdichte codiert.

Ergebnisse von Experiment 3 Es gab einen signifikanten Unterschied im Leistungszuwachs von Pre- zu Posttest zwischen den drei Übungsgruppen ($F(2,116) = 3.86$, $p < .05$; $\text{partial } \eta^2 = .062$).

Geplante Kontraste zeigten, dass der Leistungszuwachs in der Gruppe mit interaktiven Hinweisen signifikant größer war als in der Gruppe mit statischen Hinweisen ($t(116) = 2.692$; $p < .01$; $\text{partial } \eta^2 = .059$; $r = 0.242$; $d = .499$).

Der Leistungszuwachs in der Gruppe mit dynamischen Hinweisen unterschied sich hingegen nicht signifikant vom Zuwachs in der Gruppe mit statischen Hinweisen ($t(116) = 0.747$; $p > .05$).

Die Eye-Tracking-Daten von sechs Teilnehmenden waren wegen Aussetzern im Eye-Tracking-System unvollständig. Diese Datensätze wurden daher von der Analyse ausgenommen. Bei den verbleibenden Teilnehmenden konnten deutliche Unterschiede in der Sakkadenverteilung während der Übungsphase in den drei Trainingsgruppen festgestellt werden. Abb. 4.2 zeigt exemplarisch für jeweils eine Person aus der Trainingsgruppe mit dynamischen Hinweisen, interaktiven Hinweisen und statischen Hinweisen die Verteilung der Sakkaden während der Übungsphase. Die Sakkadenverteilungen der anderen Personen in der jeweiligen Übungsgruppe wiesen alle eine ähnliche Struktur auf.

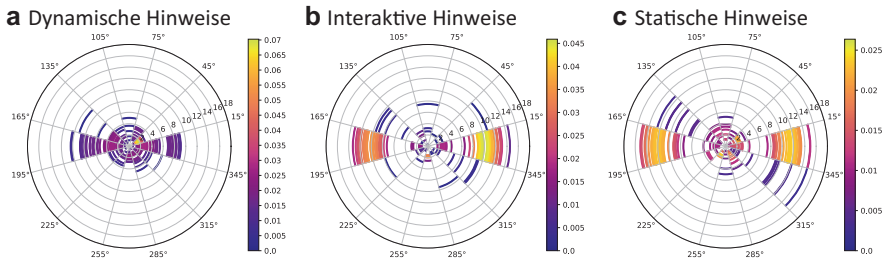


Abb. 4.2 Die Polardiagramme visualisieren jeweils für eine Person die Verteilung der Sakkaden während der Übung mit **a** dynamischen Hinweisen, **b** interaktiven Hinweisen und **c** statischen Hinweisen

In der Gruppe mit dynamischen Hinweisen enthielten die Blickmuster vermehrt kurze sakkadische Bewegungen, die in alle Raumrichtungen verteilt waren. Die längsten Sakkaden traten in der Horizontalen auf, die Auftretenswahrscheinlichkeit dieser langen Sakkaden war aber wesentlich geringer als die der kurzen Sakkaden. Dies zeigt, dass eine dynamische Animation vor allem zu kurzen Blickbewegungen führt, mit denen die Dynamik verfolgt wird, während repräsentationsübergreifende Blickbewegungen nur selten stattfinden. Das Blickverhalten in der Gruppe mit interaktiven Hinweisen und statischen Hinweisen zeichnet sich hingegen vor allem durch lange Sakkaden in der Horizontalen aus. Mit diesen Augenbewegungen führten die Probandinnen und Probanden einen repräsentationsübergreifenden Abgleich zwischen Laborsystem und Koordinatensystem durch.

4.4 Diskussion

In den drei vorgestellten Experimenten wurde die Wirkung verschiedener Hinweistypen in Arbeitsmaterialien zu unterschiedlichen physikalischen Themen untersucht. Eye-Tracking-Daten konnten dabei jeweils tiefgreifendere Einblicke in die Wirkung der gegebenen Hinweise liefern. Dabei kamen qualitative und quantitative Analysen zum Einsatz.

Im ersten Experiment wurden Ergebnisse einer qualitativen Studie vorgestellt. Dabei zeigt sich, dass qualitative Analysen mit dem Eye-Tracker tiefgreifende Erkenntnisse liefern können, die quantitativ nicht ohne Weiteres fassbar sind. So können für Gruppen von Lernenden charakteristische Blickmuster identifiziert werden, die sich bei der Bearbeitung von Aufgaben zu elektrischen Schaltungen zeigen.

Die gefundenen Unterschiede im Blickverhalten weisen auf unterschiedliche Aufmerksamkeitsverteilungen hin und legen damit auch eine unterschiedliche mentale Aktivität von Personen mit niedrigem und hohem Vorwissen nahe. Abgeleitet aus der Theorie der kognitiven Belastung (Sweller et al., 1998) sind

Personen mit niedrigem Vorwissen während der Informationsaufnahme und -verarbeitung mental grundsätzlich stärker belastet als Personen mit hohem Vorwissen. Bezogen auf die Blickpfade der Person mit niedrigem Vorwissen bedeutet das, dass die mentale Belastung primär auf die Übersetzung zwischen textueller Erklärung und Schaltungselementen zurückzuführen ist. Für darüber hinausgehende Vergleiche fehlen offensichtlich die Ressourcen. Die Person mit hohem Vorwissen verwendete deutlich weniger Aufmerksamkeit auf Übersetzungsprozesse. Dies lässt darauf schließen, dass sie deutlich schneller in der Lage ist, ein mentales Modell der einzelnen Schaltungen zu generieren und die Schaltungen dadurch ganzheitlicher erfassen bzw. bearbeiten kann. Die Blickbewegungen zwischen den Schaltungen untereinander und zwischen Schaltungen und Aufgabenstellung können ein zusätzlicher Indikator dafür sein.

Experiment 2 untersuchte, inwieweit beim Lernen mit Hinweisen in Videos die Modalität der Hinweise und die Maße der visuellen Aufmerksamkeit das Lernergebnis vorhersagen können. Ergebnisse der hierarchischen Regressionen zeigen zunächst, dass die Gruppe mit den auditiven Hinweisen bessere Lernergebnisse erzielt als die Gruppe mit den visuellen Hinweisen.

Es ist anzunehmen, dass in der Gruppe mit visuellen Hinweisen die visuelle Aufmerksamkeit zwischen den hell hervorgehobenen Bildelementen und den Textfeldern aufgeteilt werden muss. Dies führt im Vergleich zur Gruppe mit auditiven Hinweisen zu einer erhöhten kognitiven Belastung (vgl. Chandler & Sweller, 1992).

Aufgrund der aufgeteilten Aufmerksamkeit hat die Gruppe mit den visuellen Hinweisen die Bildelemente in den Videos insgesamt weniger oft und lange betrachtet, da sie zusätzlich die Textfelder verarbeiten musste. Bestätigt wird dies in der ergänzenden Analyse von Watzka et al. (2021).

Weiterhin zeigen die Ergebnisse, dass das Lernergebnis zu einem erheblichen Anteil über die Anzahl der Fixationen vorhergesagt werden kann. Dies kann als Maß für die Elaboration des Lernmaterials interpretiert werden (z. B. Alemdag & Cagiltay, 2018).

Die Ergebnisse der dritten Studie zeigen, inwieweit sich sakkadische Bewegungen bei der Wahrnehmung unterschiedlicher Hinweistypen unterscheiden. Bei der Analyse werden Unterschiede zwischen der Gruppe mit dynamischen Hinweisen und den Gruppen mit interaktiven und statischen Hinweisen deutlich. Dynamische Hinweise resultieren in kurzen Blickbewegungen, mit denen die Probandinnen und Probanden den Verlauf der Animation beobachten. In den anderen beiden Übungsgruppen sind vermehrt längere Sakkaden festzustellen. Interaktive und statische Hinweise regten also einen repräsentationsübergreifenden Abgleich zwischen den Darstellungen im Laborsystem und im Diagramm an. In Ergänzung zur Arbeit von Hoyer und Girwidz (2020) zeigte sich außerdem, dass für die untersuchte Anwendung eine Übung mit interaktiven Hinweisen zu den größten Leistungszuwächsen führt.

Aus den Ergebnissen lassen sich Schlussfolgerungen für die Entwicklung computergestützter Übungseinheiten ableiten. Es konnte gezeigt werden, dass sich dynamische Hinweise vor allem dazu eignen, die Aufmerksamkeit der Lernenden

auf die Dynamik zu konzentrieren. Zur Leistungsverbesserung in der untersuchten Aufgabe haben hingegen die Informationen stärker beigetragen, die die Lernenden durch die repräsentationsübergreifenden Sakkaden in der Übung mit interaktiven Hinweisen wahrgenommen haben.

4.5 Resümee

Eye-Tracking-Systeme liefern für Untersuchungen zur Wirkung von Hinweisen in Lernmaterialien eine Vielzahl von Daten, die auf unterschiedliche Arten ausgewertet werden können. Die drei vorgestellten Experimente geben hierzu einen Einblick. Je nach Fragestellung sind qualitative und/oder quantitative Analysen von Eye-Tracking-Daten geeignet. Es zeigt sich, dass Eye-Tracking ein sehr vielseitiges Verfahren ist, das für die Analyse der Wirksamkeit unterschiedlicher Arten von Hinweisen in Arbeitsmaterialien zu verschiedenen Fachinhalten eingesetzt werden kann.

Literatur

- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, 125, 413–428.
- Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, 101, 150–167.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62(2), 233–246.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review*, 21, 113–140.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, 20(2), 111–122.
- Glaser, M., & Schwan, S. (2015). Explaining pictures: How verbal cues influence processing of pictorial learning material. *Journal of Educational Psychology*, 107(4), 1006–1018.
- Höfler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17(6), 722–738.
- Hoyer, C., & Girwidz, R. (2018). A remote lab for measuring, visualizing and analysing the field of a cylindrical permanent magnet. *European Journal of Physics*, 39(6), 65808.
- Hoyer, C., & Girwidz, R. (2020). Animation and interactivity in computer-based physics experiments to support the documentation of measured vector quantities in diagrams: An eye tracking study. *Physical Review Physics Education Research*, 16, Article 020124.
- Hurzlmeier, M., Watzka, B., Hoyer, C., Girwidz, R. & Ertl, B. (2021). Lernergebnisse und individuelle Prozesse des Physiklernens mit auditiven und visuellen Hinweisen. *Zeitschrift für Pädagogische Psychologie*, 1–18. <https://doi.org/10.1024/1010-0652/a000331>.
- Jarodzka, H., van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70.
- Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied*, 18(2), 178–191.

- Klein, P., Viiri, J., & Kuhn, J. (2019). Visual cues improve students' understanding of divergence and curl: Evidence from eye movements during reading and problem solving. *Physical Review Physics Education Research*, 15, Article 010126.
- Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human-Computer Studies*, 65(11), 911–930.
- Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and pictures in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, 60(1), 95–109.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.
- O'Keefe, P. A., Letourneau, S. M., Homer, B. D., Schwartz, R. N., & Plass, J. L. (2014). Learning from multiple representations: An examination of fixation patterns in a science simulation. *Computers in Human Behavior*, 35, 234–242.
- Ott, N., Brünken, R., Vogel, M., & Malone, S. (2018). Multiple symbolic representations: The combination of formula and text supports problem solving in the mathematical field of propositional logic. *Learning and Instruction*, 58, 88–105.
- Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17, 19–36.
- Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction*, 36, 11–26.
- Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educational Research Review*, 23, 1–24.
- Schüler, A. (2017). Investigating gaze behavior during processing of inconsistent text-picture information: Evidence for text-picture integration. *Learning and Instruction*, 49, 218–231.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (2. Aufl., S. 263–278). Cambridge University Press.
- Watzka, B., Hoyer, C., Ertl, B., & Girwidz, R. (2021). Wirkung visueller und auditiver Hinweise auf die visuelle Aufmerksamkeit und Lernergebnisse beim Einsatz physikalischer Lernvideos. *Unterrichtswissenschaft*, 49, 627–652. <https://doi.org/10.1007/s42010-021-00118-7>.
- Xie, H., Mayer, R., Wang, F., & Zhou, Z. (2019). Coordinating visual and auditory cueing in multimedia learning. *Journal of Educational Psychology*, 111(2), 235–255.



Förderung der strukturierten Anzahlerfassung bei Kindern mit Schwierigkeiten im Rechnenlernen: Eine explorative Eye-Tracking Studie zur Evaluation schulischer Förderung in Klasse 5

Maïke Schindler und Florian Schindler

Inhaltsverzeichnis

5.1	Einleitung	75
5.2	Theoretischer Hintergrund und vorherige Forschung	76
5.3	Das Förderprojekt	77
5.4	Die Begleitforschung	79
5.5	Ergebnisse	84
5.6	Diskussion und Ausblick	85
	Literatur	89

5.1 Einleitung

Schwierigkeiten im Rechnenlernen sind ein zentrales Thema in der mathematikdidaktischen Praxis und Forschung – nicht zuletzt durch die wachsende Bedeutung inklusiver Bildung (Moser Opitz et al., 2018). Lehrkräfte wie auch Forschende sind bemüht, Kinder und Jugendliche mit Schwierigkeiten im Rechnenlernen individuell und adäquat zu unterstützen (Moser Opitz et al., 2017; Scherer et al., 2016). Dabei hat sich gezeigt, dass Interventionen teilweise erfreuliche Effekte auf Mathematikleistungen haben (vgl. Chodura et al., 2015; Kroesbergen & Van Luit, 2003). Jedoch zeigte sich auch, dass Interventionen in der *Schulpraxis* mitunter

M. Schindler (✉)

Department Heilpädagogik und Rehabilitation, Universität zu Köln, Köln, Deutschland

E-Mail: maïke.schindler@uni-koeln.de

URL: <https://www.hf.uni-koeln.de/39025>

F. Schindler

Gesamtschule Wulfen, Dorsten, Deutschland

E-Mail: schindler@gs-wulfen.de

keine oder sogar negative Effekte auf die Mathematikleistungen von Lernenden mit Schwierigkeiten haben können (vgl. Kvande et al., 2019; Moser Opitz et al., 2018).

Mit dem übergeordneten Ziel, Förderung für Kinder mit Schwierigkeiten im Rechnenlernen zu gestalten, die in der Schulpraxis effektiv ist, wurde ein Kooperationsprojekt im Sinne partizipativer Forschung (Von Unger, 2014) durchgeführt. Partizipative Forschung ist charakterisiert durch das Anliegen, soziale Wirklichkeit partnerschaftlich zu erforschen und zu verändern und geht davon aus, dass „die Wirksamkeit von Maßnahmen ... unter Praxisbedingungen nachgewiesen werden [muss]“ (ebd., S. 7). Hierin eingebettet ist die vorgestellte explorative Studie, bei der eine Förderung durch eine qualifizierte und erfahrene Lehrkraft in der Schulpraxis durchgeführt wurde. Die wissenschaftliche Begleitung erfolgte durch den universitären Akteur. Im Rahmen der Begleitforschung wurde die Förderung von $n=8$ Kindern mit Schwierigkeiten im Rechnenlernen – im Vergleich zu einer Kontrollgruppe – evaluiert. Die Begleitforschung erfolgte durch ein Prä-Post-Test Design (Beginn/Ende des Schuljahres) und fokussierte die strukturierte Anzahlerfassung am Hunderter-Rechenrahmen. Hierbei kam Eye-Tracking zum Einsatz: In vorherigen Studien hatte sich gezeigt, dass Eye-Tracking – insbesondere die Analyse von Blickbewegungsvideos – als „hochauflösendes“ Forschungsinstrument bei der strukturierten Anzahlerfassung großes diagnostisches Potenzial aufwies (Schindler, 2019; Schindler & Lilienthal, 2018; Schindler et al., 2019).

Ziel der im Beitrag dargestellten explorativen Studie war es, mögliche Fortschritte der geförderten Kinder in der strukturierten Anzahlerfassung über das Schuljahr hinweg zu erfassen. Hierzu wurden Fehlerraten, Bearbeitungszeiten und die aus Blickbewegungsvideos rekonstruierten Vorgehensweisen der geförderten Kinder zu Beginn und Ende des Schuljahres erhoben, Veränderungen über das Schuljahr hinweg ermittelt und mit jenen von Kindern ohne Förderung verglichen.

5.2 Theoretischer Hintergrund und vorherige Forschung

Mit *Schwierigkeiten im Rechnenlernen* sind im vorliegenden Beitrag Schwierigkeiten in den mathematischen Basiskompetenzen gemeint, das heißt „diejenigen mathematischen Inhalte und Lernziele ..., von welchen aufgrund von empirischen Studien, theoretischen Erkenntnissen und praktischen Erfahrungen angenommen wird, dass sie für einen gelingenden arithmetischen Lernprozess unabdingbar sind“ (Moser Opitz & Schmassmann, 2016, S. 270). Im Zentrum stehen hierbei Kenntnisse und Fähigkeiten im Umgang mit natürlichen Zahlen, das Verständnis von Rechenoperationen und des Stellenwerts – sowie Schwierigkeiten hierin (Gerster & Schulz, 2004; Wartha & Schulz, 2011). In verschiedenen Studien zeigte sich, dass Kinder auch zu Beginn der Sekundarstufe, ab Klasse 5, teilweise noch große Schwierigkeiten hierin aufweisen – vor allem Kinder an Gesamtschulen (vgl. Ehler et al., 2013; Pitters, 2019; Schindler et al., 2019).

Eine Förderung von Kindern mit Schwierigkeiten im Rechnenlernen sollte an den mathematischen Basiskompetenzen – v. a. dem Verständnis natürlicher Zahlen, der Rechenoperationen sowie des Stellenwerts – ansetzen (vgl. Moser

Opitz et al., 2018; Wartha & Schulz, 2011). Dies gilt insbesondere für eine Förderung an der wichtigen Transition von Grundschule zu weiterführender Schule (Wartha, 2009). Für die Förderung des Zahlverständnisses ist etwa das Teil-Ganze-Konzept bedeutsam: das Verständnis, dass Mengen in Teilmengen zerlegt und wieder zusammengesetzt werden können (Gerster & Schulz, 2004; Krajewski & Schneider, 2009). Neben dem Strukturieren kleinerer Mengen ist das strukturierte Erfassen größerer Mengen in ihren Teilmengen bedeutsam (Gaidoschik et al., 2017; Gerster & Schulz, 2004). Um dies zu fördern, kommen strukturierte Arbeitsmittel wie das Hunderterfeld und der Rechenrahmen zum Einsatz (Schipper, 2005; Wartha & Schulz, 2011). Empirische Studien haben aufgezeigt, dass Kinder mit Schwierigkeiten im Rechnenlernen im Umgang mit strukturierten Arbeitsmitteln eher zählend vorgehen und die Strukturen, wie den 50er, weniger nutzen als Kinder ohne Schwierigkeiten im Rechnenlernen (Rottmann & Schipper, 2002; Schindler et al., 2019). Dies weist auf die Bedeutung einer Förderung in diesem Bereich hin.

Förderprogramme, die an den mathematischen Basiskompetenzen ansetzen, zeigen teilweise positive Wirkungen (z. B. Moser Opitz et al., 2017). Jedoch deuten Studienergebnisse darauf hin, dass eine Umsetzung der Förderung durch Lehrkräfte in der Schulpraxis schwierig ist und mitunter nicht mit positiven Effekten einhergeht. In einer Interventionsstudie zum Ablösen vom zählenden Rechnen, in der eine Förderung mit hoher ökologischer Validität durch Lehrkräfte selbst durchgeführt wurde, zeigten sich etwa keine signifikanten Effekte (Moser Opitz et al., 2018). In einer groß angelegten Längsschnittstudie in Norwegen zeigte sich zudem, dass sonderpädagogische Förderung in der Schulpraxis negative Effekte auf die mathematischen Fähigkeiten der geförderten Kinder haben kann (Kvande et al., 2019). Obwohl Gründe hierfür weitestgehend unerforscht sind, werden als mögliche Ursachen fehlende Qualifikationen von Lehrkräften sowie unzureichende Verknüpfungen von Forschung und Bildungspraxis diskutiert (ebd.).

5.3 Das Förderprojekt

Das partizipative Forschungsprojekt adressierte Kinder mit Schwierigkeiten im Rechnenlernen zu Beginn der fünften Jahrgangsstufe, da Forschungsergebnisse und Praxiserfahrungen aufzeigen, dass Kinder die Grundschule teilweise mit unzureichenden mathematischen Basiskompetenzen verlassen und dass Kinder in dieser wichtigen Phase der Transition aufgefangen und angemessen gefördert werden müssen. Das Projekt erfolgte in Kooperation einer inklusiven Gesamtschule in NRW sowie insbesondere einer erfahrenen und auf Schwierigkeiten im Rechnenlernen spezialisierten sonderpädagogischen Lehrkraft auf der einen Seite und dem universitären Partner auf der anderen Seite. Die Gesamtschule hatte etwa 180 Fünftklässler*innen, die von 16 abgebenden Grundschulen zur Gesamtschule gewechselt waren. Unter diesen Fünftklässler*innen wurden jene Kinder mit Schwierigkeiten im Rechnenlernen identifiziert und anschließend gefördert.

Die Lehrkraft

Die Förderung der Kinder mit Schwierigkeiten im Rechnenlernen erfolgte in der Schule durch eine Fachlehrperson: eine Lehrperson für sonderpädagogische Förderung mit einer Schulerfahrung von zehn Jahren. Sie hatte an einem Zertifikatskurs für Förderung bei Schwierigkeiten im Rechnenlernen teilgenommen und seitdem seit sechs Jahren individuelle Diagnostik und Förderung bei Kindern mit Schwierigkeiten im Rechnenlernen (in Anlehnung an Wartha & Schulz, 2011, 2017) durchgeführt. Diese Lehrperson war für die Koordination und Durchführung der Diagnostik und Förderung aller Kinder mit Schwierigkeiten im Rechnenlernen in der fünften Jahrgangsstufe zuständig, die eine Förderung in den mathematischen Basiskompetenzen benötigten, die über die curricularen Unterrichtsinhalte hinaus ging.

Die geförderten Kinder

Von den 180 Kindern der Jahrgangsstufe 5 durchliefen jene Kinder, bei denen sich große Auffälligkeiten in den mathematischen Basiskompetenzen zeigten, eine qualitative Diagnostik bzgl. Schwierigkeiten im Rechnenlernen durch die Fachlehrkraft. Hierbei waren die mathematischen Basiskompetenzen – v. a. das Verständnis natürlicher Zahlen, der Rechenoperationen sowie des Stellenwerts (vgl. Moser Opitz et al., 2018; Wartha & Schulz, 2011) – zentral. Für die Diagnostik von Schwierigkeiten im Rechnenlernen waren die drei Hauptkriterien – das verfestigte zählende Rechnen, fehlenden Grundvorstellungen und ein unzureichendes Stellenwertverständnis (Wartha & Schulz, 2017) – ausschlaggebend. Die qualitative Diagnostik dauerte pro Kind etwa 40 min. Über den Zeitraum der fünften Jahrgangsstufe hinweg wurden von der Lehrkraft zwölf Kinder gefördert.

Inhalte der Förderung

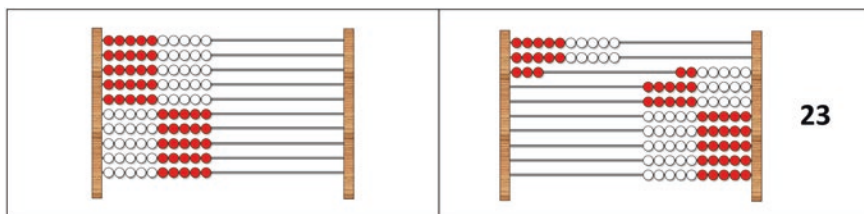
Die Kinder der Fördergruppe erhielten über den Zeitraum eines Schuljahres hinweg eine individuelle Förderung durch die Lehrkraft, die von der Lehrkraft geplant und durchgeführt wurde. Die Förderung umfasste verschiedene Themenblöcke, von denen die ersten drei in Tab. 5.1 dargestellt sind. Nach der Förderung von Grundlagen zur Zahlerfassung und zum Teile-Ganze-Konzept (Themenblock 1) erfolgte eine Förderung von Zahl- und Operationsvorstellungen am Rechenrahmen (Themenblöcke 2 und 3, vgl. Abb. 5.1).

In jedem Themenblock wurde einem Vierschritt im Hinblick auf die Arbeit mit und die Ablösung vom Material gefolgt (Wartha & Schulz, 2011):

1. Das Kind arbeitet mit dem entsprechenden Material und verbalisiert die Materialhandlung.
2. Das Kind verbalisiert die Materialhandlung mit Sicht auf das Material, während die Lehrkraft die Materialhandlung für das Kind ausführt.
3. Das Kind verbalisiert die Materialhandlung ohne Sicht auf das Material, welches sich hinter einem Sichtschutz befindet, wo die Materialhandlungen von der Lehrkraft ausgeführt werden.
4. Das Kind arbeitet mathematisch-symbolisch (vgl. ausführlich in Wartha & Schulz, 2011, 2017).

Tab. 5.1 Inhalte des Förderprogramms (in Auszügen)

Nr	Themenblock
1	Grundlagen • <i>Simultane Zahlerfassung</i>: Simultane und quasi-simultane Erfassung von Mengen bis 10. Strukturnutzung, z. B., „Kraft der 5“ • <i>Teil-Ganze-Konzept</i>: Zusammenfügen und Zerlegen von Mengen bis 10 <i>Material</i>: Holzwürfel, Finger, Papier und Bleistift
2	Zahlvorstellung und Strukturen am Rechenrahmen • <i>Erkunden des Rechenrahmens</i>: Strukturen wie 5er, 10er, 50er und 100er • <i>Anzahlerfassung</i>: Erfassen und Darstellen von Anzahlen am Rechenrahmen durch Strukturnutzung (5, 10, 50, 100). „Blitzsehen“ (mit kurzer Darbietungszeit) <i>Material</i>: Rechenrahmen, Papier und Bleistift
3	Addition und Subtraktion am Rechenrahmen • <i>Auf und ab zum Zehner</i>: Ergänzen bzw. Vermindern zum Zehner • <i>Addition und Subtraktion</i>: mit und ohne Zehnerübergang (z. B. $26 + 7$, $46 - 9$). Verbalisieren und notieren der Zwischenschritte (z. B. für $26 + 7$: $26 + 3 = 30$, $30 + 4 = 34$) <i>Material</i>: Rechenrahmen, Papier und Bleistift

**Abb. 5.1** Rechenrahmen (idealisiertes Bild) inklusive Beispiel

Dabei wurde das Ziel verfolgt, durch die Verinnerlichung der Arbeitsmittel mentale Vorstellungen von Zahlen und Operationen aufzubauen.

Die Kinder durchliefen die drei Themenblöcke in individuellem Tempo unterschiedlich schnell – wobei sie hierin je etwa 13 bis 20 Fördereinheiten erhielten. Es schloss sich bei allen Kindern eine Förderung mit Mehrsystemmaterial zum Operations- und Stellenwertverständnis an. Dies kann aus Platzgründen hier nicht eingehender dargestellt werden.

5.4 Die Begleitforschung

Ziel und Forschungsfragen

Ziel der Begleitforschung war es, die möglichen Fortschritte der Kinder der Fördergruppe bei der strukturierten Anzahlerfassung am Rechenrahmen zu erfassen. Es wurde der folgenden Forschungsfrage nachgegangen:

Gibt es Veränderungen der Kinder der Fördergruppe in der strukturierten Anzahlerfassung am Rechenrahmen über das Schuljahr hinweg?

Design der Studie und Eye-Tracking Erhebung

Aufgrund des explorativen Charakters der Studie und des Aufwands des Einsatzes von Eye-Tracking in der Schulpraxis wurde ein Prä-Post-Test Design mit Förder- und Kontrollgruppe (s. u.) gewählt: Die Kinder nahmen zu Beginn des Schuljahres, vor Beginn der Förderung, an einer Prä-Testung (t1) sowie am Ende des Schuljahres an einer Post-Testung (t2) teil (Abb. 5.2). Zum Zeitpunkt der Post-Testung arbeiteten alle Kinder seit mindestens zehn Wochen nicht mehr aktiv mit dem Rechenrahmen. Somit erhielt keines der Kinder zum Zeitpunkt der Testungen aktiv eine Förderung am Rechenrahmen.

Teilnehmende Kinder

Für die Kontrollgruppe, die keine Förderung erhielt, wurden zwölf Kinder ausgewählt. Es handelte sich um Kinder, die im Schulunterricht keine Schwierigkeiten im Rechnenlernen gezeigt hatten. Elf der zwölf Kinder (bzw. deren Erziehungsberechtigte) waren mit der Teilnahme einverstanden und nahmen entsprechend teil (Tab. 5.2). Von den 12 geförderten Kindern erklärten sich neun Kinder (bzw. deren Erziehungsberechtigte) einverstanden, an der Begleitforschung teilzunehmen, wobei ein Kind in der Post-Erhebung durch einen längeren Krankenhausaufenthalt ausfiel. Hierdurch werden bei der Begleitstudie acht teilnehmende Kinder in der Fördergruppe berücksichtigt (Tab. 5.2).

Mit allen Kindern der Förder- und Kontrollgruppe wurde zum Zwecke der Begleitforschung ein standardisiertes Verfahren zur Überprüfung der

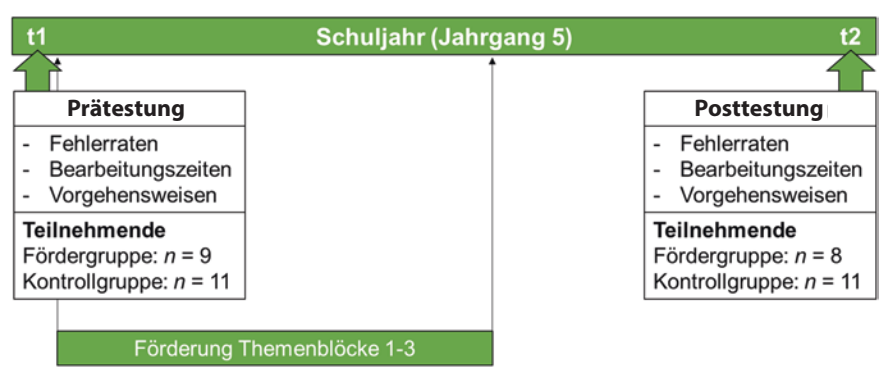


Abb. 5.2 Design der Studie

Tab. 5.2 Teilnehmende Kinder. Angaben zu Alter, Geschlecht, Muttersprache und Rechenleistung

	Fördergruppe	Kontrollgruppe
Alter: <i>M (SD)</i>	10;9 (0;6)	10;5 (0;4)
Geschlecht: Mädchen	6 von 8	4 von 11
Muttersprache: deutsch	5 von 8	9 von 11
HRT T-Wert: <i>M (SD)</i>	26.0 (7.4)	48.4 (6.8)

arithmetischen Fähigkeiten eingesetzt, um die qualitative Diagnostik von Schwierigkeiten im Rechnenlernen (s. o.) abzusichern: der Heidelberger Rechentest (HRT, Haffner et al., 2005). Es handelt sich hierbei um einen Speed-Test, bei dem die Kinder innerhalb kurzer Zeit möglichst viele Aufgaben (z. B. im Bereich der Addition oder Subtraktion im Kopf) lösen sollen. „Dahinter steckt die Annahme, dass Kinder mit ineffektiven, umständlichen oder unsicheren Lösungsstrategien weniger Aufgaben pro Zeiteinheit lösen können als geschickte, kompetente und geübte Rechner“ (ebd., S. 13): Zählende Rechner benötigen bspw. länger als Kinder, die geschickt rechnen oder auf automatisierte Aufgaben zurückgreifen. Der HRT wurde mit allen teilnehmenden Kindern als Gruppentest durchgeführt. Dabei wurde nur die erste Hälfte des HRT eingesetzt, welche auch allein zur Erfassung der elementaren Rechenleistung genutzt werden kann (Haffner et al., 2005, siehe auch Landerl, 2013; Schleifer & Landerl, 2011).

Eye Tracker

Die Datenerhebung erfolgte mittels der Eye-Tracking Brille Tobii Pro Glasses 2. Diese hat eine hoch-auflösende Sichtfeldkamera, binokulare Augenkameras und Eye-Tracking Sensoren, sowie ein Mikrofon, das den Ton während der Aufnahme aufzeichnet. Die Sichtfeld-Kamera zeichnet ein Full HD Video (1920 × 1080 Pixel) des Sichtfeldes mit 82 Grad (horizontal) × 52 Grad (vertikal) auf. Die Blickbewegungen werden mittels Infrarot-Illumination mit einer Bildfrequenzrate von 50 Hz und einer Accuracy von durchschnittlich 0.3 bis 1.1 Grad (Tobii Pro, 2017) erhoben, indem die korneale Reflektion der Pupillen gemessen wird. Die Eye-Tracking Brille ist mit 45 g leicht und wenig einschränkend. Die Aufgaben wurden auf einem 24 Zoll Full-HD Monitor (1920 × 1080 Pixel) mit einer Bildfrequenzrate von 60 Hz präsentiert. Das Kind saß auf einem Stuhl in etwa 50 cm Entfernung vom Bildschirm. Der Einsatz einer Eye-Tracking Brille ermöglichte dabei die Erfassung von Blicken auch außerhalb des Bildschirms sowie eine synchronisierte Aufzeichnung des Videos der Szene (inkl. Gesten), der Blickbewegungen sowie der Antworten der Kinder (durch die Audioaufzeichnung). Keines der Kinder benötigte bei der Teilnahme an der Studie Sehhilfen.

Aufgaben und Durchführung

Für die Studie wurde eine Abbildung des Rechenrahmens erstellt (Abb. 5.1). Die Aufgaben bestanden darin, dargestellte Anzahlen zu benennen. Es wurden zehn Anzahlen dargestellt, bei denen die Ziffern so gewählt wurden, dass jeder Zehner und Einer je einmal vorkam. Die Anzahlen waren (der Größe nach geordnet): 7, 15, 20, 31, 43, 54, 68, 76, 89, 92.

Die Testungen fanden in einem ruhigen Raum in der Schule der Kinder statt. Zunächst wurde eine 1-Punkt Kalibrierung durchgeführt. Bevor den Kindern die o. g. Anzahlen präsentiert wurden, sahen sie ein Foto eines Rechenrahmens, den sie je betrachteten, sowie anschließend die digitale Abbildung (Abb. 5.1). Danach erfolgten zwei Übungsaufgaben. Die Kinder erhielten die Instruktion, die Anzahlen korrekt und so zügig wie möglich zu benennen. Zwischen den Aufgaben erschien ein Fixationsstern, den die Kinder anschauen sollten, bevor die nächste Aufgabe erschien.

Daten

Um die Veränderungen der Vorgehensweisen zu betrachten, wurden Fehlerraten, Bearbeitungszeiten und Vorgehensweisen der Kinder untersucht. Bei Fehlerraten und Bearbeitungszeiten handelt es sich um gängige Größen in Zusammenhang mit der Untersuchung mathematischen Aufgabenbearbeitungen – die Untersuchung von Vorgehensweisen mittels Eye-Tracking ist eine Methode, die in den letzten Jahren etabliert wurde. *Fehler* wurden durch die mündlichen Antworten der Kinder (z. B., „76“ korrekt – „86“ inkorrekt) erhoben. Die *Bearbeitungszeiten* wurden vom Erscheinen der jeweiligen Aufgabe bis zur mündlichen Antwort gemessen. Für die Analyse der *Vorgehensweisen* wurden mittels der Software Tobii Pro Lab Blickbewegungsvideos produziert, bei denen keine Blicke ausgeschlossen wurden (Filter: raw) (siehe <https://youtu.be/C2mlnKxHBeg> für ein Beispiel).

Datenanalyse

Fehlerraten und Veränderungen der Fehlerraten Um die Veränderungen der Fehlerraten zwischen t1 und t2 zu vergleichen, wurde zunächst für alle 380 Aufgabenbearbeitungen (19 Kinder $\times$ 10 Aufgaben $\times$ 2 Testzeitpunkte, keine Datenverluste) bestimmt, ob ein Fehler vorlag. Es wurden die Fehleranzahlen pro Kind in Prä- und Post-Testung sowie die Differenzen zwischen Prä- und Post-Test ermittelt.

Bearbeitungszeiten und Veränderungen der Bearbeitungszeiten Zunächst wurde für alle 380 Aufgabenbearbeitungen in Prä- und Post-Testung die Bearbeitungszeit ermittelt. Anschließend wurde für jede Aufgabenbearbeitung jedes Kindes die paarweise Differenz aus Prä- und Post-Testungs-Bearbeitungszeit gebildet (z. B. Schüler Nidal bei Aufgabe 76: Prä: 21.037 ms, Post: 8.352 ms, Differenz ($t_2 - t_1$): -12.685 ms).

Vorgehensweisen und Veränderungen der Vorgehensweisen Um die Veränderungen in den Vorgehensweisen der Schüler*innen zu bestimmen, wurden zunächst Vorgehensweisen für 366 Aufgabenbearbeitungen (19 Kinder $\times$ 10 Aufgaben $\times$ 2 Testzeitpunkte, 14 Datenverluste) bestimmt. Die Vorgehensweisen wurden deduktiv-kategoriengeleitet zugewiesen – basierend auf einem Kategoriensystem, das von Schindler et al. (2019) für die Analyse der Anzahlerfassung am Rechenrahmen auf Grundlage von Blickbewegungsvideos entwickelt wurde. Es wurden folgende Vorgehensweisen unterschieden: (V1) *Einer zählen*: Das Kind zählt jede einzelne Kugel; (V2) *Fünfer zählen*: Das Kind zählt Fünfer. Dabei werden bspw. bei 92 18 Fünfer gezählt und 2 addiert, (3) *Reihen zählen*: Das Kind zählt Reihen. Bei 92 werden etwa neun Reihen nacheinander abgezählt, (4) *Simultanerfassung größerer Einheit*: Das Kind nutzt eine größere Einheit zur Anzahlerfassung, z. B. wird bei 92 der Fünfziger genutzt und die weiteren Reihen gezählt; (5) *Subtraktion oder letzte Reihe*: Das Kind fokussiert auf die zur 100 fehlenden Punkte und subtrahiert – oder es nutzt, etwa bei 92, die Kenntnis, dass es sich um die letzte Reihe handelt. Die Vorgehensweisen wurden den Auf-

gabenbearbeitungen von zwei Codern getrennt von einander zugewiesen und die Interraterreliabilität nach Cohen (1988) ermittelt. Cohen's kappa lag bei 0.96, was als nahezu perfekt erachtet werden kann (Landis & Koch, 1977).

Um die Veränderungen der Vorgehensweisen zwischen t1 und t2 zu evaluieren, wurden für jedes Kind und jede Aufgabe überprüft, ob die genutzte Vorgehensweise zu t1 und t2 dieselbe war (keine Veränderung), ob zu t2 eine günstigere Vorgehensweise genutzt wurde (positive Veränderung) oder ob zu t2 eine weniger günstige Vorgehensweise vorlag (negative Veränderung). Grundlage hierfür war eine ordinale Einschätzung der Vorgehensweisen – von V1 aufsteigend bis V5: Bei 92 wurde es etwa als eher ungünstig eingeschätzt, wenn alle Kugeln einzeln gezählt wurden (V1), als etwas günstiger, wenn alle Fünfer gezählt wurden (V2), als wiederum günstiger, wenn die Reihen gezählt wurden (V3), als noch günstiger, wenn auf den 50er zurückgegriffen wurde (V4) und als erneut günstiger, wenn die Nähe zur 100 genutzt wurde (V5). Schließlich wurde für jede Aufgabe und jedes Kind ermittelt, ob keine Veränderung, eine positive oder eine negative Veränderung vorlag. Aufgrund der genannten Datenverluste lag der paarweise Vergleich für insgesamt 176 Aufgabenbearbeitungen vor.

Statistische Analysen

Für die statistischen Analysen wurde die Software IBM SPSS 26 genutzt.

Veränderungen der Fehlerraten Die Gesamtanzahl der Fehler sowie der Fehleratendifferenzen war recht niedrig. Zunächst wurde ein Shapiro-Wilks Test auf Normalverteilung durchgeführt (Shapiro & Wilk, 1965). Hierbei zeigte sich, dass die Fehleratendifferenzen nicht normalverteilt waren. Daher wurde der Mann Whitney U Test genutzt, welcher sich als nicht-parametrischer Test auch für kleine Gruppengrößen eignet.

Veränderungen der Bearbeitungszeiten Um die Gruppen bezüglich der Bearbeitungszeitendifferenzen zu vergleichen, wurde eine mixed ANOVA mit Messwiederholung durchgeführt. Die Aufgaben (7, 15, 20, 31, 43, 54, 68, 76, 89, 92) stellten den Innersubjektfaktor und die Gruppe (Fördergruppe, Kontrollgruppe) den Zwischensubjektfaktor dar.

Veränderungen in den Vorgehensweisen Um die Gruppen bezüglich der globalen Veränderungen in den Vorgehensweisen zu vergleichen (positive Veränderung, negative Veränderung, keine Veränderung), wurde der exakte Freeman-Halton Test durchgeführt, da die Voraussetzungen für den Chi Quadrat Test nicht erfüllt waren. Für die Effektstärke wurde Cramér's V berechnet. Um Gruppenunterschiede separat für positive, negative und keine Veränderungen zu untersuchen, wurden zwei Mann-Whitney U Tests sowie ein Mediantest (bei fehlender Varianzhomogenität in Kombination mit kleiner Gruppengröße, vgl. Bortz & Lienert, 2008) durchgeführt. Die Resultate wurden nach Bonferroni-Holm alphaadjustiert.

5.5 Ergebnisse

Fehlerraten

Bereits im Prä-Test machten die Kinder nur wenige Fehler: Die Fehlerrate der Kinder der Fördergruppe lag pro Kind bei durchschnittlich 0.100 Fehlern pro Aufgabe, jene der Kontrollgruppe bei 0.091 Fehlern pro Aufgabe. Die Kinder machten im Post-Test durchschnittlich 0.038 (Fördergruppe) bzw. 0.027 (Kontrollgruppe) Fehler pro Aufgabe.

Die durchschnittlichen Fehlerratendifferenzen betrugen für die Fördergruppe -0.063 und für die Kontrollgruppe -0.064 Fehler pro Kind pro Aufgabe. Die Fehlerraten bei der strukturierten Anzahlerfassung waren somit in beiden Gruppen insgesamt in ähnlichem Maße rückläufig. Die Fehlerratendifferenzen wiesen zwischen den Gruppen keinen bedeutsamen Unterschied auf ($U=40.500$, $p=.778$, $r=-0.05$).

Bearbeitungszeiten

Für die Bearbeitungszeitdifferenzen wurde eine 2 (Gruppe: Fördergruppe, Kontrollgruppe) $\times 10$ (Aufgabe: 7, 15, 20, 31, 43, 54, 68, 76, 89, 92) mixed ANOVA mit Messwiederholung durchgeführt. Da eine Verletzung der Voraussetzung der Sphärizität vorlag, wurde eine Greenhouse-Geisser Korrektur der Freiheitsgrade vorgenommen. Homogenität der Fehlervarianzen zwischen den Gruppen war gemäß dem Levene-Test für alle Variablen erfüllt ($p>.05$). Es konnte keine statistisch signifikante Interaktion zwischen den Gruppen und den Aufgaben ($F(3.593, 61.076)=1.009$, $p=.404$, $\eta_p^2=.056$) festgestellt werden. Es gab keinen signifikanten Effekt für die Gruppen ($F(1, 17)=3.181$, $p=.092$, $\eta_p^2=.158$) und keinen signifikanten Effekt für die Aufgaben ($F(3.593, 61.076)=2.523$, $p=.056$, $\eta_p^2=.129$). Die Veränderungen in den Bearbeitungszeiten zwischen Prä- und Post-Test unterschieden sich somit weder zwischen den Aufgaben noch zwischen den Gruppen signifikant.

Die deskriptive Betrachtung deutet jedoch an, dass die Kinder der Fördergruppe gerade die Aufgaben mit Anzahlen >50 offenbar zu t2 tendenziell schneller erkannten als zu t1 (Abb. 5.3). Bei post-hoc Analysen von Gruppenunterschieden für die einzelnen Aufgaben zeigten sich jedoch auch für diese Aufgaben keine signifikanten Gruppeunterschiede (mit Ausnahme der Aufgabe 76 mit $F(1, 17)=5.467$, $p=.032$, $\eta_p^2=.243$).

Vorgehensweisen

Zunächst wurden für alle Aufgabenbearbeitungen der Kinder in Prä- und Post-Testung Vorgehensweisen aus Blickbewegungsvideos bestimmt. Die deskriptive Betrachtung der ermittelten Vorgehensweisen zeigt auf (Abb. 5.4), dass die Kinder der Fördergruppe offenbar vor der Förderung häufig Reihen zählten (V3), während sie in der Post-Testung häufig größere Einheiten wie den 50er nutzten (V4).

Basierend auf den ermittelten Vorgehensweisen wurde für jedes Kind für jede Aufgabe bestimmt, ob von t1 zu t2 eine positive Veränderung, eine negative Veränderung oder keine Veränderung vorlag. Abb. 5.5 visualisiert die Unterschiede in den Veränderungen der Vorgehensweisen von Kindern mit und ohne

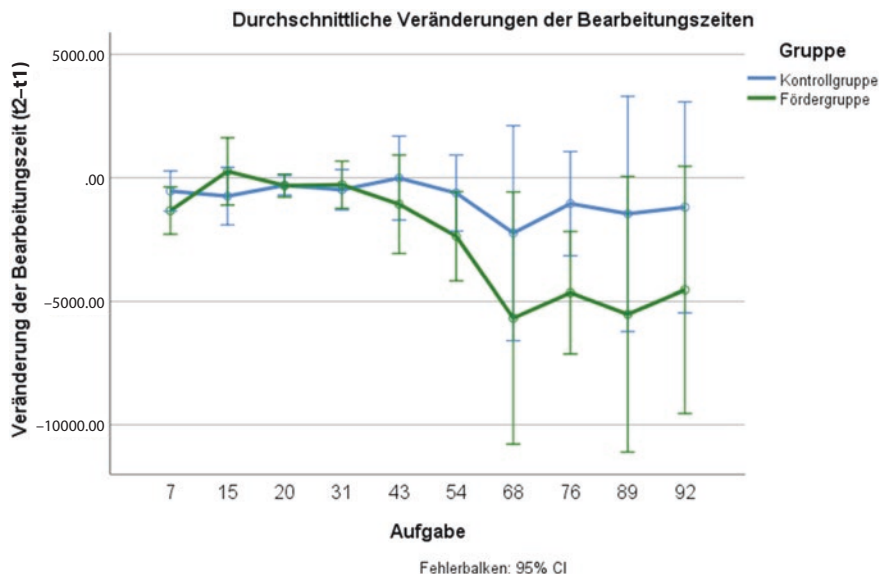


Abb. 5.3 Veränderungen der Bearbeitungszeiten zwischen Prä- und Post-Test (in ms)

Förderung: Während die Kinder ohne Förderung in der Nutzung von Vorgehensweisen überwiegend keine Veränderungen, einzelne negative Veränderungen und wenige positive Veränderungen aufwiesen, zeigten die Kinder mit Förderung überwiegend positive Veränderungen, keine negativen Veränderungen und vergleichsweise weniger Stagnation der Nutzung von Vorgehensweisen. Der visuelle Eindruck bestätigt sich im Intergruppenvergleich: Die Veränderungen in den Vorgehensweisen unterscheiden sich signifikant zwischen Förder- und Kontrollgruppe (Freeman-Halton Test: $p < .001$, $Cramér's V = .377$, mittlerer Effekt).

Im Detail zeigte sich, dass positive Veränderungen bei Kindern der Fördergruppe ($Md = 5$) signifikant häufiger auftraten als bei Kindern ohne Förderung ($Md = 1$, $U = 10.000$, $p = 0.12$, $r = -.650$, starker Effekt). Negative Veränderungen zeigten sich bei Kindern der Fördergruppe ($Md = 0$) signifikant seltener als bei Kindern ohne Förderung ($Md = 1$, Exakter Mediantest: $\chi^2 = 6.378$, $p = .036$, $r = -.553$, starker Effekt). Ebenso kamen keine Veränderungen bei Kindern der Fördergruppe ($Md = 4$) signifikant seltener vor als bei Kindern ohne Förderung ($Md = 6$, $U = 18.500$, $p = .036$, $r = -.489$, mittlerer Effekt). Dies deutet insgesamt auf Veränderungen zugunsten intendierter günstiger Vorgehensweisen bei der Fördergruppe hin.

5.6 Diskussion und Ausblick

Ziel der im Beitrag dargestellten Studie war es, mögliche Fortschritte von geförderten Kindern ($n = 8$) in der strukturierten Anzahlerfassung über ein Schuljahr hinweg zu erfassen. Es wurden Fehlerraten, Bearbeitungszeiten sowie die aus

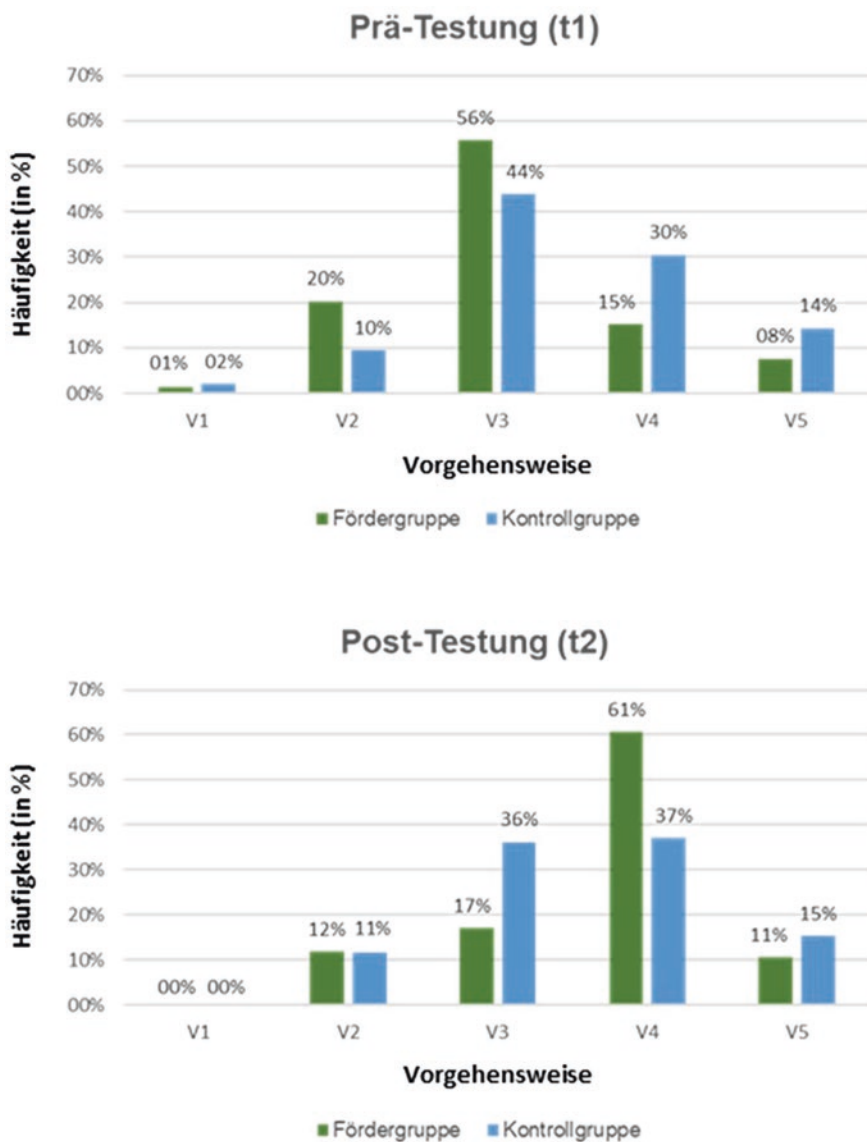
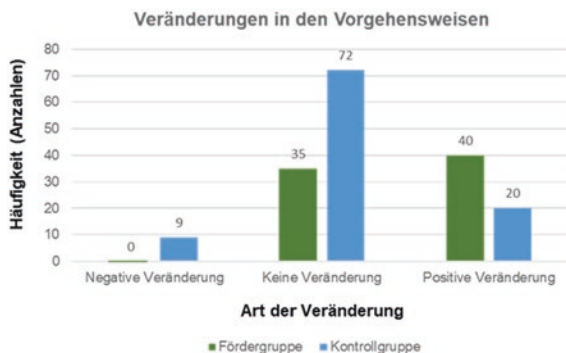


Abb. 5.4 Vorgehensweisen in Prä- und Post-Testung

Blickbewegungsvideos ermittelten Vorgehensweisen der geförderten Kinder in einem Prä- und Post-Test-Design erhoben und die Veränderungen mit jenen von Kindern ohne Förderung ($n = 11$) verglichen.

Insgesamt wurde mit der Studie eine hohe ökologische Validität angestrebt: Es wurde eine schulische Förderung, die durch eine Lehrkraft der entsprechenden

Abb. 5.5 Veränderungen der Vorgehensweisen (Anzahlen) zwischen Prä- und Post-Test



Schule erfolgte, in einem partizipativen Forschungsprojekt evaluiert. Dabei wurde nicht etwa ein Förderprogramm durch den universitären Akteur vorgegeben oder teilnehmende Kinder aus verschiedenen Schulen akquiriert, sondern das Projekt involvierte die Personen und die Förderung, wie sie in der betreffenden Schule vorkamen. Dies zog auf Forschungsseite natürlich Limitationen nach sich. Hierzu zählt u. a. die kleine Teilnehmendenzahl – sowie die entsprechenden Einschränkungen bzgl. der statistischen Analysen. Die Anzahl der Kinder mit Förderung war im Kooperationsprojekt durch die Kapazitäten der Schule beschränkt: Es gab nicht mehr Kinder, die eine entsprechende Förderung erhielten. Jedoch hatte die vorliegende Studie zunächst einen explorativen Charakter und die gewonnen Erkenntnisse zur Effektivität der durchgeführten Förderung können in Zukunft in Studien mit größeren Fallzahlen einfließen. Eine weitere Limitation besteht darin, dass die Kinder der Kontrollgruppe keine Schwierigkeiten im Rechnenlernen zeigten. Hätte die Kontrollgruppe aus Kindern mit Schwierigkeiten im Rechnenlernen bestanden (z. B. durch Aufspaltung der 8 Kinder mit Schwierigkeiten im Rechnenlernen in Förder- und Kontrollgruppe), hätte dies für gewisse Kinder in der Praxis ein Vorenthalten der einjährigen schulischen Förderung bedeutet. Dies erschien uns ethisch nicht vertretbar – zumal eine Förderung im sechsten Schuljahr (i. S. einer Wartekontrollgruppe) nicht umsetzbar gewesen wäre. Jedoch zeigte sich in der Prä-Testung auch bei der Kontrollgruppe, dass die strukturierte Anzahlerfassung noch nicht durchweg gesichert war, da die Kinder teilweise noch Fehler machten und weniger elaborierte Vorgehensweisen nutzten.

Bezogen auf die Erforschung der Förderung von Kindern mit Schwierigkeiten im Rechnenlernen hat die dargestellte explorative Studie aufgezeigt, dass eine Förderung der strukturieren Anzahlerfassung am Rechenrahmen in der Praxis effektiv sein kann. Um im Rahmen der Studie mögliche Effekte von Spontanentwicklung oder des regulären Schulunterrichts auszuschließen, wurden die Fortschritte der geförderten Kinder mit jenen von Kindern ohne Förderung verglichen. Dabei zeigte sich beim Vergleich der Vorgehensweisen der Anzahlerfassung und tendenziell auch beim Vergleich der Bearbeitungszeiten, dass die Kinder mit Förderung größere Fortschritte

machten als Kinder ohne Förderung – mit teilweise großen Effekten. So nutzten die Kinder der Fördergruppe nach der Intervention bei der Anzahlerfassung häufiger den 50er (sogar häufiger als die Kontrollgruppe ohne Schwierigkeiten im Rechnenlernen) und griffen weniger häufig auf das langwierige Zählen von Reihen zurück. Vor dem Hintergrund der kleinen Anzahl geförderter Kinder und der Tatsache, dass die Studie klar statistisch unterpower war, sind diese Erkenntnisse nennenswert. Dabei ist weiterhin zu beachten, dass die Kinder bei der Post-Testung bereits seit mindestens zehn Wochen nicht mehr aktiv mit dem Rechenrahmen arbeiteten, was für die Nachhaltigkeit der schulischen Förderung spricht und die Effekte unterstreicht. Für die Zukunft wäre es derweil erstrebenswert, ähnliche Studien mit Prä-Post- und Follow-Up-Design durchzuführen, wobei der Post-Test unmittelbar nach der Intervention stattfindet. Dies konnte in unserer explorativen Studie organisatorisch nicht realisiert werden.

Bezogen auf die verwendete Methode – das Eye-Tracking – konnte die vorliegende Studie aufzeigen, dass Eye-Tracking gewinnbringend für die Evaluation von Fördermaßnahmen sein kann. Es zeigte sich insbesondere, dass die Analyse von Blickbewegungsvideos fruchtbar ist, um die Vorgehensweisen von Kindern zu erfassen und darauf aufbauend mögliche Fortschritte einer Förderung zu ermitteln. Die Ergebnisse deuten etwa darauf hin, dass die Kinder mit Schwierigkeiten im Rechnenlernen vor der Förderung als häufigste Vorgehensweise Reihen zählten, während sie nach der Förderung häufiger den Fünfiger nutzen. So zählten bspw. bei der Anzahl 54 in der Prä-Testung sieben von acht Kindern Reihen, während in der Post-Testung sieben von acht Kindern den 50er nutzten und nur ein Kind die Reihen einzeln abzählte. Durch die Analyse der Eye-Tracking Daten konnten so Effekte der Förderung aufgezeigt werden, die durch die anderen Maße (Fehlerraten und Bearbeitungszeiten) nicht ersichtlich waren. Dabei ist erwähnenswert, dass die Fehlerraten bereits im Prä-Test niedrig waren und daher keine wesentlichen Fortschritte bei den Fehlerraten erwartbar waren und dass die Potenziale beim Fortschritt eher in der Weiterentwicklung der Vorgehensweisen zu liegen schienen als in den Fehlerraten. Die Bearbeitungszeiten hingegen zeigten die Tendenzen in den Entwicklungen an: So zeigten sich etwa Trends von verkürzten Bearbeitungszeiten bei den geförderten Kindern bei größeren Anzahlen: Dies spiegelt wider, dass die Kinder im Zusammenhang mit der Förderung im Post-Test weniger häufig Reihen zählten (was mitunter langwierig ist), sondern eher auf die 50er-Struktur zurückgriffen.

Unsere explorative Studie hat aufgezeigt, dass das Eye-Tracking sich – bei angemessenen Aufgaben und -design – zur Evaluation von Fördermaßnahmen eignet und wir hoffen, dass zukünftige Forschung Eye-Tracking hierzu vermehrt einsetzt. Während in unserer Studie – aufgrund ihres explorativen Charakters und der Umsetzbarkeit in der Schulpraxis – ein Prä-Post-Test Design gewählt wurde, sind für weiterführende Studien fortlaufende Evaluierungen mittels Eye-Tracking denkbar, die ein engeres Monitoring der Entwicklungen von Vorgehensweisen über eine Fördermaßnahme hinweg ermöglichen.

Die vorliegende explorative Studie hat – bei den genannten Einschränkungen – vielversprechende Ergebnisse in Bezug auf die Förderung von Kindern mit Schwierigkeiten im Rechnenlernen und den Einsatz von Eye-Tracking zur Evaluation von Fördermaßnahmen aufgezeigt. Wir sehen sie als Ausgangspunkt für weitere Forschung in diesem Bereich und hoffen, dass nachfolgende Studien unsere Arbeit fortführen werden.

Literatur

- Bortz, J., & Lienert, G. A. (2008). *Kurzgefasste Statistik für die klinische Forschung* (3. Aufl.). Springer.
- Chodura, S., Kuhn, J. T., & Holling, H. (2015). Interventions for children with mathematical difficulties. *Zeitschrift für Psychologie*, 223(2), 129–144.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science*. Erlbaum.
- Ehlert, A., Fritz, A., Arndt, D., & Leutner, D. (2013). Arithmetische Basiskompetenzen von Schüler*innen und Schülern in den Klassen 5 bis 7 der Sekundarstufe. *Journal für Mathematik-Didaktik*, 34(2), 237–263.
- Gaidoschik, M., Fellmann, A., Guggenbichler, S., & Thomas, A. (2017). Empirische Befunde zum Lehren und Lernen auf Basis einer Fortbildungsmaßnahme zur Förderung nicht-zählenden Rechnens. *Journal für Mathematikdidaktik*, 38(1), 93–124.
- Gerster, H.-D., & Schultz, R. (2004). *Schwierigkeiten beim Erwerb mathematischer Konzepte im Anfangsunterricht. Bericht zum Forschungsprojekt Rechenschwäche – Erkennen, Beheben, Vorbeugen*. PH Freiburg.
- Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *Heidelberger Rechentest (HRT 1–4)*. Hogrefe.
- Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction*, 19, 513–526.
- Kroesbergen, E. H., & Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education*, 24(2), 97–114.
- Kvande, M. N., Bjørklund, O., Lydersen, S., Belsky, J., & Wichstrøm, L. (2019). Effects of special education on academic achievement and task motivation: A propensity-score and fixed-effects approach. *European Journal of Special Needs Education*, 34(4), 409–423.
- Landerl, K. (2013). Development of numerical processing in children with typical and dyscalculic arithmetic skills—a longitudinal study. *Frontiers in Psychology*, 4, 459.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Moser Opitz, E., Freeseemann, O., Prediger, S., Grob, U., Matull, I., & Hußmann, S. (2017). Remediation for students with mathematics difficulties: An intervention study in middle schools. *Journal of Learning Disabilities*, 50(6), 724–736.
- Moser Opitz, E., Grob, U., Wittich, C., Häsel-Weide, U., & Nührenböcker, M. (2018). Fostering the computation competence of low achievers through cooperative learning in inclusive classrooms: A longitudinal study. *Learning Disabilities: A Contemporary Journal*, 16(1), 19–35.
- Moser Opitz, E., & Schmassmann, M. (2016). Grundoperationen. In U. Heimlich & F. B. Wember (Hrsg.), *Didaktik des Unterrichts im Förderschwerpunkt Lernen. Ein Handbuch für Studium und Praxis* (3. Aufl.). Kohlhammer.
- Pitters, J. (2019). *Faktoren der Entwicklung mathematischer Basiskompetenzen. Eine Untersuchung mit Schüler*innen und Schülern der fünften Jahrgangsstufe an Integrierten Gesamtschulen und Oberschulen*. Carl von Ossietzky Universität Oldenburg.
- Rottmann, T., & Schipper, W. (2002). Das Hunderter-Feld – Hilfe oder Hindernis beim Rechnen im Zahlenraum bis 100? *Journal für Mathematik-Didaktik*, 23, 51–74.

- Scherer, P., Beswick, K., DeBlois, L., Healy, L., & Moser Opitz, E. (2016). Assistance of students with mathematical learning difficulties: how can research support practice?. *ZDM–Mathematics Education*, 48(5), 633–649.
- Schindler, M. (2019). Anzahlerfassung am digitalen Rechenrahmen und Hunderterfeld bei Kindern mit und ohne Rechenschwäche: Vergleich von Eye-Tracking und Lautem Denken. In A. Frank, S. Krauss, & K. Binder (Hrsg.), *Beiträge zum Mathematikunterricht 2019* (S. 1247–1250). WTM.
- Schindler, M., Bader, E., Lilienthal, A. J., Schindler, F., & Schabmann, A. (2019). Quantity recognition in structured whole number representations of students with mathematical difficulties: An eye-tracking study. *Learning Disabilities: A Contemporary Journal*, 17(1), 5–28.
- Schindler, M., & Lilienthal, A. J. (2018). Eye-tracking for studying mathematical difficulties—also in inclusive settings. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Hrsg.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Bd. 4, S. 115–122). PME.
- Schipper, W. (2005). *Lernschwierigkeiten erkennen – verständnisvolles Lernen fördern*. IPN.
- Schleifer, P., & Landerl, K. (2011). Subitizing and counting in typical and atypical development. *Developmental Science*, 14, 280–291.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.
- Tobii Pro. (2017). Eye tracker data quality report: Accuracy, precision and detected gaze under optimal conditions—controlled environment. <https://www.tobii.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-pro-glasses-2-accuracy-and-precision-test-report.pdf>.
- Von Unger, H. (2014). *Partizipative Forschung. Einführung in die Forschungspraxis*. Springer.
- Wartha, S. (2009). Rechenstörungen in der Sekundarstufe. Die Bedeutung des Übergangs von der Grundschule zur weiterführenden Schule. In A. Heinze & M. Grüßing (Hrsg.), *Mathematiklernen vom Kindergarten bis zum Studium. Kontinuität und Kohärenz als Herausforderung für den Mathematikunterricht* (S. 157–180). Waxmann.
- Wartha, S., & Schulz, A. (2011). *Aufbau von Grundvorstellungen (nicht nur) bei besonderen Schwierigkeiten im Rechnen*. IPN.
- Wartha, S., & Schulz, A. (2017). *Rechenproblemen vorbeugen. Grundvorstellungen aufbauen. Zahlen und Rechnen bis 100*. Cornelsen.

Biologie lernen mit Concept Maps – Lässt sich die Expertise im Umgang mit Concept Maps von den Augen ablesen?

Sina Lenski  und Jörg Großschedl 

Inhaltsverzeichnis

6.1	Einleitung	91
6.2	Theoretischer Hintergrund	92
6.3	Hypothesen	95
6.4	Material und Methode	95
6.5	Ergebnisse	100
6.6	Diskussion	101
	Literatur	104

6.1 Einleitung

Empirische Studien legen nahe, dass Lernende bei der Beschäftigung mit biologischen Inhalten besonderer Unterstützung bedürfen (Wendt et al., 2016). Als geeignetes Unterstützungsinstrument werden Concept Maps betrachtet (z. B. Fischer & Mandl, 2000). Sie bestehen aus Begriffen (Konzepten), die über beschriftete Pfeile (Relationen) in eine sinnvolle Beziehung gebracht werden, wobei die Pfeilrichtung die Leserichtung anzeigt (Novak & Gowin, 1984). Concept Maps können entweder als vorgefertigte Informationsquelle (Rezeption von Informationen) bereitgestellt oder von den Lernenden aus den verfügbaren

S. Lenski · J. Großschedl (✉)
Institut für Biologiedidaktik, Universität zu Köln, Köln, Deutschland
E-Mail: j.grossschedl@uni-koeln.de

S. Lenski
E-Mail: s.lenski@uni-koeln.de

fachlichen Informationen selbst konstruiert werden. Sie können dabei unterschiedlich strukturiert sein (z. B. netzartig oder hierarchisch; s. unter „Concept Maps, Eye-Tracking und Expertise“ Abschn. 6.2).

Unerfahrene Lernende scheinen mit der Konstruktion und Rezeption von Concept Maps jedoch häufig überfordert zu sein (z. B. Mintzes et al., 2011). Dies bringt auch der von Blankenship und Dansereau (2000) geprägte Begriff „Map Shock“ zum Ausdruck. Dieser bezeichnet Schwierigkeiten von Lernenden, Informationen aus vorgefertigten Maps zu entnehmen und ihre Topografie sinnvoll zu entschlüsseln. Ob die Konstruktion und Rezeption von Concept Maps besonderer Trainingsmaßnahmen bedarf, ist jedoch bislang weitgehend ungeklärt (Großschedl & Tröbst, 2018).

Die vorliegende Studie geht aus einer Follow-up-Erhebung einer größer angelegten Trainingsstudie (Lenski et al., in Vorbereitung) hervor, in der untersucht wurde, ob sich ein Concept-Map-Training positiv auf eine nachfolgende Lerngelegenheit mit Concept Maps auswirkt. Dabei bildeten die Lernleistung und die kognitive Belastung der Schüler*innen das Forschungsinteresse. Diese Variablen informieren jedoch nur eingeschränkt über die zugrunde liegenden, internal ablaufenden Prozesse der Informationsverarbeitung. Eye-Tracking gewährt hierbei tiefere Einblicke, da sich diese internalen Prozesse der Informationsverarbeitung aus den Blickbewegungscharakteristika von Personen ableiten lassen (Gegenfurtner et al., 2011). Expertise in einer Domäne geht mit Blickbewegungscharakteristika einher, welche eine Unterscheidung zwischen Expert*innen und Noviz*innen ermöglichen (Funke & Sperring, 2006; s. unter „Eye-Tracking und Expertise“ in Abschn. 6.2). Dementsprechend wurde in der vorliegenden Studie untersucht, ob Schüler*innen, die ein Concept-Map-Training erhielten, im Vergleich zu Schüler*innen, die an einem Kontrolltraining teilnahmen, Blickbewegungscharakteristika aufweisen, wie sie für Experten typisch sind. Neben einer objektiven Evaluation von Expertiseunterschieden stellt Eye-Tracking damit Informationen bereit, wie Experten Concept Maps „lesen“. Dies wiederum kann hilfreich sein, um Noviz*innen im richtigen Umgang mit Concept Maps zu trainieren.

6.2 Theoretischer Hintergrund

Eye-Tracking und Expertise

Das *Novizen-Experten-Paradigma* (Funke & Sperring, 2006) postuliert, dass Expert*innen andere Herangehensweisen zur Lösung von Aufgaben verfolgen als Noviz*innen. Blickbewegungsanalysen mittels Eye-Tracking-Systemen werden seit einigen Jahrzehnten in den verschiedensten Anwendungsbereichen zur Aufklärung von Lösungsstrategien und Expertiseunterschieden eingesetzt (Gegenfurtner et al., 2011). Schachexpert*innen erkennen relevante Bereiche

für den nächsten Spielzug beispielsweise schneller als Noviz*innen (Sheridan & Reingold, 2014), und erfahrene Radiologen identifizieren Krebs in einer Mammografie im Gegensatz zu ihren jungen Berufskolleg*innen bereits innerhalb von Sekundenbruchteilen (Kundel et al., 2007).

Gegenfurtner et al. (2011) untersuchten in ihrer Metaanalyse den Stand der Expertiseforschung im Bereich des Betrachtens von Visualisierungen (z. B. Diagramme). Dabei konnten sie zeigen, dass Expert*innen im Vergleich zu Noviz*innen u. a. aufgrund kürzerer Fixationsdauern ($r=0.27$), höherer Fixationszahlen ($r=0.53$) sowie längerer Sakkaden ($r=0.30$) Visualisierungen schneller erfassen können. Bezüglich der Länge von Scanpaths wurde gezeigt, dass steigende Expertise mit kürzeren Scanpaths einhergeht (Krupinski, 2000).

Diese Befunde lassen sich nach Gegenfurtner et al. (2011) durch verschiedene theoretische Annahmen erklären:

1. Auf Grundlage der *Eye-Mind-Hypothese* wird davon ausgegangen, dass Blickbewegungscharakteristika kognitive Prozesse abbilden (Just & Carpenter, 1984).
2. Die Theorie des *Long-Term Working Memory* (Ericsson & Kintsch, 1995) geht davon aus, dass Vorwissen die Kapazitäten für die Informationsverarbeitung aufgrund des Erwerbs von Abrufstrukturen erhöht. Dies ermöglicht es Expert*innen, neue Informationen schneller und effizienter im Langzeitgedächtnis zu encodieren und schließlich gezielt abzurufen. Bezogen auf Eye-Tracking-Messungen sollten Expert*innen sich daher durch eine schnellere Informationsverarbeitung mit kürzeren Fixationsdauern, höheren Fixationszahlen und kürzeren Scanpaths auszeichnen. Zudem ist bekannt, dass Expert*innen Probleme gründlicher analysieren als Noviz*innen (Patel & Groen, 1986), was auf die höheren Kapazitäten für die Informationsverarbeitung zurückgeführt werden könnte. Basierend hierauf wäre zu erwarten, dass sich Expert*innen ausführlicher mit den Stimuli beschäftigen und somit die Gesamtbetrachtungsdauer länger ausfällt als bei Noviz*innen.
3. Das *Holistic Model of Image Perception* (Kundel et al., 2007) beschreibt, dass Expert*innen Informationen zunächst global analysieren und sich dann der feineren Analyse widmen. Dies sollte durch längere Sakkaden für relevante Bereiche zum Ausdruck kommen.
4. Im Rahmen der *Information-Reduction Hypothesis* (Haider & Frensch, 1999) wird zudem postuliert, dass Expert*innen die Menge der verarbeiteten Informationen reduzieren, indem sie selektiv aufgabenrelevante Informationen priorisieren und irrelevante Informationen ausblenden. Übertragen auf die Blickbewegungen sollten Expert*innen demnach geringere Fixationszahlen und kürzere Fixationsdauern in aufgabenirrelevanten Bereichen (z. B. White Spaces; Moser et al., 2020) und dafür höhere Fixationszahlen sowie längere Fixationsdauern in aufgabenrelevanten Bereichen aufweisen.

Concept Maps, Eye-Tracking und Expertise

Bislang existieren nur wenige Eye-Tracking-Studien zur Untersuchung von Expertiseunterschieden bei der Rezeption und Konstruktion von Concept Maps (Amadiou et al., 2009; Amadiou & Salmerón, 2014; Dogusoy-Taylan & Cagiltay, 2014). Anders als in der vorliegenden Studie, wird der Expertisegrad in diesen Studien durch das vorhandene Fachwissen (Vorwissen) und nicht durch die Expertise im Lernen mit Concept Maps definiert. Während sich Dogusoy-Taylan und Cagiltay (2014) sowie Amadiou und Salmerón (2014) der Konstruktion von Concept Maps widmeten, stand bei Amadiou et al. (2009) die Rezeption von hierarchisch und netzartig strukturierten Concept Maps im Vordergrund.

Concept Maps können einerseits mit wenigen Querverbindungen und einer klaren Symmetrie strukturiert sein (hierarchisch) oder über wesentlich mehr Querverbindungen zwischen den Konzepten verfügen und unsymmetrisch (netzartig) strukturiert sein (Novak & Gowin, 1984; Beispiele befinden sich in Abb. 6.1). Es wird angenommen, dass hierarchisch strukturierte Concept Maps ein höheres Ausmaß an Strukturiertheit bieten im Vergleich zu netzartigen Concept Maps, was die Orientierung in den Maps / die Navigation durch die Maps für Noviz*innen erleichtern kann (Blankenship & Dansereau, 2000). Damit einhergehend konnten

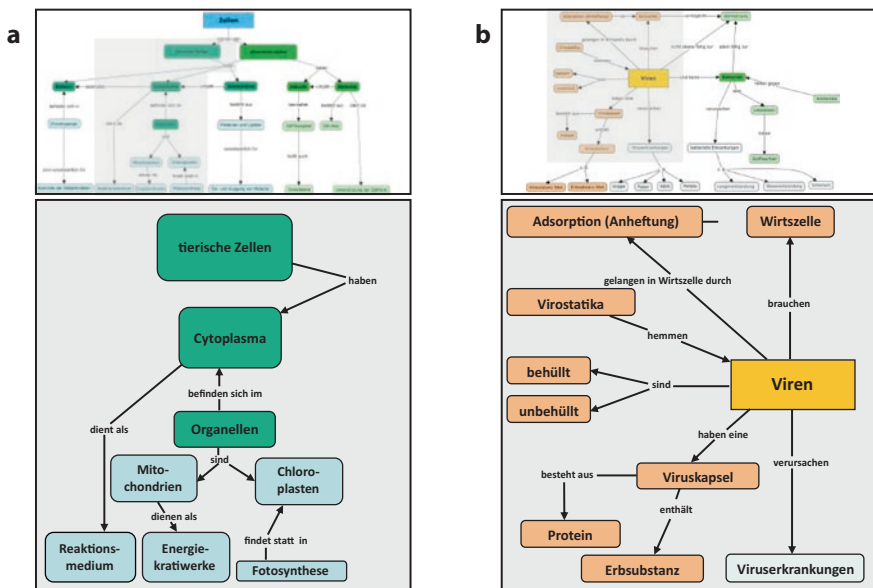


Abb. 6.1 Unterschiedlich strukturierte Concept Maps. Ausschnitte einer hierarchisch strukturierten Concept Map zum Thema „Zellen“ (a) und einer netzartig strukturierten Concept Map zum Thema „Viren“ (b). Die in den Concept Maps (oben) grau markierten Bereiche werden in der unteren Bildhälfte vergrößert dargestellt

Amadiou et al. (2009) zeigen, dass Schüler*innen mit wenig Vorwissen (Faktenwissen) bei der Rezeption von hierarchisch strukturierten Concept Maps einen höheren Lernerfolg erzielen als bei der Rezeption netzartiger Maps. Mit zunehmendem Vorwissen nimmt der Vorteil von den hierarchisch strukturierten Concept Maps jedoch kontinuierlich ab. Laut *Aptitude-Treatment-Interaction Effect* (Cronbach & Snow, 1977) wird ein stark strukturiertes Lernformat Lernenden mit geringeren Fähigkeiten besser gerecht, wohingegen eine starke Vorstrukturierung Lernende mit höheren Fähigkeiten einschränkt.

6.3 Hypothesen

Basierend auf dem skizzierten Stand der Forschung ergeben sich die folgenden Hypothesen:

- Im Vergleich zu Noviz*innen zeichnen sich Expert*innen bei der Rezeption von Concept Maps im Mittel durch eine längere Gesamtbetrachtungsdauer (H1.1), kürzere Fixationsdauern (H1.2), höhere Fixationszahlen (H1.3), kürzere Scanpaths (H1.4) sowie längere Sakkaden (H1.5) aus.
- Expert*innen weisen im Mittel höhere Fixationsdauern auf relevanten Bereichen (Konzepte und Relationen; H2.1) bzw. geringere Fixationsdauern auf irrelevanten Bereichen von Concept Maps (White Spaces; H2.2) auf als Noviz*innen.
- Expert*innen weisen im Mittel sowohl für die hierarchisch strukturierte als auch für die netzartig strukturierte Concept Map eine kürzere Fixationsdauer auf als Noviz*innen (H3.1). Aufgrund der höheren Strukturiertheit, die hierarchisch strukturierte Concept Maps den Lernenden bieten, wird ein *Aptitude-Treatment-Interaction Effect* erwartet: So sollten sich die mittleren Fixationsdauern von Expert*innen und Noviz*innen bei der Rezeption von Informationen aus hierarchisch strukturierten Concept Maps weniger stark voneinander unterscheiden als bei der Rezeption von Informationen aus netzartig strukturierten Concept Maps (H3.2).

6.4 Material und Methode

Stichprobe

Es wurden 62 Schüler*innen der Sekundarstufe I (8. und 9. Jahrgangsstufe) rekrutiert. Aufgrund von Kalibrierfehlern und aus schulorganisatorischen Gründen (Nichterscheinen zum Messtermin oder vorzeitiger Abbruch der Messung) mussten zehn Schüler*innen ausgeschlossen werden. Die finale Stichprobe bestand aus $N=52$ Schüler*innen (69 % weiblich; $M_{\text{Alter}} = 13.9$ Jahre, $SD = 0.52$). Die Gruppe der Expert*innen bestand aus $n=14$ Schüler*innen, während die Gruppe der Noviz*innen $n=38$ Schüler*innen umfasste.

Instrument und Ablauf

In einer vorausgehenden Trainingsstudie erhielt eine Gruppe von Schüler*innen ein intensives Concept-Map-Training (Expert*innen), während eine zweite Gruppe lediglich an einer Kurzeinführung über Concept Maps teilnahm (Noviz*innen; Kontrollgruppe). Im Anschluss an die Trainingsstudie wurden die Schüler*innen eingeladen, an der vorliegenden Eye-Tracking-Studie¹ teilzunehmen.

Nach einer kurzen Begrüßung und Aufklärung über den Ablauf erhielten die Schüler*innen Instruktionen zur Eye-Tracking-Messung. Anschließend wurde das infrarotlichtbasierte Eye-Tracking-System SMI RED250 mobile (SensoMotoric Instruments GmbH, Teltow, Deutschland) auf einem 17-Zoll-Bildschirm (Auflösung 1920×1080 Pixel) für jede Schülerin und jeden Schüler durch 5-Punkt-Kalibrierung kalibriert. Aufgrund einer Accuracy von 0.4° mussten die Kopfbewegungen laut Hersteller nicht eingeschränkt werden. Die Daten wurden bei 60 Hz aufgezeichnet, und der Betrachtungsabstand betrug ca. 40–50 cm. Die Schüler*innen wurden zufällig einem von zwei Settings (A und B) zugeordnet. Beide Settings waren inhaltlich identisch und bestanden aus zwei Concept Maps. Bevor diese den Schüler*innen gezeigt wurden, wurden sie durch einen Einstieg / eine Problemstellung (Abb. 6.2) in einen sinnvollen Kontext eingebettet. Unmittelbar nach der Beschäftigung mit jeder Concept Map folgte ein Kurztest, in dem die Lernenden Fragen über die Inhalte der jeweiligen Concept Map beantworten mussten (Abb. 6.2). Die Kurztests wurden jeweils vor der Rezeption der Concept Maps angekündigt und sollten eine konzentrierte/ernsthafte Auseinandersetzung der Schüler*innen mit den Concept Maps unterstützen. Sie wurden nicht weiter in die Auswertung mit einbezogen und stellten kein Lernerfolgsmaß dar. Die Settings A und B unterschieden sich in der Reihenfolge der Concept Maps. In Setting A wurde zuerst die netzartig strukturierte Concept Map (zum Thema „Viren“) und nachfolgend die hierarchisch strukturierte Concept Map (zum Thema „Zellen“) präsentiert. In Setting B erfolgte die Präsentation in umgekehrter Reihenfolge.

Dieses Design wurde gewählt, um Reihenfolgeeffekte zu vermeiden (Causse et al., 2019). Nachdem die Messung von der Testleitung gestartet worden war, konnten die Schüler*innen die ausgearbeiteten Concept Maps entsprechend ihrer eigenen Logik und ihrem eigenen Zeitbedarf ohne weitere strukturierende Vorgaben selbstgesteuert rezipieren und den Beginn des Kurztests bestimmen. Die Eye-Tracking-Daten wurden lediglich für die Rezeption der Concept Maps, nicht aber für den Einstieg / die Problemstellung und die Kurztests (Abb. 6.2) erhoben (minimale Messdauer: 0.39 min, maximale Messdauer: 4.54 min; s. Gesamtbetrachtungsdauer in Tab. 6.1). Die gesamte Untersuchung nahm etwa 15–20 min

¹ Bei der Umsetzung wurden die ethischen Standards der Deklaration von Helsinki beachtet, u. a. Information der Schüler*innen, Schulleiter*innen und Erziehungsberechtigten, Arbeit mit anonymisierten Daten, risikofreier Charakter der Studie (Taupitz, 2001).

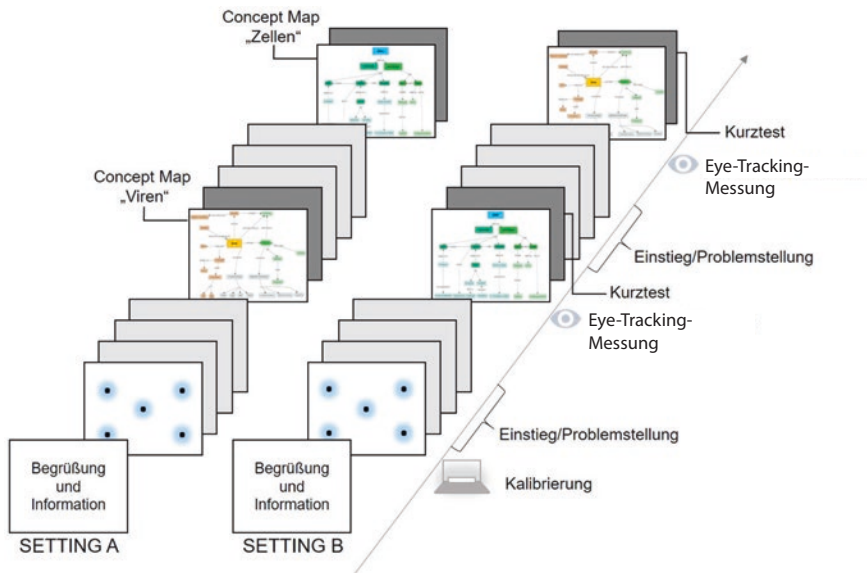


Abb. 6.2 Design und Ablauf der Eye-Tracking-Messung. Die Schüler*innen wurden einem von zwei Settings (A und B) zugeordnet. Die netzartig und die hierarchisch strukturierten Concept Maps zum Thema „Viren“ bzw. „Zellen“ wurden in den beiden Settings in unterschiedlicher Reihenfolge präsentiert

in Anspruch. Um identische Versuchsbedingungen während der Eye-Tracking-Messung zu gewährleisten, wurden der Ablauf der Untersuchung und der Versuchsaufbau standardisiert (z. B. Anpassung der Lichtverhältnisse).

Stimulusmaterial

Das Stimulusmaterial bestand aus zwei Concept Maps, welche basierend auf den allgemeinen Prinzipien von Novak und Gowin (1984) mit dem Softwarepaket CmapTools Version 6.01 ² konstruiert wurden. Die Concept Map „Zellen“ wurde hierarchisch (21 Konzepte und 19 Relationen) und die Concept Map „Viren“ netzartig strukturiert (24 Konzepte und 18 Relationen; Abb. 6.1).

Concept-Map-Training und Kontrolltraining

Das Concept-Map-Training und das Kontrolltraining umfassten insgesamt 135 min und bestanden aus zwei Teilen. Eine 25-minütige Kurzeinführung stellte

²Download unter <https://cmap.ihmc.us..>

Tab. 6.1 Deskriptive Daten von Expert*innen und Noviz*innen

Maß	Expert*innen (<i>n</i> = 14)		Noviz*innen (<i>n</i> = 38)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Gesamtbetrachtungsdauer (ms)	138222.16	9277.33	15072.53	44515.99
Fixationsdauer (ms)	373.43	59.88	456.04	163.15
Fixationsdauer (ms) auf relevanten Bereichen*	379.75	66.45	479.44	172.80
Fixationsdauer (ms) auf irrelevanten Bereichen**	359.16	128.08	366.59	186.44
Fixationsdauer (ms) auf hierarchischen Concept Maps	394.37	76.16	493.48	212.26
Fixationsdauer (ms) auf netzartigen Concept Maps	352.48	48.85	418.61	145.23
Fixationszahl (<i>n</i>)	335.82	171.20	183.49	147.74
Scanpath-Länge (px)	47857.54	21232.87	30308.00	19196.41
Sakkadendauer (ms)	24.82	1.98	41.35	34.03

* Verwendung eines AOI-Sets, welches relevante Bereiche (Relationen und Konzepte) umfasst

** Bereiche außerhalb der relevanten Bereiche (White Spaces; für mehr Informationen s. unter „Datenanalyse“)

den ersten Teil dar (für mehr Informationen s. Lenski et al., [in Vorbereitung](#)). In diesem Teil wurde deklaratives Strategiewissen über Concept Maps vermittelt (u. a. Aufbau und Nutzen von Concept Maps). Im zweiten Teil wurden die Expert*innen in der Nutzung von Concept Maps intensiv trainiert, wobei prozedurales Wissen über die Nutzung von Concept Maps vermittelt wurde. Das verwendete Trainingsmaterial deckt Themen verschiedener nichtbiologischer Inhaltsbereiche ab (Lenski & Großschedl, 2021). Die Noviz*innen nahmen stattdessen an einem Kontrolltraining³ teil, das neben der Kurzeinführung keine weitere Lerngelegenheit für das Lernen mit Concept Maps vorsah.

³Im Kontrolltraining wurde das Spiel „Eierfall“ gespielt. Dazu wurden die Schüler*innen in Gruppen eingeteilt und entwickelten gemeinsam und unter Zeitdruck einen Plan, um ein rohes Ei für einen Wurf aus 4 m vorzubereiten. Ziel war es, kooperatives Arbeiten zu fördern. Gemeinsam sollten Ideen und Lösungen entwickelt werden, um das Ei mit den vorgegebenen Materialien vor dem Sturz zu schützen. Diese Beschäftigung sollte sicherstellen, dass das Kontrolltraining denselben zeitlichen Umfang einnahm wie das Concept-Map-Training.

Datenanalyse

Zur Datenaufbereitung und -analyse wurden die Messungen von Setting A und Setting B mittels BeGaze in einen Datensatz zusammengeführt. In Anlehnung an Lykins et al. (2006) wurde nur die Position des rechten Auges analysiert. IViewX wurde zur Datenerfassung verwendet. Fixierungen wurden unter Verwendung des SMI-Standardalgorithmus auf Dispersionsbasis (100 Pixel und Minstdauer von 80 ms) erkannt.

Zur Testung der Hypothesen H1.1 bis H1.5 wurde die Gesamtbetrachtungsdauer, die Fixationsdauer, die Fixationszahl, die Länge der Scanpaths und die Sakkadenlänge für beide Concept Maps gemittelt.

Um H2.1 zu testen, wurden die Fixationsdauern auf einem zuvor definierten AOI-Set, das ausschließlich relevante Bereiche (Relationen und Konzepte) markierte, analysiert. Hierbei wurden Überlappungen von AOIs vermieden und Abstände zwischen den verschiedenen AOIs gelassen (Holmqvist et al., 2011). Die Bereiche, welche nicht in den AOIs enthalten waren (White Spaces), wurden für die Berechnung der Fixationsdauern auf irrelevanten Bereichen (H2.2) verwendet.

Zur Beantwortung von Hypothese H3.1 und H3.2 wurden die Fixationsdauern auf der netzartigen und der hierarchischen Concept Map getrennt voneinander betrachtet.

Deskriptiv- und inferenzstatistische Auswertung

Obwohl Kolmogorov-Smirnov-Tests nur bei einigen Variablen auf eine Verletzung der Normalverteilung hindeuteten, legten weitere Einschränkungen der Daten (z. B. Verletzung der Varianzhomogenität der Gruppen, stark variierende Gruppengrößen) eine Verwendung nichtparametrischer Verfahren für die Hypothesentestung nahe (Bortz & Lienert, 2008). So wurden die unabhängigen Stichproben der Expert*innen und Noviz*innen mittels Mediantest verglichen. Der Mediantest ist dem Mann-Whitney-U-Test bei unterschiedlich großen Stichproben vorzuziehen, da der Mann-Whitney-U-Test an Schärfe verliert, wenn die zu vergleichenden Stichproben unterschiedlich groß sind (Bortz & Lienert, 2008). Zur Vermeidung einer Alpha-Fehler-Kumulierung durch mehrfaches Testen innerhalb derselben Stichprobe wurde eine Bonferroni-Holm-Korrektur vorgenommen (Hemmerich, 2021). Dies betrifft in der vorliegenden Studie all jene Vergleiche, die sich auf Gruppenunterschiede in der Fixationsdauer beziehen (H1.2, H2.1, H2.2, H3.2 und H3.2) und somit der Untersuchung übergreifender Hypothesen mittels verwandter Variablen dienen. Zur Beantwortung von H3.2 wurde im Rahmen eines Solomon-2-Gruppen-Planes ein Mediantest für Paardifferenzen mit der unabhängigen Variable „Gruppe“ (Concept-Map-Training vs. Kontrolltraining) durchgeführt. Um die Fixationsdauern auf hierarchisch und netzartig strukturierten Concept Maps vergleichbar zu machen, wurden die zugehörigen Variablen zunächst z-standardisiert (Bortz & Schuster, 2011). Anschließend wurde

die abhängige Variable berechnet, die sich aus der Differenz der z-standardisierten Variablen ergab. Effektstärken wurden nach Cohen (1984) berechnet und interpretiert.

6.5 Ergebnisse

H1.1 Vergleicht man die Gesamtbetrachtungsdauer von Expert*innen und Noviz*innen fällt auf, dass Expert*innen die Concept Maps insgesamt länger betrachteten (Tab. 6.1) als Noviz*innen. Dieser Unterschied erwies sich hypothesenkonform als statistisch bedeutsam, wobei der Effekt als groß zu interpretieren ist: $\chi^2(1) = 8.17$, $p = .004$, $d = 1.63$.

H1.2 Expert*innen wiesen bei der Rezeption der Concept Maps eine kürzere Fixationsdauer auf als Noviz*innen (Tab. 6.1). Dieser Unterschied erwies sich jedoch als nicht signifikant: $\chi^2[1] = 4.79$, $p = .087$ (nach Bonferroni-Holm korrigiert). Somit muss diese Hypothese verworfen werden.

H1.3 Hypothesenkonform wiesen Expert*innen eine höhere Fixationszahl (Tab. 6.1) auf als Noviz*innen: $\chi^2(1) = 7.92$, $p = .005$, $d = 0.95$. Der gefundene Effekt ist als groß zu bewerten.

H1.4 Konträr zu dieser Hypothese zeigten Expert*innen signifikant längere Scanpaths in Pixeln (px; Tab. 6.1) als Noviz*innen: $\chi^2(1) = 7.92$, $p = .005$, $d = 0.87$. Auch hier liegt ein großer Effekt vor.

H1.5 Bezüglich der Sakkadendauer (Tab. 6.1) traten keine signifikanten Unterschiede zwischen Expert*innen und Noviz*innen auf: $\chi^2[1] = 0.23$, $p = .630$. Somit muss die Hypothese an dieser Stelle wiederum verworfen werden.

H2.1 Wider Erwarten unterschieden sich Expert*innen und Noviz*innen nicht hinsichtlich ihrer mittleren Fixationsdauern auf den relevanten Bereichen der Concept Maps (Tab. 6.1): $\chi^2[1] = 4.79$, $p = .087$ (nach Bonferroni-Holm korrigiert).

H2.2 Ebenfalls erwartungswidrig unterschieden sich Expert*innen und Noviz*innen nicht hinsichtlich ihrer mittleren Fixationsdauern auf den irrelevanten Bereichen (White Spaces) der Concept Maps (Tab. 6.1): $\chi^2[1] = 4.10$, $p = .075$ (nach Bonferroni-Holm korrigiert).

H3.1 Expert*innen wiesen auf hierarchisch strukturierten Concept Maps eine signifikant kürzere Fixationsdauer auf als Noviz*innen (Tab. 6.1): $\chi^2[1] = 7.92$, $p = .025$ (nach Bonferroni-Holm korrigiert), $d = 0.61$. Dies konnte auch für die netzartig strukturierte Concept Map ermittelt werden (Tab. 6.1): $\chi^2[1] = 7.92$, $p = .025$ (nach Bonferroni-Holm korrigiert), $d = 0.61$.

H3.2 Es wurde erwartet, dass sich die mittleren Fixationsdauern von Expert*innen und Noviz*innen bei der Rezeption von Informationen aus hierarchisch strukturierten Concept Maps weniger stark voneinander unterscheiden als bei der Rezeption von Informationen aus netzartig strukturierten Concept Maps. Ein Mediantest für Paardifferenzen ergab, dass dies nicht der Fall war, weshalb wir unsere Hypothese verwerfen: $\chi^2[1]=2.44$, $p=.118$ (nach Bonferroni-Holm korrigiert).

Insgesamt ist anzumerken, dass die Werte der Noviz*innen für die meisten Parameter eine größere Streuung aufwiesen als die der Expert*innen (Tab. 6.1).

6.6 Diskussion

Interpretation der Ergebnisse

In dieser Studie wurde untersucht, ob sich die während eines Concept-Map-Trainings erworbene Expertise bei der Rezeption von Concept Maps in Blickbewegungscharakteristika manifestiert. Hierzu wurden die Blickbewegungen von trainierten Schüler*innen (Expert*innen) und untrainierten Schüler*innen (Noviz*innen) verglichen (H1.1 bis H1.5). Weiterhin wurde untersucht, ob sich Expert*innen und Noviz*innen in der Betrachtung relevanter und irrelevanter Concept-Map-Bereiche (H2.1 und H2.2) und unterschiedlich strukturierter Concept Maps unterscheiden (H3.1 und 3.2).

Der Vergleich ausgewählter Blickbewegungen ergab hypothesenkonform, dass Expert*innen die Concept Maps länger betrachteten (H1.1) und sich durch höhere Fixationszahlen (H1.3) auszeichneten. Dieser Befund steht in Einklang mit metaanalytischen Befunden von Gegenfurtner et al. (2011) zum Betrachten von Visualisierungen. Die längere Betrachtungsdauer lässt darauf schließen, dass sich Expert*innen ausführlicher mit den bereitgestellten Concept Maps auseinandersetzen. Die höhere Fixationszahl lässt sich mithilfe der Theorie des *Long-Term Working Memory* erklären, wonach sich Expert*innen durch eine schnellere Informationsverarbeitung auszeichneten als Noviz*innen. Der erwartete Unterschied in der Fixationsdauer zwischen Expert*innen und Noviz*innen deutet sich zwar an, erwies sich jedoch als nicht signifikant (H1.2).

Anders als erwartet, wiesen Expert*innen längere Scanpaths auf als Noviz*innen (H1.4). Dies steht im Widerspruch der Theorie des *Long-Term Working Memory* und der hieraus abgeleiteten Hypothese. Gestützt wird der Befund jedoch durch andere empirische Studien (z. B. Jaarsma et al., 2014), bei denen ebenfalls längere Scanpaths mit steigendem Expertisegrad einhergingen. Geschlussfolgert wurde hier, dass Noviz*innen nicht „das ganze Bild“, sondern nur Teile davon verarbeiteten. Dies würde mit den Annahmen des *Holistic Model of Image Perception* übereinstimmen und könnte in der vorliegenden Untersuchung ein Anzeichen für einen Map Shock sein, welcher sich laut Blankenship und Dansereau (2000) in einer unvollständigen Betrachtung der Map äußern kann.

Im Widerspruch zu diesem Modell und der hierauf basierenden Hypothese steht das Ergebnis, dass Expert*innen kürzere Sakkadendauern aufwiesen als Noviz*innen (H1.5). Kok und Jarodzka (2017) merkten allerdings an, dass die Sakkadendauer von Expert*innen und Noviz*innen je nach Art der visuellen Aufgabe variiert, und stellten die Eignung der Sakkadendauer zur Bestimmung des Expertisegrades infrage.

In Anlehnung an die *Information-Reduction Hypothesis* wurde erwartet, dass sich Expert*innen und Noviz*innen in ihren mittleren Fixationsdauern auf relevanten (H2.1) und irrelevanten Bereichen der Concept Maps (H2.2) unterscheiden. In beiden Fällen konnten jedoch keine entsprechenden Unterschiede beobachtet werden. Um Expertiseunterschiede sichtbar zu machen, reichte es wahrscheinlich nicht aus, lediglich zwischen Teilen der Concept Map und White Spaces zu unterscheiden. Hier hätten eventuell in Bezug auf die Inhalte der Concept Maps genauere Aufgabenstellungen formuliert und dazu passende AOIs definiert werden müssen.

Betrachtet man die Fixationsdauern von Expert*innen und Noviz*innen auf den hierarchischen und netzartigen Concept Maps, wird eine erwartungskonform kürzere Fixationsdauer der Expert*innen sichtbar (H3.1). Dieser Unterschied wird jedoch schwächer, wenn keine differenzierte Analyse nach Concept-Map-Struktur erfolgt (H1.2). Da sich dies weder theoretisch noch empirisch herleiten lässt, könnte es durch die geringe Stichprobengröße und hiermit verbundene geringe Power der verwendeten statistischen Verfahren erklärt werden.

Hinsichtlich der mittleren Fixationsdauern auf hierarchisch und netzartig strukturierten Concept Maps wurde ein *Aptitude-Treatment-Interaction Effect* erwartet (H3.2). Mittlere Fixationsdauern von Expert*innen und Noviz*innen sollten sich bei der Rezeption von Informationen aus hierarchisch strukturierten Concept Maps weniger stark voneinander unterscheiden als bei der Rezeption von Informationen aus netzartig strukturierten Concept Maps. Dies war jedoch nicht der Fall, was darauf hindeutet, dass Noviz*innen nicht stärker von der hierarchischen Struktur profitierten als Expert*innen.

Zu erwähnen ist, dass viele Messwerte der Noviz*innen eine größere Streuung aufwiesen als die der Expert*innen. Einerseits könnte dies als Artefakt interpretiert werden, das sich aus einer größeren Stichprobe von Noviz*innen ergibt. Andererseits könnte eine geringere Streuung der Messwerte auf Expertenseite auch auf geringere Kompetenzunterschiede in der Rezeption von Informationen aus Concept Maps hindeuten, was auf die Wirksamkeit des Trainings hindeuten könnte. Sollten sich die Befunde dieser Studie in nachfolgenden Studien erhärten, könnten auf die Expertise abgestimmte Hinweise zur praktischen Nutzung von Concept Maps abgeleitet werden.

Limitationen der vorliegenden Studie

Eine Limitation dieser Studie stellt die unterschiedlich große Stichprobenverteilung in den Gruppen dar ($n=38$ Noviz*innen vs. $n=14$ Expert*innen). Diese ergab sich hauptsächlich durch ein unterschiedlich großes Interesse an der Studienteilnahme. Wie bei vielen Eye-Tracking-Studien ist einschränkend zu erwähnen, dass die Power der angewandten statistischen Verfahren vergleichsweise gering ist (Gegenfurtner et al., 2011). Wir empfehlen daher eine Replikation der vorliegenden Studie an einer größeren und für die Populationen der Expert*innen und Noviz*innen repräsentativeren Stichprobe. Da in dieser Studie durch Vorabgespräche mit den Lehrkräften sichergestellt wurde, dass die Inhalte der Concept Maps noch nicht im Unterricht behandelt wurden, jedoch kein Vorwissenstest durchgeführt wurde, sollte in künftigen Studien das fachliche Vorwissen einbezogen werden.

Fazit und Implikationen für zukünftige Forschung

Im Rahmen der vorliegenden Studie wurden wichtige Hinweise dafür gefunden, dass Schüler*innen, die ein Concept-Map-Training besuchten, eine größere Expertise in der Rezeption von Informationen aus Concept Maps aufwiesen als Schüler*innen, die im Rahmen eines Kontrolltrainings lediglich eine Kurzeinführung erhielten. Dieses Ergebnis deckt sich auch in den Leistungsdaten der vorausgehenden Trainingsstudie (Lenski et al., in Vorbereitung) – trainierte Schüler*innen erzielten eine höhere Lernleistung als untrainierte Schüler*innen. Gemeinsam mit diesen Befunden, untermauert die vorliegende Studie die Annahme, dass ein Concept-Map-Training die Expertise in der Rezeption von Concept Maps steigert und diese Expertise auch notwendig ist, um Concept Maps gewinnbringend nutzen zu können. Sie leistet somit einen wichtigen Beitrag zur Debatte um die Wirkung und Relevanz von Trainingsmaßnahmen für die Nutzung von Concept Maps und trägt zu einem wichtigen Erkenntnisfortschritt an der Schnittstelle zwischen Expertiseforschung und anwendungsorientierter biologie-didaktischer Forschung bei.

Neben quantitativen Untersuchungen mit repräsentativeren Stichprobenzahlen könnten sich zukünftige Studien der qualitativen Untersuchungen der Unterschiede in den Blickbewegungen von Expert*innen und Noviz*innen bei der Rezeption von Concept Maps widmen. Hier wäre insbesondere interessant, ob sich bei Noviz*innen weitere Hinweise für einen Map Shock finden lassen. Es könnten zudem nach für Experten typische Scanpaths gesucht werden mit dem Ziel, diese in Form von sog. Eye Movement Modeling Examples (van Marlen et al., 2018) Noviz*innen zur Verfügung zu stellen. Perspektivisch könnten Noviz*innen so bei der Navigation durch die Inhalte der Concept Maps unterstützt werden.

Das in dieser Studie eingesetzte Concept-Map-Konstruktionstraining samt aller Materialien ist online verfügbar (Lenski & Großschedl, 2021).

Danksagung und Angaben zur finanziellen Förderung Diese Studie ist Teil des Projekts „Biologie lernen durch Concept Mapping: Bedeutung eines Lernstrategietrainings für kognitive Belastung, kognitive Prozesse und Lernleistung“, welches von der Deutschen Forschungsgemeinschaft (DFG) gefördert wurde (Projektnummer GR 4763/2). Für die Unterstützung bei der Datenerhebung bedanken wir uns bei unseren Abschlussarbeitskandidat*innen Benjamin Bayer, Sven Nam Karsten und Regina Drevs.

Literatur

- Amadiou, F., & Salmerón, L. (2014). Concept maps for comprehension and navigation of hyper-texts. In D. Ifenthaler & R. Hanewald (Hrsg.), *Digital knowledge maps in education* (S. 41–59). Springer.
- Amadiou, F., van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction*, 19, 376–386. <https://doi.org/10.1016/j.learninstruc.2009.02.005>.
- Blankenship, J., & Dansereau, D. F. (2000). The effect of animated node-link displays on information recall. *Journal of Experimental Education*, 68(4), 293–308. <https://doi.org/10.1080/00220970009600640>.
- Bortz, J., & Lienert, G. A. (2008). *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben*. Springer.
- Bortz, J., & Schuster, C. (2011). *Statistik für Human- und Sozialwissenschaftler*. Springer.
- Causse, M., Lancelot, F., Maillant, J., Behrend, J., Cousy, M., & Schneider, N. (2019). Encoding decisions and expertise in the operator's eyes: Using eye-tracking as input for system adaptation. *International Journal of Human-Computer Studies*, 125, 55–65.
- Cohen, A. (1984). Effect of girdling date on fruit size of Marsh Seedless grapefruit. *Journal of Horticultural Science*, 59(4), 567–573.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington Publishers.
- Dogusoy-Taylan, B., & Cagiltay, K. (2014). Cognitive analysis of experts' and novices' concept mapping processes: An eye tracking study. *Computers in Human Behavior*, 36, 82–93.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102(2), 211.
- Fischer, F., & Mandl, H. (2000). Strategiemodellierung mit Expertenmaps. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (S. 37–54). Hogrefe.
- Funke, J., & Spring, M. (2006). Methoden der Denk- und Problemlöseforschung. In J. Funke (Hrsg.), *Denken und Problemlösen (= Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Band 8)* (S. 675–744). Hogrefe.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552.
- Großschedl, J., & Tröbst, S. (2018). Biologie lernen durch Concept Mapping: Bedeutung eines Lernstrategietrainings für kognitive Belastung, kognitive Prozesse und Lernleistung – Kurzdarstellung des DFG-Projekts. *Zeitschrift für Didaktik der Biologie (ZDB) – Biologie Lehren und Lernen*, 22(1), 20–30.
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 172–190.
- Hemmerich, W. (2021). *Rechner zur Adjustierung des α -Niveaus*. StatistikGuru. <https://statistik-guru.de/rechner/adjustierung-des-alphaniveaus.html>. Zugegriffen: 11. Jan 2021

- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Jaarsma, T., Jarodzka, H., Nap, M., van Merriënboer, J. J. G., & Boshuizen, H. P. A. (2014). Expertise under the microscope: Processing histopathological slides. *Medical Education*, 48(3), 292–300.
- Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Hrsg.), *New methods in reading comprehension research* (S. 151–182). Erlbaum.
- Kok, E. M., & Jarodzka, H. (2017). Before your very eyes: The value and limitations of eye tracking in medical education. *Medical Education*, 51(1), 114–122.
- Krupinski, E. A. (2000). The importance of perception research in medical imaging. *Radiation Medicine – Medical Imaging and Radiation Oncology*, 18(6), 329–334.
- Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology*, 242(2), 396–402.
- Lenski, S., Elsner, S., & Großschedl, J. (eingereicht). *Comparing construction and study of concept maps – An intervention study on cognitive, metacognitive and emotional effects of training & learning*. *Frontiers in Education*.
- Lenski, S., Elsner, S. & Großschedl, J. (in Vorbereitung). *Working with concept maps in biology learning – Development and evaluation of a construction and a study training*.
- Lykins, A. D., Meana, M., & Kambe, G. (2006). Detection of differential viewing patterns to erotic and non-erotic stimuli using eye-tracking methodology. *Archives of Sexual Behavior*, 35(5), 569–575.
- Mintzes, J. J., Canas, A., Coffey, J., Gorman, J., Gurley, L., Hoffman, R., et al. (2011). Comment on „Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping“ [Technical Comment]. *Science*, 334, 453–c–453–d.
- Moser, T., Lohmeyer, Q., Meboldt, M., Distler, O., & Becker, M. O. (2020). Visual assessment of digital ulcers in systemic sclerosis analysed by eye tracking: Implications for wound assessment. *Clinical and Experimental Rheumatology*, 38(3), 137–139.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press.
- Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10(1), 91–116.
- Sheridan, H., & Reingold, E. M. (2014). Expert vs. novice differences in the detection of relevant information during a chess game: Evidence from eye movements. *Frontiers in Psychology*, 5, 941.
- Taupitz, J. (2001). Die Neufassung der Deklaration von Helsinki des Weltärztebundes vom Oktober 2000. *MedR – Medizinrecht*, 19(6), 277–286.
- van Marlen, T., van Wermeskerken, M., Jarodzka, H., & van Gog, T. (2018). Effectiveness of eye movement modeling examples in problem solving: The role of verbal ambiguity and prior knowledge. *Learning and Instruction*, 58, 274–283.
- Wendt, H., Bos, W., Selter, C., Köller, O., Schwippert, K., & Kasper, D. (2016). *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann.

Haben Kinder mit Rechenschwierigkeiten auch Schwierigkeiten im Bereich Geometrie? – Eine Eye-Tracking- Studie zur Identifikation von Dreiecken bei Schüler*innen im inklusiven Kontext

Anna Lisa Simon, Benjamin Rott und Maike Schindler

Inhaltsverzeichnis

7.1	Einleitung	107
7.2	Theoretischer Hintergrund und Forschungsfrage	108
7.3	Studie	112
7.4	Ergebnisse	115
7.5	Diskussion und Ausblick	120
	Literatur	122

7.1 Einleitung

„Die beste Prävention von Rechenstörungen ist ein guter Mathematikunterricht.“
(Schipper, 2005, S. 27)

Rechenschwierigkeiten (RS) sind ein wichtiges Thema in der mathematik-
didaktischen Forschung und Praxis, denn ohne Fördermaßnahmen können sie bis
ins Erwachsenenalter bestehen bleiben und „einen nachhaltigen negativen Einfluss

A. L. Simon (✉) · B. Rott · M. Schindler
Universität zu Köln, Köln, Deutschland
E-Mail: anna.simon@uni-koeln.de

B. Rott
E-Mail: benjamin.rott@uni-koeln.de

M. Schindler
E-Mail: maike.schindler@uni-koeln.de

auf die gesamte Bildungs-, Berufs- und Persönlichkeitsentwicklung“ haben (von Aster, 2017, S. 482). Daher besteht ein verstärktes Interesse, Unterstützungsbedarfe von Schüler*innen mit RS zu ermitteln, um diese angemessen unterrichten und fördern zu können (Schipper, 2005).

RS werden als Schwierigkeiten in den arithmetischen (Basis-)Bereichen verstanden. Das Mathematiklernen und -lehren in der Grundschule beschränkt sich jedoch nicht ausschließlich auf die Arithmetik. Einen weiteren wichtigen mathematischen Inhaltsbereich stellt die Geometrie dar, bei dem im Primarstufenbereich ebene Figuren, wie Dreiecke, eine zentrale inhaltliche Rolle einnehmen (z. B. MSW NRW, 2008). Eines der Ziele des Geometrieunterrichts der Primarstufe ist es, dass Schüler*innen ebene Figuren benennen und Fachbegriffe wie „Seite“ und „Ecke“ zu deren Beschreibung verwenden können (MSW NRW, 2008). In verschiedenen Studien wurde untersucht, wie gut Kinder geometrische Figuren, z. B. Dreiecke, identifizieren können: Während Kindergartenkinder prototypische Dreiecke und prototypische Nichtdreiecke (z. B. Quadrat) bereits identifizieren können (Aslan & Aktaş Arnas, 2007; Tsamir et al., 2008), orientieren sich Kinder zum Ende der Grundschulzeit oder zu Beginn der weiterführenden Schule zunehmend an definierenden Eigenschaften und sind in der Lage, zunehmend auch untypische Dreiecke richtig zu identifizieren (Satlow & Newcombe, 1998). Ob dies auch Kindern mit RS gelingt oder ob diese hier vermehrt Schwierigkeiten haben, ist jedoch noch nicht erforscht.

Im Rahmen des Forschungsprojekts „Mathematik im inklusiven Kontext: Lernausgangslagen von Kindern mittels Eye-Tracking Diagnostizieren“ – kurz MI-T – wird daher der Frage nachgegangen, ob und ggf. wie sich Schüler*innen mit RS bei der Identifikation ebener Figuren von Schüler*innen ohne RS unterscheiden. Das vorliegende Kapitel präsentiert eine empirische Studie aus dem Projekt MI-T und geht der Forschungsfrage nach, ob sich Schüler*innen mit und ohne RS bei der Identifikation von Dreiecken unterscheiden. Hierfür wird neben der Analyse von Fehlerraten und Bearbeitungszeiten bei der Identifikation von Dreiecken Eye-Tracking genutzt.

7.2 Theoretischer Hintergrund und Forschungsfrage

Rechenschwierigkeiten

Die Begriffe, mit denen Schwierigkeiten beim Mathematiklernen benannt werden, sind breit gefächert. Je nach Verständnis und wissenschaftlicher Ausrichtung ist von Dyskalkulie, Rechenstörung, mathematischer Lernschwäche, Rechenschwäche und anderen Begriffen die Rede (Moser Opitz, 2007; Moser Opitz et al., 2017; Scherer et al., 2016). So vielfältig wie die Begriffe sind auch die Probleme bei der Eingrenzung und Bestimmung des Gegenstands (Landerl et al., 2017). Im vorliegenden Kapitel wird der Begriff „Rechenschwierigkeiten“ verwendet, womit – in Anlehnung an Moser Opitz et al. (2017) und Scherer et al. (2016) – Schwierigkeiten in den arithmetischen Basisbereichen gemeint sind. Solche Schwierigkeiten können sich sowohl auf konzeptueller (z.

B. beim Verständnis des Dezimalsystems und von Stellenwerten) als auch auf prozeduraler Ebene (z. B. beim flexiblen Einsatz von Rechenstrategien) zeigen. Mathematische Schwierigkeiten in der Sekundarstufe können oft auf Probleme im Verständnis bestimmter grundlegender arithmetischer Konzepte zurückgeführt werden, die in der Grundschule gelehrt werden (Moser Opitz et al. 2017; Scherer et al. 2016). Untersuchungen zu Kindern mit mathematischen Schwierigkeiten deuten u. a. auf Schwierigkeiten mit dem visuell-räumlichen Arbeitsgedächtnis sowie dem visuell-räumlichen Kurzzeitgedächtnis hin (z. B. Szucs et al., 2013).

Geometrie

Neben der Arithmetik (Zahlen und Operationen) stellt die Geometrie (Raum und Form) einen der zentralen Inhaltsbereiche der Mathematik der Primarstufe dar (z. B. MSW NRW, 2008). Ein zentrales, allgemeines Ziel des Geometrieunterrichts ist „mit Hilfe der Geometrie die (Um-)Welt zu erschließen“ (Weigand et al., 2018, S. 4). Geometrische Begriffe tragen dazu bei, die (Um-)Welt ordnen und interpretieren zu können (Weigand et al., 2018). Einer der zentralen Inhalte innerhalb des Bereichs „Raum und Form“ ist „Ebene Figuren“ (MSW NRW, 2008): Ziel ist es, dass Schüler*innen ebene Figuren, z. B. Dreiecke, benennen und Fachbegriffe wie „Seite“ und „Ecke“ zu deren Beschreibung verwenden können. Mathematische Begriffe werden durch Definitionen bestimmt. Tall und Vinner (1981) beschreiben diese formale „concept definition“ als die Wörter, die zur Spezifizierung eines mathematischen Begriffs oder Konzepts benutzt werden und die in der mathematischen Community anerkannt wird.

Über die Kenntnis einer Definition hinaus spielt zum Verständnis eines Begriffs das „concept image“ (Tall und Vinner, 1981) eine wichtige Rolle. Unter dem „concept image“ versteht man „the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes“ (Tall und Vinner, 1981, S. 152). Laut Tall und Vinner wird dieses im Laufe der Jahre durch Erfahrungen aufgebaut und verändert. Lernende bilden Vorstellungen über den „Begriffsinhalt“, den „Begriffsumfang“ und das „Begriffsnetz“ aus (Weigand et al., 2018, S. 85). Dies geschieht darüber, dass sie „Merkmale oder Eigenschaften eines Begriffs und deren Beziehungen“ kennen, „einen Überblick über die Gesamtheit aller Objekte erhalten, die unter einem Begriff zusammengefasst werden“ und „Beziehungen des Begriffs zu anderen Begriffen aufzeigen können“ (Weigand et al., 2018, S. 85). Bei geometrischen Figuren gibt es die definierenden Eigenschaften, dass sie „geschlossen“ sind und eine „bestimmte Anzahl“ von „geraden Seiten“ haben (Satlow & Newcombe, 1998). Darüber hinaus gibt es nichtdefinierende Eigenschaften, die beliebig verändert werden können, ohne den Status einer Figur als gültiger Repräsentant zu beeinträchtigen (Satlow & Newcombe, 1998).

Dreiecke

Alle Repräsentanten des Begriffs Dreieck sind mathematisch äquivalent, d. h., sie erfüllen die Begriffsdefinition und enthalten alle definierenden Eigenschaften (Hershkowitz, 1989). Unterschiede zwischen diesen Repräsentanten bestehen visuell, also im Aussehen. Dabei werden jene Repräsentanten, die typisch und

höchst repräsentativ sind und die jeweiligen Prototypen abbilden (gleichseitige und gleichschenklige Dreiecke), von Schüler*innen besser erkannt als andere (Hershkowitz, 1989; Tsamir et al., 2008; zur Prototypentheorie s. Rosch, 1973). Die prototypischen Repräsentanten enthalten neben den definierenden Eigenschaften weitere spezifische, nichtdefinierende Eigenschaften, welche dominant sind und die Aufmerksamkeit des Betrachters auf sich ziehen (Hershkowitz, 1989, S. 73). Im Falle von gleichseitigen und gleichschenkligen Dreiecken sind dies Symmetrie, Seitenverhältnis und Ausrichtung. Alle anderen Dreiecke, die davon in den nichtdefinierenden Eigenschaften wie bspw. „skewness“ (die Asymmetrie bzw. Schrägheit einer Figur), „aspect ratio“ (das Seitenverhältnis innerhalb der Figur) und „orientation“ (die Ausrichtung der Figur) abweichen, werden als nicht-prototypisch verstanden (Aslan & Aktaş Arnas, 2007; Clements et al., 1999).

Identifikation von Dreiecken

Studien zur Identifikation von Dreiecken beziehen sich häufig auf den vorschulischen Bereich. Kinder im Vorschulalter orientieren sich vorwiegend visuell am Aussehen einer Figur. Sie identifizieren die prototypischen Dreiecke schnell als Repräsentanten und lehnen häufig fälschlicherweise untypische Dreiecke ab (Aslan & Aktaş Arnas, 2007; Tsamir et al., 2008; Unterhauser, 2020). Nichtdreiecke werden fälschlicherweise als Dreiecke akzeptiert, bspw. wenn diese einem prototypischen Dreieck ähnlich sehen, aber etwa gerundete Seiten oder Ecken haben (Clements et al., 1999; Satlow & Newcombe, 1998; Tsamir et al., 2008; Unterhauser, 2020). Das Ablehnen untypischer Dreiecke als Dreiecke zeigt, dass mathematisch irrelevante, nichtdefinierende Eigenschaften die kindlichen Entscheidungen beeinflussen. So identifizieren Kinder Dreiecke auch nicht als solche, weil sie für „zu lang“ befunden werden oder weil „die Spitze nicht nach oben zeigt“ (Clements & Sarama, 2000, S. 483). Ein möglicher Grund könnte in der Anzahl der nichtdefinierenden Eigenschaften des Dreiecks liegen, die dazu führen könnten, dass die Kinder im Vergleich zu anderen geometrischen Formen, wie Kreisen und Quadraten, mehr Fehler machen. So nimmt der Erfolg der Kinder bei der Identifikation einer Figur ab, wenn die Anzahl der nichtdefinierenden Eigenschaften der Figur zunimmt (Aslan & Aktaş Arnas, 2007).

Es zeigt sich, dass die Identifikation von Dreiecken den Kindern mehr Schwierigkeiten bereitet als die Identifikation anderer ebener Figuren, wie Kreise oder Quadrate (Aslan & Aktaş Arnas, 2007; Clements et al., 1999). Diese anderen (prototypischen) ebenen Figuren werden von Kindergartenkindern schnell und sicher als Nichtdreiecke identifiziert (Aslan & Aktaş Arnas, 2007; Tsamir et al., 2008). Aslan und Aktaş Arnas (2007) haben aufgezeigt, dass bereits im Vorschulalter die Dreiecke mit zunehmendem Alter weniger aufgrund ihres Erscheinungsbilds beurteilt werden, wobei Kinder Figuren als visuell Ganzes wahrnehmen und sich dabei stark auf Prototypen stützen („visual level“ nach van Hiele; Clements & Battista, 1992a), und der Anteil eigenschaftsbezogener Antworten zunimmt („descriptive/analytic level“ nach van Hiele; Clements & Battista, 1992a; Tsamir et al., 2008). Generell zeigt sich ein Trend, dass Kinder mit steigendem Alter mehr Figuren richtig identifizieren (Clements et al., 1999). Kinder zum Ende der Grund-

schulzeit oder zu Beginn der weiterführenden Schule orientieren sich zunehmend an definierenden Eigenschaften. Sie lehnen Nichtdreiecke häufiger ab als untypische Dreiecke (Satlow & Newcombe, 1998). Ein Vergleich zweier Studien von Vorschulkindern und Sechstklässler*innen deutet darauf hin, dass trotz des großen Altersunterschieds die Unterschiede in Bezug auf die Identifikation nicht so groß waren wie erwartet (Clements & Sarama, 2000): Bei der Bearbeitung derselben Aufgaben zeigten sich Erfolgsraten von 64 % bei Vorschulkindern und von 81 % bei Sechstklässler*innen.

Die existierenden Studien zur Identifikation von Dreiecken zeigen auf, welche Dreiecke bzw. Nichtdreiecke wie gut von Kindern erkannt werden. Mögliche Unterschiede zwischen Kindern mit und ohne RS hinsichtlich der Dreiecksidentifikation sind unseres Wissens nach noch nicht erforscht. Kinder mit RS haben Schwierigkeiten in den arithmetischen Basisbereichen (Moser Opitz et al., 2017; Scherer et al., 2016). Mögliche Beeinträchtigungen des visuell-räumlichen Arbeits- und Kurzzeitgedächtnisses bei Kindern mit mathematischen Schwierigkeiten (Szucs et al., 2013) könnten bei Kindern mit RS auch mit Schwierigkeiten bei Aufgaben zur Identifikation ebener Figuren einhergehen. Im Rahmen des Forschungsprojekts MI-T wird daher der Frage nachgegangen, ob sich Schüler*innen mit und ohne RS bei der Identifikation ebener Figuren¹ – hier speziell Dreiecke – unterscheiden.

Es wird der folgenden Forschungsfrage nachgegangen:

*Unterscheiden sich Schüler*innen mit und ohne RS bei der Identifikation von Dreiecken?*

Um dies zu erforschen, werden neben den Fehlerraten die Bearbeitungszeiten der Schüler*innen untersucht. Des Weiteren wird Eye-Tracking verwendet, um weitere Einblicke in die Bearbeitungsprozesse zu erlangen. Die Eye-Tracking-Forschung stützt sich vor allem auf die sog. Eye-Mind-Hypothese (Just & Carpenter, 1976; Schindler & Lilienthal, 2019), die annimmt, dass das, was das Auge fixiert, und die Informationsverarbeitung im Gehirn eng miteinander verbunden sind (Holmqvist et al., 2011). Im Allgemeinen kann direkt über den visuellen Fokus die sog. offene Aufmerksamkeit beobachtet werden (Carrasco, 2011), d. h. jener Bereich, der foveal fokussiert wird. Die Dauer der Fixationen kann die Wahrnehmungsaufnahme und -verarbeitung widerspiegeln (Holmqvist et al., 2011, S. 379). Dabei gibt es verschiedene Fixationsdauern und unterschiedliche Interpretationen dieser Fixationsdauern (Holmqvist et al., 2011, S. 378). In der Regel werden längere Fixationsdauern mit einer „deeper and more effortful cognitive processing“ (Holmqvist et al., 2011, S. 381) in Verbindung gebracht. In

¹Der Inhalt der Erhebung wurde in Anlehnung an den Grundschullehrplan NRW ausgewählt. Da die Kinder zu Beginn des 5. Schuljahres sind, kann so sichergestellt werden, einen Inhalt zu untersuchen, den alle Kinder im Unterricht kennengelernt haben.

der vorliegenden Studie wurden die Mean Fixation Duration (MFD) und die Total Fixation Duration (TFD) erhoben. Die MFD steht für die durchschnittliche Fixationsdauer auf einer Area of Interest (AOI) (Tobii Pro AB, 2020). Die TFD steht für die Summe aller Fixationsdauern auf einer AOI (Tobii Pro AB, 2020). Längere Summen von Fixationsdauern auf einer AOI – d. h. in der vorliegenden Studie eine längere „dwell time“ (Holmqvist et al., 2011, S. 389) – können ein Hinweis auf „difficulty in extracting ... information“ (Holmqvist et al., 2011, S. 388) sein.

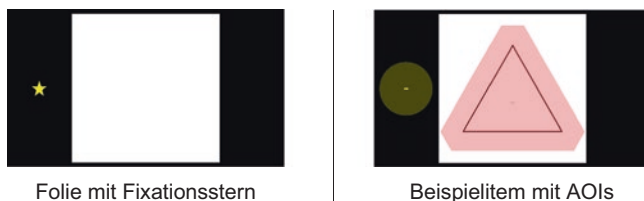
Frühere Studien (z. B. Schindler & Lilienthal, 2019) haben aufgezeigt, dass Eye-Tracking im Bereich der Geometrie Einblicke in Bearbeitungsprozesse von Lernenden ermöglicht. In anderen mathematischen Inhaltsbereichen hat sich gezeigt, dass Eye-Tracking die Möglichkeit bietet, Unterschiede zwischen Kindern mit und ohne RS (z. B. Schindler et al., 2019) zu erfassen und dass Eye-Tracking – gerade für Schüler*innen mit RS – ein großes diagnostisches Potenzial besitzt (Anzahlerfassung: Schindler, 2019; Zahlenstrahlaufgaben: Simon & Schindler, 2020). Auch wenn Eye-Tracking in den letzten Jahren an Bedeutung in der mathematikdidaktischen Forschung gewonnen hat, gibt es im Bereich der Geometrie generell bisher nur relativ wenige Studien, die Eye-Tracking nutzen (vgl. Lilienthal & Schindler, 2019; Strohmaier et al., 2020).

7.3 Studie

Stichprobe

Es wurden $N = 174$ Kinder einer inklusiven Gesamtschule in NRW zu Beginn der 5. Klasse untersucht. Zunächst wurde ein standardisierter Rechentest, der Heidelberger Rechentest (HRT) (Haffner et al., 2005), mit allen Kindern durchgeführt. Der HRT zieht zur Diagnostik von RS den Prozentrang (PR) heran, der Auskunft darüber gibt, „wie viel Prozent der Vergleichsstichprobe (Eichstichprobe) gleich gut oder schlechter abgeschnitten haben“ (Haffner et al., 2005, S. 19). Basierend auf den Ergebnissen des HRT wurden bei 72 Schüler*innen RS ($PR \leq 10$) festgestellt, bei 58 wurden keine RS ($PR \geq 26$) festgestellt. 44 Kinder befanden sich im „Risikobereich“ (Haffner et al., 2005, S. 20) für RS ($11 \leq PR \leq 25$) und können somit weder der Gruppe der Kinder mit noch der Gruppe der Kinder ohne RS zugeordnet werden. Dementsprechend wurden die Kinder im Risikobereich für die Analysen nicht berücksichtigt. Zusätzlich wurden Daten von acht Schüler*innen mit RS und von zehn Schüler*innen ohne RS aufgrund geringer Trackinggenauigkeit² nicht berücksichtigt. Im vorliegenden Kapitel werden die Gruppen der Kinder mit RS ($n = 64$) und ohne RS ($n = 48$) vergleichend betrachtet.

² Wir haben als Grenzwert eine Trackinggenauigkeit (Accuracy + Precision) $< 2^\circ$ gewählt.

**Abb. 7.1** Items aus der Studie

	prototypisch	nichtprototypisch		
Dreiecke		gedrehte Dreiecke mit prototypischem Seitenverhältnis 	unregelmäßige Dreiecke mit spitzem bzw. stumpfem Winkel 	gleichschenklige Dreiecke mit extrem spitzem bzw. stumpfem Winkel
#	3	7	4	4
Nicht-dreiecke		Nichtdreiecke mit Lücke 	Nichtdreiecke mit gerundeten Seiten bzw. Ecken 	Nichtdreiecke mit mehr als drei Ecken
#	0	4	6	6

Abb. 7.2 Auswahl an Figuren, die in der Erhebung verwendet wurden (# = Anzahlen der Items); die Kombination „Nichtdreiecke × prototypisch“ (ausgegraut) wurde bewusst nicht in die Studie mit einbezogen

Durchführung und Technik

In Einzeluntersuchungen in einem ruhigen Raum der Schule wurde jedem Kind eine Serie von 34 Dreiecken³ gezeigt (Repräsentanten und Nichtrepräsentanten; Abb. 7.2). Vor jedem präsentierten Dreieck sollten die Kinder einen Fixationsstern fixieren (Abb. 7.1), um sicherzugehen, dass alle Schüler*innen die Aufgabenbearbeitung von demselben Ort aus beginnen. Die Schüler*innen wurden gefragt, ob die präsentierte Figur ein Dreieck sei, und mündlich instruiert, so schnell und korrekt wie möglich mit „ja“ oder „nein“ zu antworten. Die Dreiecke wurden auf einem 24-Zoll-Full-HD-Bildschirm (Bildwiederholungsrate 60 Hz, Auflösung 1920 × 1080 Pixel) präsentiert. Die Größe des prototypischen Dreiecks in Abb. 7.2 betrug in Weite und Höhe 680 × 589 Pixel. Es wurde der bildschirmbasierte Eye-Tracker Tobii Pro X3-120 (120 Hz, Accuracy 0.4°, binokular, infra-

³Im Rahmen des Forschungsprojekts MI-T wurden verschiedene Aufgaben zu unterschiedlichen mathematischen Inhaltsbereichen genutzt. Die in diesem Kapitel thematisierten Aufgaben zur Identifikation von Dreiecken stellten den ersten Aufgabenblock dar.

rot) verwendet. Es gab keine Bewegungseinschränkungen für die Schüler*innen. Vor jeder Datenerhebung wurde eine 5-Punkt-Kalibrierung durchgeführt. Die durchschnittliche Accuracy betrug 0.7° . Die Entfernung der Schüler*innen vom Monitor betrug ca. 60 cm. Mithilfe der Software Tobii Pro Lab (Version 1.118.21001) wurden die Bearbeitungszeiten, die MFD und die TFD erhoben (Tobii-I-VT-Fixationsfilter: Geschwindigkeitsgrenzwert: $30^\circ/\text{s}$, minimale Dauer einer Fixation: 60 ms). Die für die Analyse der Eye-Tracking-Daten relevanten AOIs sind die jeweilige AOI auf dem Dreieck und die AOI an der Stelle, wo der Fixationsstern eingeblendet war (Abb. 7.1). Die AOI des Fixationssterns wurde konstruiert, um zu überprüfen, ob es diesbezüglich Unterschiede zwischen Kindern mit und ohne RS gibt. Es wurde ein größerer Bereich als die tatsächliche Größe der Dreiecke/des Sterns gewählt, um Trackingungenauigkeiten zu berücksichtigen. Die mündlichen Antworten der Kinder wurden mit einem Audioaufnahmegerät aufgezeichnet. Insgesamt gab es 3808 Aufgabenbearbeitungen. Aufgrund einer Aufgabenbearbeitung, bei der ein Kind keine Antwort gab, flossen in die Analyse der Fehlerrate 3807 Aufgabenbearbeitungen ein. Für die Analyse der Bearbeitungszeiten, der MFD und der TFD standen aufgrund technikbedingter Datenverluste 3768 Aufgabenbearbeitungen zur Verfügung.

Items

Die Entwicklung der Items (Abb. 7.2) erfolgte in Anlehnung an Aslan und Aktaş Arnas (2007), Clements et al. (1999), Satlow und Newcombe (1998) und Tsamir et al. (2008). Die Einteilung der Items in prototypische und nichtprototypische Dreiecke und nichtprototypische Nichtdreiecke erfolgte in Anlehnung an Tsamir et al. (2008)⁴. Dabei sind die „prototypischen Dreiecke“ die – wie vorangehend aufgeführt – gleichseitigen und gleichschenkligen Dreiecke (wie in Abb. 7.2 abgebildet). Die „nichtprototypischen Dreiecke“ sind untypische Repräsentanten, die in nichtdefinierenden Eigenschaften (Ausrichtung, Schrägheit/Asymmetrie und Seitenverhältnis) vom prototypischen Dreieck abweichen. Die „nichtprototypischen Nichtdreiecke“ haben eine dreiecksähnliche Form, bei der definierende Eigenschaften eines Dreiecks verletzt sind (Abb. 7.2): Diese Figuren haben bspw. gerundete Ecken/Seiten, sind nicht geschlossen oder haben mehr als drei Ecken/Seiten. In einem nächsten Schritt wurde die Einteilung der Dreiecke weiter differenziert in prototypische Dreiecke, gedrehte Dreiecke mit prototypischem Seitenverhältnis, unregelmäßige Dreiecke mit spitzem bzw. stumpfem Winkel, gleichschenklige Dreiecke mit extrem spitzem bzw. stumpfem Winkel, Nichtdreiecke mit Lücke, Nichtdreiecke mit gerundeten Seiten bzw. Ecken, Nichtdreiecke mit mehr als drei Ecken (Abb. 7.2). Auf das Einbeziehen prototypischer Nichtdreiecke, also anderer prototypischer geometrischer Formen, wie bspw. ein Quadrat, wurde verzichtet, da diese bereits von Kindergartenkindern einfach und sicher als

⁴Tsamir und Kolleg*innen verwenden den Begriff „intuitiv“. Wir behalten an dieser Stelle den Begriff „prototypisch“ bei.

Nichtdreiecke erkannt werden (Aslan & Aktaş Arnas, 2007; Tsamir et al., 2008). Die präsentierten Items wurden in einer zufälligen Reihenfolge präsentiert.

Im Folgenden werden die gedrehten Dreiecke mit prototypischem Seitenverhältnis als „gedrehte Dreiecke“, die unregelmäßigen Dreiecke mit spitzem bzw. stumpfem Winkel als „unregelmäßige Dreiecke“ und die gleichschenkligen Dreiecke mit extrem spitzem bzw. stumpfem Winkel als „extrem spitze bzw. schmale Dreiecke“ bezeichnet.

Methode und statistische Analyse

Anhand der Antworten der Kinder wurde die Fehlerrate ermittelt. Darüber hinaus wurden die Bearbeitungszeiten, die MFD und die TFD (mit der Software Tobii Pro Lab) erhoben. Für die statistischen Analysen wurde die Statistik- und Analysesoftware IBM SPSS 26 verwendet. Um die beiden Gruppen hinsichtlich der Fehlerraten, Bearbeitungszeiten, der MFD und der TFD zu vergleichen, wurden ANOVAs mit Messwiederholung mit der Art der Messung (prototypische Dreiecke, gedrehte Dreiecke, unregelmäßige Dreiecke, extrem spitze bzw. schmale Dreiecke, Nichtdreiecke mit Lücke, Nichtdreiecke mit gerundeten Seiten bzw. Ecken, Nichtdreiecke mit mehr als drei Ecken) als Innersubjektfaktor und der Gruppe (mit RS, ohne RS⁵) als Zwischensubjektfaktor durchgeführt. Aufgrund der unterschiedlichen Anzahl an Items, die unter den verschiedenen Dreiecksarten zusammengefasst werden, wurden für die Analysen die Mittelwerte der Items genutzt. Dadurch konnten bei den Analysen alle Dreiecksarten gleich stark gewichtet werden. Die Dreiecksarten werden in den Abbildungen zur Ergebnisdarstellung, wie nachfolgend dargestellt, durch Zahlen abgebildet: prototypische Dreiecke (1), gedrehte Dreiecke (2), unregelmäßige Dreiecke (3), extrem spitze bzw. schmale Dreiecke (4), Nichtdreiecke mit Lücke (5), Nichtdreiecke mit gerundeten Seiten bzw. Ecken (6), Nichtdreiecke mit mehr als drei Ecken (7).

7.4 Ergebnisse

Fehlerraten

Bei 3807 Aufgabenbearbeitungen machten die Kinder insgesamt 679 Fehler (17.84 %). Dies ist vergleichbar mit der Fehlerrate von Sechstklässler*innen (19 %) einer früheren Studie zur Dreiecksidentifikation von Clements und Battista (1992b, zitiert in Clements & Sarama, 2000). Während in der vorliegenden Studie die mittlere Fehlerrate der Kinder ohne RS 16.73 % betrug (273 Fehler/1632 Aufgabenbearbeitungen), lag diese bei den Kindern mit RS bei 18.67 % (406 Fehler/2175 Aufgabenbearbeitungen). Es zeigte sich trotz der leicht unterschiedlichen Gesamtfehlerrate und der teilweise unterschiedlichen Fehlerraten in Bezug

⁵Die Schüler*innen mit RS werden in den folgenden Abbildungen als „RS“ und die Schüler*innen ohne RS als „non-RS“ bezeichnet.

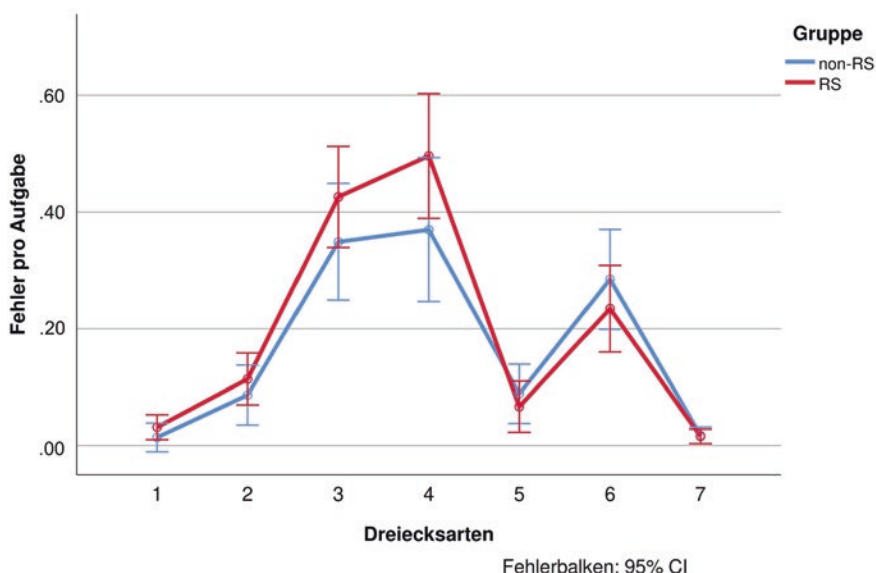


Abb. 7.3 Durchschnittliche Fehlerraten unterschieden nach Dreiecksarten

auf einzelne Dreiecksarten (Abb. 7.3) keine statistisch signifikante Interaktion zwischen den Gruppen und den Dreiecksarten (Greenhouse-Geisser $F(1.68, 185.12) = 1.49$, $p = .230$, $\eta_p^2 = .01$). Es gab keinen signifikanten Unterschied zwischen beiden Gruppen – weder für alle Dreiecke zusammen ($F(1, 110) = 1.63$, $p = .205$, $\eta_p^2 = .02$) noch für die einzelnen Dreiecksarten (für alle Dreiecksarten $p > .05$; adjustierte p -Werte nach Bonferroni-Holm). Es gab einen signifikanten Effekt für die Dreiecksarten (Greenhouse-Geisser $F(1.68, 185.12) = 49.98$, $p < .001$, $\eta_p^2 = .31$). Dass es signifikante Unterschiede zwischen den verschiedenen Dreiecksarten gibt, entspricht der Erwartung und wird aus Platzgründen nicht vertiefend dargestellt.

Bearbeitungszeiten

Für die Bearbeitungszeiten gab es keine signifikante Interaktion zwischen den Gruppen und den Dreiecksarten (Greenhouse-Geisser $F(4.15, 452.61) = 2.25$, $p = .061$, $\eta_p^2 = .02$). Es gab einen signifikanten Effekt für die Dreiecksarten (Greenhouse-Geisser $F(4.15, 452.61) = 13.77$, $p < .001$, $\eta_p^2 = .11$). Es trat kein signifikanter Effekt für die Gruppen ($F(1, 109) = 2.43$, $p = .122$, $\eta_p^2 = .02$) auf. Es zeigten sich in den Post-hoc-Analysen jedoch signifikante Unterschiede zwischen den Schüler*innen mit und ohne RS bei einer nichtprototypischen Dreiecksart. Die Schüler*innen mit RS hatten eine signifikant längere Bearbeitungszeit bei den gedrehten Dreiecken (2) ($t(109) = -2.74$, $p = .049$; adjustierter p -Wert nach Bonferroni-Holm). Die gruppenspezifischen Bearbeitungszeiten für die einzelnen Dreiecksarten sind in Abb. 7.4 dargestellt.

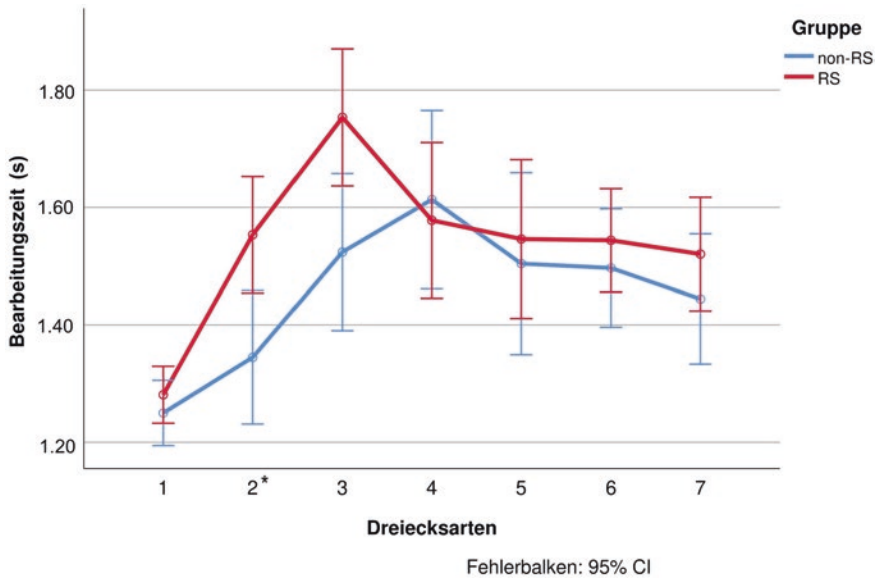


Abb. 7.4 Durchschnittliche Bearbeitungszeiten unterschieden nach Dreiecksarten (signifikante Gruppenunterschiede sind mit einem * gekennzeichnet)

TFD auf den Dreiecken

Für die TFD auf den Dreiecken (Abb. 7.5) gab es keine signifikante Interaktion zwischen den Gruppen und den Dreiecksarten (Greenhouse-Geisser $F(4.44, 484.17) = 1.66, p = .151, \eta_p^2 = .02$). Es gab einen signifikanten Effekt für die Dreiecksarten (Greenhouse-Geisser $F(4.44, 484.17) = 17.32, p < .001, \eta_p^2 = .14$). Es zeigte sich kein signifikanter Gruppenunterschied – weder für alle Dreiecke zusammen ($F(1, 109) = 3.27, p = .073, \eta_p^2 = .03$) noch für die einzelnen Dreiecksarten (für alle Dreiecksarten $p > .05$; adjustierte p -Werte nach Bonferroni-Holm).

MFD auf den Dreiecken

Für die MFD auf den Dreiecken (Abb. 7.6) gab es keine signifikante Interaktion zwischen den Gruppen und den Dreiecksarten (Greenhouse-Geisser $F(4.21, 459.05) = .31, p = .881, \eta_p^2 < .01$). Es gab keinen signifikanten Effekt für die Dreiecksarten (Greenhouse-Geisser $F(4.21, 459.05) = 1.87, p = .110, \eta_p^2 = .02$). Es zeigte sich kein signifikanter Gruppenunterschied – weder für alle Dreiecke zusammen ($F(1, 109) = 1.16, p = .283, \eta_p^2 = .01$) noch für die einzelnen Dreiecksarten (für alle Dreiecksarten $p > .05$; adjustierte p -Werte nach Bonferroni-Holm).

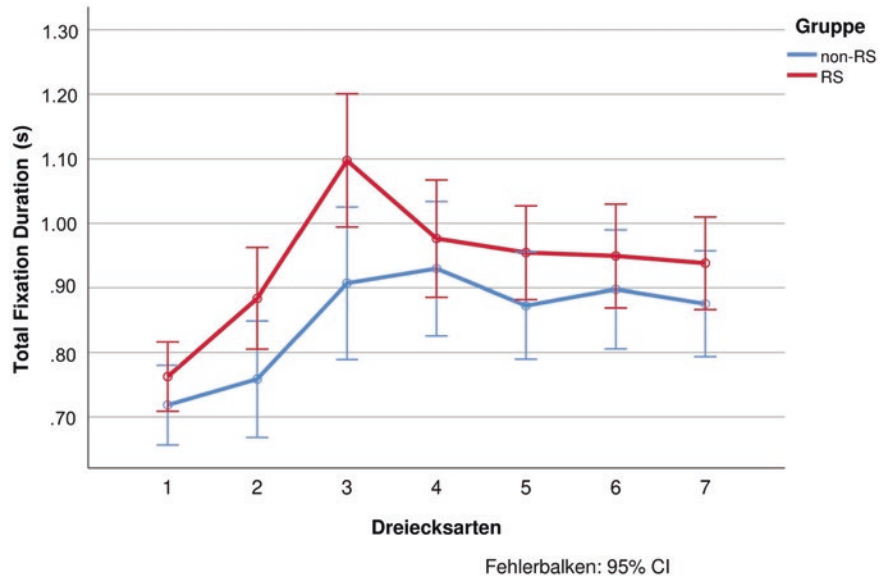


Abb. 7.5 Durchschnittliche TFD auf den Dreiecken unterschieden nach Dreiecksarten

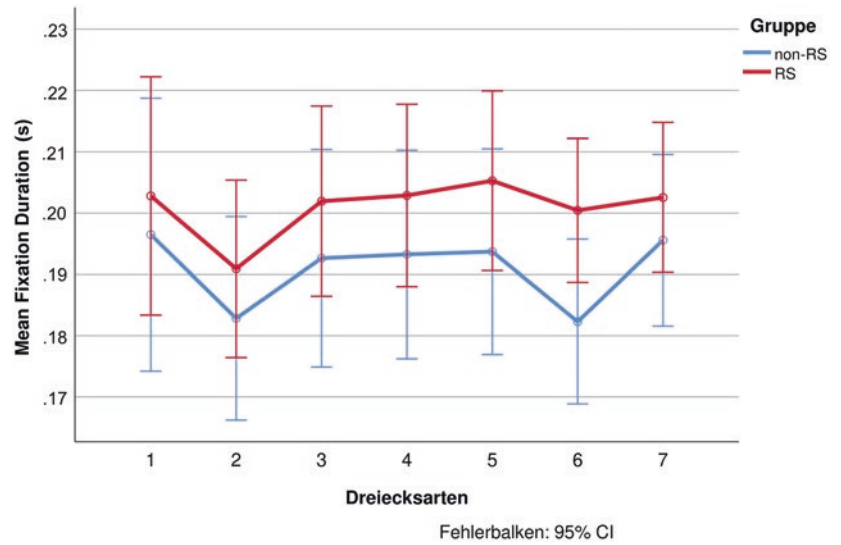


Abb. 7.6 Durchschnittliche MFD auf den Dreiecken unterschieden nach Dreiecksarten

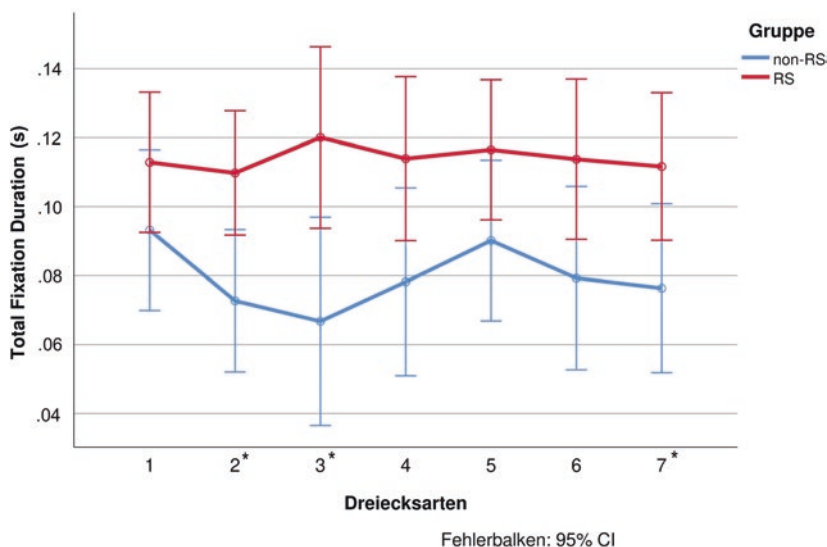


Abb. 7.7 Durchschnittliche TFD auf dem Ort des Fixationssterns unterschieden nach Dreiecksarten (signifikante Gruppenunterschiede sind mit einem * gekennzeichnet)

TFD auf dem Ort des Fixationssterns

Für die TFD auf dem Ort des Fixationssterns (Abb. 7.7) gab es keine signifikante Interaktion zwischen den Gruppen und den Dreiecksarten (Greenhouse-Geisser $F(4.71, 513.60) = 1.53, p = .183, \eta_p^2 = .01$). Es gab keinen signifikanten Effekt für die Dreiecksarten (Greenhouse-Geisser $F(4.71, 513.60) = 1.25, p = .288, \eta_p^2 = .01$). Für alle Aufgaben zusammen zeigte sich ein signifikanter gruppenspezifischer Unterschied ($F(1, 109) = 5.28, p = .023, \eta_p^2 = .05$). Die Schüler*innen mit RS hatten signifikant längere TFD auf dem Ort des Fixationssterns als die Kinder ohne RS. Unterschiede nach Dreiecksarten (Abb. 7.7) zeigten sich hier bei den gedrehten Dreiecken (2) ($t(109) = -2.68, p = .009$), den unregelmäßigen Dreiecken (3) ($t(109) = -2.64, p = .010$) und den Nichtdreiecken mit mehr als drei Ecken (7) ($t(109) = -2.15, p = .034$).

MFD auf dem Ort des Fixationssterns

Für die MFD auf dem Ort des Fixationssterns (Abb. 7.8) gab es keine signifikante Interaktion zwischen den Gruppen und den Dreiecksarten (Greenhouse-Geisser $F(4.96, 540.86) = 1.19, p = .313, \eta_p^2 = .01$). Es gab keinen signifikanten Effekt für die Dreiecksarten (Greenhouse-Geisser $F(4.96, 540.86) = 1.58, p = .166, \eta_p^2 = .01$). Für alle Aufgaben zusammen zeigte sich ein signifikanter gruppenspezifischer Unterschied ($F(1, 109) = 11.02, p = .001, \eta_p^2 = .09$). Die Schüler*innen mit RS hatten eine signifikant längere MFD auf dem Ort des Fixationssterns als die Kinder ohne RS. Unterschiede zeigten sich hier bei fünf

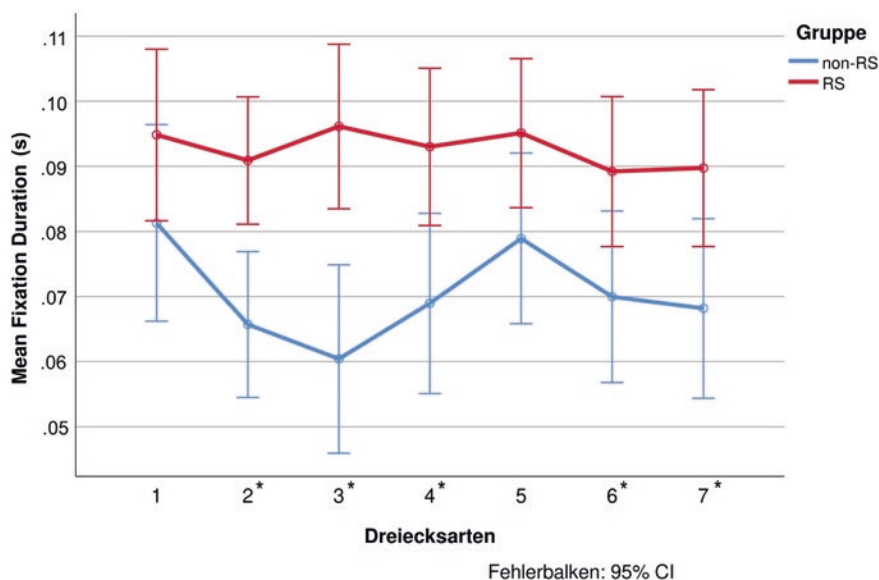


Abb. 7.8 Durchschnittliche MFD auf dem Ort des Fixationssterns unterschieden nach Dreiecksarten (signifikante Gruppenunterschiede sind mit einem * gekennzeichnet)

Dreiecksarten: (2) ($t(109) = -3.36$, $p = .001$), (3) ($t(109) = -3.69$, $p < .001$), (4) ($t(105.65) = -2.75$, $p = .007$), (6) ($t(109) = -2.18$, $p = .032$) und (7) ($t(109) = -2.33$, $p = .022$).

7.5 Diskussion und Ausblick

Das Ziel der vorliegenden Studie war es zu untersuchen, ob Schüler*innen mit RS, bei denen also Schwierigkeiten im arithmetischen Bereich festgestellt wurden, sich auch bei der Identifikation von Dreiecken von Kindern ohne RS unterscheiden. Damit wurde der übergeordneten Fragestellung nachgegangen, ob Kinder mit RS auch in mathematischen Bereichen wie der Geometrie Schwierigkeiten und Unterstützungsbedarfe haben. Dazu wurden bei $n = 64$ Kindern mit RS und $n = 48$ Kindern ohne RS Fehlerraten, Bearbeitungszeiten, MFD und TFD verschiedener AOIs bei der Identifikation von Dreiecken vergleichend analysiert.

Empirische Erkenntnis

In den *Fehlerraten* zeigten sich keine signifikanten Gruppenunterschiede zwischen den Kindern mit und ohne RS. Hinsichtlich der *Bearbeitungszeiten* zeigte sich, dass die Kinder mit und ohne RS sich insgesamt nicht unterschieden – die Kinder mit RS hatten jedoch bei einer Dreiecksart (gedrehten prototypischen Dreiecken)

signifikant längere Bearbeitungszeiten als die Kinder ohne RS. Die Analyse der *MFD* und *TFD auf den Dreiecken* zeigte jedoch auf, dass die Kinder mit RS – in dem Fall, in dem die Bearbeitungszeiten sich signifikant unterschieden, – die Dreiecke nicht signifikant länger fixierten als die Kinder ohne RS. Hinweise auf eine „deeper and more effortful cognitive processing“ (Holmqvist et al., 2011, S. 381) der Kinder mit RS wurden somit nicht gefunden. Die Analyse der *MFD* und *TFD auf dem Ort des Fixationssterns* (wo dieser zuvor eingeblendet war) zeigte hingegen signifikante Gruppenunterschiede auf. Die Kinder mit RS hatten längere Fixationsdauern auf dem Ort des Fixationssterns, bevor sie zum Dreieck schauten. Dies zeigte sich insgesamt über alle Dreiecksarten hinweg und auch insbesondere für jene Dreiecksart, bei der sich der signifikante Gruppenunterschied in den Bearbeitungszeiten gezeigt hatte. Dies deutet auf eine längere Reaktionszeit hin (Fischer et al., 2004).

Methodische Erkenntnis

Bei den Aufgabebearbeitungen dieser Studie, mit einer durchschnittlich sehr kurzen Bearbeitungszeit von ca. 1.5 s, scheint die Reaktionszeit einen großen Teil der Bearbeitungszeit auszumachen. Dementsprechend müssen Gruppenunterschiede in Bearbeitungszeiten bei derart kurzen Aufgabendauern mit Vorsicht betrachtet werden. Die Bearbeitungszeit scheint bei dieser Art der Aufgaben kein guter Indikator für inhaltliche Unterschiede bei der Aufgabebearbeitung zu sein. Das Eye-Tracking, mit der Möglichkeit, die „Verweildauer“ auf verschiedenen Elementen der Aufgaben zu betrachten, bietet Möglichkeiten der näheren Untersuchung.

Ausblick und weiterführende Forschung

Zusammenfassend zeigten sich keine Unterschiede bei der Identifikation von Dreiecken zwischen den Schüler*innen, bei denen im Vorfeld der Erhebung RS festgestellt wurden, und den Schüler*innen ohne RS. Weiterführende Forschung sollte hier untersuchen, ob sich diese Ergebnisse bestätigen lassen oder ob sich Unterschiede zwischen Kindern mit und ohne RS bei der Identifikation ebener Figuren, wie von Dreiecken, zeigen. Studien im Bereich der Geometrie (Schindler & Lilienthal, 2019) und in anderen mathematischen Inhaltsbereichen (z. B. Schindler & Lilienthal, 2018; Schindler et al., 2019) konnten zeigen, dass der Einsatz von Eye-Tracking sich dazu eignet, Vorgehensweisen von Schüler*innen zu analysieren: Aus *gaze overlaid*-Videos, in denen die Blicke des jeweiligen Kindes als wandernder Punkt in einem Video visualisiert sind, können mitunter Vorgehensweisen identifiziert werden. Dies erwies sich gerade für Kinder mit RS als besonders aufschlussreich (z. B. Schindler, 2019; Simon & Schindler, 2020). Für die Dreiecksidentifikation sollte ebenso in zukünftiger Forschung ergründet werden, welche Vorgehensweisen die Kinder nutzen und ob es Unterschiede in der Nutzung dieser Vorgehensweisen zwischen Kindern mit und ohne RS gibt. Dies könnte dazu beitragen, die Unterstützungsbedarfe von Schüler*innen mit RS besser zu ermitteln, um diese angemessen unterrichten und fördern zu können (Schipper, 2005).

Literatur

- Aslan, D., & Aktaş Arnas, Y. (2007). Three- to six-year-old children's recognition of geometric shapes. *International Journal of Early Years Education*, 15(1), 83–104.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51, 1484–1525.
- Clements, D. H., & Battista, M. T. (1992a). Geometry and spatial reasoning. In D. A. Grouws (Hrsg.), *Handbook of research on mathematics teaching and learning* (S. 420–464). Macmillan.
- Clements, D. H., & Battista, M. T. (1992b). *The Development of a Logo-Based Elementary School Geometry Curriculum*. Final Report: NSF Grant No. MDR-8651668. Buffalo, N.Y./Kent, Ohio: State University of New York at Buffalo/Kent State University.
- Clements, D. H., & Sarama, J. (2000). Young children's ideas about geometric shapes. *Teaching Children Mathematics*, 6(8), 482–488.
- Clements, D. H., Swaminathan, S., Zeitler Hannibal, M. A., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education*, 30, 192–212.
- Fischer, M. H., Warlop, N., Hill, R. L., & Fias, W. (2004). Oculomotor bias induced by number perception. *Experimental Psychology*, 51(2), 91–97.
- Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *Heidelberger Rechentest (HRT 1–4)*. Hogrefe.
- Hershkowitz, R. (1989). Visualization in geometry - Two sides of the coin. *Focus on Learning Problems in Mathematics*, 11, 61–76.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8, 441–480.
- Landerl, K., Vogel, S., & Kaufmann, L. (2017). *Dyskalkulie: Modelle, Diagnostik, Intervention*. Ernst Reinhardt.
- Lilienthal, A. J., & Schindler, M. (2019). Eye tracking research in mathematics education: A PME literature review. In M. Graven, H. Venkat, A.A. Essien & P. Vale (Hrsg.), *Proceedings of the 43rd Conference of the IGPME* (Bd. 4, S. 62). PME.
- Moser Opitz, E. (2007). *Rechenschwäche/Dyskalkulie. Theoretische Klärungen und empirische Studien an betroffenen Schüler*innen und Schülern*. Haupt.
- Moser Opitz, E., Freeseemann, O., Prediger, S., Grob, U., Matull, I., & Hußmann, S. (2017). Remediation for students with mathematics difficulties: An intervention study in middle schools. *Journal of learning disabilities*, 50(6), 724–736.
- MSW NRW (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen). (2008). *Lehrplan Mathematik für die Grundschulen des Landes Nordrhein-Westfalen*. Ritterbach Verlag.
- Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, 4, 328–350.
- Satlow, E., & Newcombe, N. (1998). When is a triangle not a triangle? Young children's conceptions of geometric shapes. *Cognitive Development*, 13, 547–559.
- Scherer, P., Beswick, K., Deblois, L., Healy, L., & Moser Opitz, E. (2016). Assistance of students with mathematical learning difficulties: How can research support practice? *ZDM Mathematics Education*, 48(5), 633–649.
- Schindler, M. (2019). Anzahlerfassung am digitalen Rechenrahmen und Hunderterfeld bei Kindern mit und ohne Rechenschwäche: Vergleich von Eye-Tracking und Lautem Denken. In *Beiträge zum Mathematikunterricht 2019*. WTM.
- Schindler, M., Bader, E., Lilienthal, A. J., Schindler, F., & Schabmann, A. (2019). Quantity recognition in structured whole number representations of students with mathematical difficulties: An eye-tracking study. *Learning Disabilities: A Contemporary Journal*, 17(1), 5–28.
- Schindler, M., & Lilienthal, A. J. (2018). Eye-tracking for studying mathematical difficulties – also in inclusive settings. In Bergqvist, E., Österholm, M., Granberg, C., & Sumpter, L. (Hrsg.), *Proceedings of the 42nd Conference of the IGPME* (Bd. 4, S. 115–122). PME.

- Schindler, M., & Lilienthal, A. J. (2019). Domain-specific interpretation of eye-tracking data: Towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics*, 101, 123–139.
- Schipper, W. (2005). *Lernschwierigkeiten erkennen - verständnisvolles Lernen fördern*. IPN-Leibniz.
- Simon, A. L., & Schindler, M. (2020). A comparative analysis of eye tracking and thinking aloud in number line estimation tasks: A study on students with and without mathematical difficulties. In M. Inprasitha, N. Changsri, & N. Boonsena (Hrsg.), *Interim Proceedings of the 44th Conference of the IGPME* (S. 537–545). PME.
- Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, 104, 147–200.
- Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, 49(10), 2674–2688.
- Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12, 151–169.
- Tobii Pro AB. (2020). *Pro Lab User Manual*. <https://www.tobii.com/siteassets/tobii-pro/user-manuals/Tobii-Pro-Lab-User-Manual/?v=1.145>.
- Tsamir, P., Tirosh, D., & Levenson, E. (2008). Intuitive nonexamples: The case of triangles. *Educational Studies in Mathematics*, 69(2), 81–95.
- Unterhauser, E. (2020). *Geometrisches Begriffsverständnis in der frühen Bildung: Eine Interviewstudie zu den Begriffen Vier- und Dreieck bei Kindergartenkindern*. Springer.
- Von Aster, M. (2017). Dyskalkulie. Wenn Kinder nicht rechnen lernen. *Monatsschrift Kinderheilkunde*, 165(6), 482–489.
- Weigand, H.-G., Filler, A., Hölzl, R., Kuntze, S., Ludwig, M., Roth, J., Schmidt-Thieme, B., & Wittmann, G. (2018). *Didaktik der Geometrie für die Sekundarstufe I*. Springer.

Blickbewegungen beim grafischen Ableiten – Lassen sich Fehler durch Eye-Tracking-Daten vorhersagen und elaborieren?

Kathrin Kennel, Sebastian Becker, Pascal Klein, Stefan Küchemann,
Jochen Kuhn und Stefan Ruzika

Inhaltsverzeichnis

8.1	Einleitung und Forschungslage	125
8.2	Theoretischer Hintergrund	126
8.3	Forschungsfragen und Hypothesen	129
8.4	Material und Methode	129
8.5	Ergebnisse	132
8.6	Diskussion	135
	Literatur	141

8.1 Einleitung und Forschungslage

Große Heterogenität von Lernenden erfordert einen adaptiven Unterricht mit Differenzierung (Vock & Gronostaj, 2017). Adaptive Lernsysteme können dazu beitragen, individualisierte Lernwege zu realisieren. Hierzu ist es notwendig, dass diese Systeme eine gute Diagnosefähigkeit besitzen, sodass ein Lernermodell entwickelt werden kann, auf dessen Basis adaptive Reaktionen erfolgen.

K. Kennel · S. Ruzika

Wirtschafts- und Schulmathematik, Fachbereich Mathematik, Technische Universität
Kaiserslautern, Kaiserslautern, Deutschland

E-Mail: kkennel@mathematik.uni-kl.de

S. Ruzika

E-Mail: ruzika@mathematik.uni-kl.de

S. Becker (✉) · S. Küchemann · J. Kuhn

Didaktik der Physik, Fachbereich Physik, Technische Universität Kaiserslautern,
Kaiserslautern, Deutschland

E-Mail: sbeckerg@uni-koeln.de

Zur Erstellung des Lernermodells, welches im Idealfall Aufschluss über den vollständigen Wissensstand bzw. über Lerndefizite gibt, werden im Bereich der Prozessdaten bislang meist nur die getätigten Eingaben mit Maus und Tastatur verwendet. Eye-Tracking-Daten können zusätzlich wertvolle Informationen darüber liefern, welchen Bereichen für wie lange und in welcher Reihenfolge Aufmerksamkeit geschenkt wird. So können Wahrnehmungs- und Verarbeitungsprozesse sichtbar gemacht werden, und es kann ggf. schon im Lösungsprozess eingegriffen werden (Leber & Skuballa, 2014). Studien belegen, dass sich das Blickverhalten von Schüler*innen, die eine Aufgabe richtig lösen, von denen, die eine Aufgabe falsch lösen, unterscheidet (Gegenfurtner et al., 2011). Die Nutzung von Eye-Tracking-Daten in adaptiven Lernsystemen ist daher ein vielversprechender Ansatz zur Erhöhung der Diagnosefähigkeit solcher Systeme. Bisher gibt es aber kaum Ansätze, die diesen Weg verfolgen (Scheiter et al., 2019). Speziell für das Fach Mathematik ist uns kein adaptives Lernsystem bekannt, welches Eye-Tracking-Daten mitberücksichtigt. Es gibt jedoch Vorarbeiten hierzu (z. B. Schindler et al., 2020). Die hier beschriebene Studie stellt den ersten Schritt hin zu einem adaptiven Lernsystem für das Fach Mathematik dar, welches sich thematisch mit dem grafischen Ableiten befasst und Eye-Tracking-Daten zur besseren Diagnose nutzen soll, um ein elaboriertes Feedback bzw. visuelle Hilfen zu ermöglichen. Wenn Blickdaten Aufschluss darüber geben können, für welche Antwort sich eine Probandin bzw. ein Proband entscheidet, so ist beispielsweise denkbar, dass ein adaptives Lernsystem schon im Lösungsprozess gezielt unterstützen kann.

8.2 Theoretischer Hintergrund

Adaptive Lernsysteme

Adaptive Lernsysteme sammeln Daten über einen Lerner, analysieren diese und reagieren angepasst mit dem Ziel, individuelles Lernen bestmöglich zu unterstützen. Die Art der Adaption reicht vom individuellen Feedback über die Auswahl passender Folgeaufgaben bis hin zu einer an den Nutzer angepassten Darstellung der Lerninhalte (Brusilovsky, 1998). Der Kern eines solchen Systems ist die Diagnosefähigkeit. Im Four-Process Adaptive Cycle von Shute und Zapata-

S. Küchemann

E-Mail: s.kuechemann@lmu.de

J. Kuhn

E-Mail: jochen.kuhn@physik.uni-muenchen.de

P. Klein

Physik und ihre Didaktik, Fakultät für Physik, Georg-August-Universität Göttingen, Göttingen, Deutschland

E-Mail: pascal.klein@uni-goettingen.de

Rivera (2012) wird ein solches System modellhaft dargestellt. Es beinhaltet die vier Prozesse Sammeln, Analysieren, Auswählen und Präsentieren. Mit dem Durchlaufen der ersten beiden Prozesse wird ein Lernermodell gebildet, welches ständig in Echtzeit aktualisiert wird. Auf dieser Grundlage kann das System eine für den Lerner passende Rückmeldung wählen und auf dem Bildschirm anzeigen (Shute & Zapata-Rivera, 2012).

Grafisches Ableiten

Die Forschungsfragen werden anhand von Aufgaben zum grafischen Ableiten untersucht. Dieses erfordert in der Regel ein formelfreies Schließen vom Funktionsgraphen auf den entsprechenden Graphen der Ableitungsfunktion. Damit lässt sich der Vorgang der qualitativen Analysis zuordnen, welche nach Hussmann und Prediger (2010) einen bedeutsamen Beitrag zum Aufbau inhaltlicher Vorstellungen im Analysisunterricht leistet. Im Lehrplan wird die Behandlung sowohl für den Grund- als auch für den Leistungskurs explizit gefordert (z. B. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur Rheinland-Pfalz, 2015).

Der Vorgang des grafischen Ableitens verlangt Schüler*innen ein großes Maß an inhaltlichem Verständnis ab (Klinger, 2017) und führt häufig zu Schwierigkeiten (z. B. Aspinwall et al., 1997; Ubuz, 2007; Asiala et al., 1997). Der Wechsel zwischen den Ebenen der Funktion und der Ableitungsfunktion sowie der damit verbundene Aspektwechsel zwischen Kovariations- und Zuordnungsvorstellung stellen eine große Herausforderung für Schüler*innen dar (Hahn & Prediger, 2008). Durch die Vertauschung der Ebenen kann die sog. Slope-Height Confusion (z. B. Leinhardt et al., 1990) eine Verwechslung von Steigung und Höhe auftreten. Ein weiterer typischer Fehler im Kontext des grafischen Ableitens ist die Annahme von Ähnlichkeiten (Nemirovsky & Rubin, 1992). Sind Abszisse und Ordinate des Koordinatensystems unterschiedlich skaliert, kann auch das zu Hürden führen, „da hier nicht mehr notwendigerweise eine größere Steigung einer linearen Funktion mit einem steiler angezeigten Graphen einhergeht“ (Klinger & Thurm, 2016, S. 226). Diese Nichtbeachtung der Achsenbeschriftung wird hier als Skalierungsfehler bezeichnet.

Nach unserem Wissen existiert kein reines Testinstrument zum grafischen Ableiten, allerdings ist es sehr wohl Bestandteil existierender Tests (vgl. Beichner, 1994). Folgende Aspekte sind Teil des Vorgangs und müssten in einem Test abgeprüft werden: Um zu einem gegebenen Funktionsgraphen den Graphen der Ableitungsfunktion skizzieren zu können, müssen Lernende in der Lage sein, den Extremstellen der Funktion die Nullstellen der Ableitungsfunktion zuzuordnen. Weiterhin muss aus dem Wachstumsverhalten der Funktion in einem Intervall auf das Vorzeichen der Funktionswerte der Ableitungsfunktion in diesem Intervall geschlossen werden. Neben dem „ob“ muss auch beurteilt werden, „wie stark“ ein Graph steigt oder fällt. Hierzu ist es erforderlich, dass der Lerner sich die Tangente an den Graphen vorstellen kann. Das Wissen über den Grad einer Funktion kann beim grafischen Ableiten eine große Hilfe sein.

Studien zum Blickverhalten von Experten und Novizen

Die Informations-Reduktions-Hypothese (Haider & Frensch, 1999) besagt, dass geübte Personen in der Lage sind, zwischen aufgabenrelevanten und aufgabenirrelevanten Informationen zu unterscheiden und ihre Verarbeitung auf die aufgabenrelevanten Informationen einzuschränken. Basierend auf dieser Hypothese ist davon auszugehen, dass sich das Blickverhalten von Experten und Novizen beim Aufgabenlösen folgendermaßen unterscheidet:

- Experten haben mehr Fixationen von längerer Dauer auf für die Aufgabe relevante Bereiche.
- Experten haben weniger Fixationen von kürzerer Dauer auf für die Aufgabe irrelevante Bereiche.

In einer Metaanalyse von Gegenfurtner et al. (2011) konnten diese Annahmen verifiziert werden. Madsen et al. (2012) zeigten in ihrer Studie, dass es beim Lösen physikalischer Aufgaben zwischen Teilnehmenden, die richtig oder falsch antworten, zu Unterschieden in der visuellen Aufmerksamkeit kommt. Bei fünf von sechs Problemen konnte auch hier gezeigt werden, dass Probanden und Probandinnen, die eine Aufgabe richtig lösen, länger in thematisch relevante Bereiche schauen (ausgewertetes Eye-Tracking-Maß: TVD = Total Visit Duration). Außerdem fanden sie heraus, dass die Aufmerksamkeit von Novizen oftmals stark durch Fehlkonzepte geleitet wird. Hierzu betrachteten sie auch Bereiche, welche in einer vorgelagerten Interviewstudie von Teilnehmenden mit falscher Antwort besonders in den Fokus genommen wurden und mit Fehlkonzepten aus der Literatur assoziiert waren. Bei fünf von sechs Problemen schauten Probanden und Probandinnen, die eine Aufgabe falsch lösten, länger in diese Bereiche. Weiterhin untermauern die Studien von Jarodzka et al. (2010), Canham und Hegarty (2010) sowie Haider und Frensch (1999), dass Personen mit höherer Expertise relevanten Bereichen mehr Beachtung schenken.

Nutzung von Eye-Tracking-Daten in adaptiven Lernsystemen

Bisher gibt es nur wenige Versuche, Eye-Tracking-Daten in adaptiven Lernsystemen zu integrieren (Scheiter et al., 2019). Conati und Merten untersuchten 2007 erstmals die Eignung von Blickdaten für die Bildung des Lernermodells. Auch der motivationale Zustand von Lernenden lässt sich mithilfe von Eye-Tracking-Daten charakterisieren (D'Mello et al., 2012). Scheiter et al. (2019) entwickelten ein adaptives System, das Eye-Tracking-Daten in Echtzeit auswertet und auf dieser Grundlage personalisierte Unterstützung bereitstellt. Für das Fach Mathematik ist uns kein adaptives Lernsystem bekannt, das Eye-Tracking-Daten auswertet.

8.3 Forschungsfragen und Hypothesen

Die Integration von Eye-Tracking in adaptive Lernsysteme könnte die Diagnosefähigkeit solcher Systeme erhöhen. Eine Voraussetzung hierfür ist, dass sich die Blickbewegungen von erfolgreichen und weniger erfolgreichen Schüler*innen unterscheiden. Damit neben einem antwortbezogenen Feedback auch Feedbackkomponenten angeboten werden können, die die Ursache eines Fehlers adressieren, ist es notwendig, Zusammenhänge zwischen auftretenden Fehlern und Blickmustern zu betrachten.

Die Forschungsfragen lauten:

1. Schauen Schüler*innen, die eine Aufgabe richtig lösen, signifikant länger in relevante Bereiche als Schüler*innen, die die Aufgabe falsch lösen?
2. Gibt es einen Zusammenhang zwischen dem Auftreten von Fehlern und Blickdauern?

Darüber hinaus soll die Pilotstudie Aufschluss über folgende Punkte geben:

- Wie müssen Items aufgebaut sein, damit ein System mithilfe der Auswertung von Blickdaten vorhersagen kann, ob der Nutzer im Begriff ist, eine falsche Lösung abzugeben, und somit eine Hilfestellung benötigt?
- Welches Eye-Tracking-Maß eignet sich, um erfolgreiche von weniger erfolgreichen Schüler*innen zu unterscheiden?

8.4 Material und Methode

Teilnehmende und Datenerhebung

Die Studie wurde in den Mathematik-Oberstufenkursen (Grund- und Leistungskurse der MSS 12 und 13) eines rheinland-pfälzischen Gymnasiums durchgeführt. Das Thema wurde von allen Oberstufenkursen in Klassenstufe 11 behandelt. Es nahmen insgesamt 27 Schüler*innen (11 männlich, 16 weiblich) an der Studie teil, welche alle (teils mithilfe einer Korrektur) über ein normales Sehvermögen verfügten. Die Teilnahme war freiwillig, wurde nicht extrinsisch motiviert und erfolgte entweder in einer Freistunde oder während einer Unterrichtsstunde (mit Erlaubnis der Lehrperson). Die Durchführung nahm eine Unterrichtsstunde (45 min) in Anspruch und erfolgte in der Schulbibliothek, in der bis zu vier Teilnehmende parallel die Aufgaben unter Aufzeichnung ihrer Blickdaten lösten. Zu Beginn des Experiments bekamen die Probanden und Probandinnen eine kurze Einweisung und machten anonymisierte Angaben zum Alter, zum Geschlecht und zur Mathematiknote. Anschließend nahmen die Teilnehmenden jeweils vor einem Computerbildschirm Platz, und es wurde eine 9-Punkt-Kalibrierung durchgeführt.

Die Probanden und Probandinnen bearbeiteten 29 Items zum Thema „Grafisches Ableiten“ in vorgegebener Reihenfolge und ohne Unterbrechung. Sie bekamen kein Feedback, und es war nicht möglich zu einer vorherigen Aufgabe zurückzuspringen. Nach jeder Lösungsabgabe (durch Klick oder mittels Zettel und Stift) wurde die Antwortsicherheit abgefragt.

Material

Das Testinstrument zum grafischen Ableiten wurde selbst erstellt und enthält 26 Items zum Thema und drei einführende Items, welche den grundsätzlichen Umgang mit Graphen abprüfen (Skalierung der y-Achse, Ablesen einer Stelle bzw. eines Intervalls). Die Aufgaben lassen sich in folgende Kategorien einteilen:

- 0 Umgang mit Graphen (3 Items)
- 1 Steigungen bestimmen und vergleichen (10 Items)
- 2 Zusammenhänge erkennen (7 Items)
- 3 Graph der Ableitungsfunktion ermitteln (9 Items)

Die Items der Kategorie 1 reichen vom qualitativen Vergleich von Steigungen (größer/kleiner) hin zur quantitativen Ermittlung von Steigungen linearer und nichtlinearer Funktionen. Letzteres erfordert ein adäquates Verständnis einer Tangente. Aufgaben der Kategorie 2 sollen prüfen, ob Zusammenhänge zwischen dem Funktionsgraphen und dem Graphen der Ableitungsfunktion verstanden werden. Beispielsweise muss die Ableitungsfunktion in Intervallen, in denen die Funktion streng monoton fällt, kleiner null sein. Kategorie 3 beinhaltet die eigentliche Aufgabe des grafischen Ableitens. Für eine Funktion muss jeweils der passende Graph der Ableitungsfunktion gefunden werden. Bei den neun Items wurden Funktionen unterschiedlichen Grades gewählt. Die Kriterien, die hierbei zur Lösungsfindung herangezogen werden müssen, unterscheiden sich.

Die Graphen aller Items wurden selbst erstellt. Folgende Items wurden aus der Literatur entnommen und neu dargestellt: Item 0.1 (Kerslake, 1981), Item 1.4 (Planinic et al., 2013), Item 1.5 und 1.6 (Ceuppens et al., 2019) und Item 3.3 (Ubuz, 2007).

Bei 17 der Items handelt es sich um Single-Choice-Aufgaben, wobei die oben beschriebenen häufig auftretenden Fehler in den Distraktoren aufgegriffen werden (Abb. 8.1 rechts). Aufgrund der Vermutung, dass eine Vorgabe von Antwortmöglichkeiten das Blickverhalten beeinflusst, sind die restlichen zwölf Aufgaben offen gestellt. Außerdem kann die Wahl des Antwortformats die Anzahl korrekter Antworten beeinflussen (Berg & Smith, 1994). Abb. 8.1 (links) zeigt ein offen gestelltes Item, welches so konstruiert wurde, dass bestimmte Fehler (analog zu den Distraktoren) im Graphen wiederzufinden sind. So ist beispielsweise davon auszugehen, dass die Teilnehmenden, die vermehrt in den hellgrau eingefärbten Bereich schauen, die Skalierung der y-Achse nicht beachten und sich zur Bestimmung einer Stelle mit der angegebenen Steigung nur an den

Kästchen orientieren. Schüler*innen, die vorrangig der schwarz eingefärbten AOI Beachtung schenken, verwechseln wahrscheinlich die Steigung mit dem Funktionswert.

Eye-Tracking-Apparat

Die Items wurden den Probanden und Probandinnen auf einem 22-Zoll-Monitor präsentiert. Die Auflösung des Computerbildschirms betrug 1920×1080 Pixel mit einer Bildwiederholfrequenz von 75 Hz. Die Blickbewegungen wurden mit einem stationären Remote-Eye-Tracking-System (Tobii X3-120, binokular) mit einer Frequenz von 120 Hz und einer Accuracy von durchschnittlich $0.2\text{--}0.9^\circ$ (Tobii AB, 2015) erfasst. Vor jeder Datenerhebung wurde eine 9-Punkt-Kalibrierung durchgeführt. Der mittlere Abstand zwischen Bildschirm und Teilnehmendem betrug 65 cm. Zur Identifikation von Fixationen und Sakkaden wurde ein I-VT-Algorithmus mit den Schwellenwerten $8500^\circ/\text{s}^2$ für die Beschleunigung und $30^\circ/\text{s}$ für die Geschwindigkeit angewendet (Salvucci & Goldberg, 2000).

Datenanalyse/Methode der Auswertung

Die Datenerhebung und die Festlegung von AOIs erfolgte mit der Software Tobii Pro Studio. Aufgrund geringer Trackinggenauigkeit wurde ein Proband nachträglich von der Analyse ausgeschlossen, da die Fixationen beim Lesen der Aufgabenstellung deutlich vom Text abwichen. Somit liegt eine Stichprobengröße von $n=26$ vor. Zur Festlegung der relevanten Bereiche jeder Aufgabe diente ein Expertenrating. Hierzu markierten acht Mathematiker diejenigen Bereiche, die sie für die Lösung der Aufgabe als wichtig einstufen. Die am häufigsten markierten Bereiche wurden zu einer AOI-Gruppe zusammengefasst und verwendet.

Die Datenanalyse erfolgte mit der Programmiersprache R. Die codierten Antworten (0 für falsche Lösung, 1 für richtige Lösung) wurden in der dichotomen Variablen „Korrektheit“ gespeichert, welche die Teilnehmenden für jedes Item in zwei Gruppen einteilt. Antworten, bei denen die Teilnehmenden angaben, geraten zu haben, wurden hierbei als falsch markiert. Item 1.2 wurde ausgeschlossen, da dieses Item nur von einer Person nicht gelöst wurde (Tab. 5.1).

Folgende Eye-Tracking-Maße wurden betrachtet: Fixation Count (FC), Total Fixation Duration (TFD), Total Visit Duration (TVD), Time To First Fixation (TTFF) und Mean Fixation Duration (FD). Zusätzlich wurde für jedes Item eine AOI um den gesamten Graphen (ohne Aufgabenstellung und Antwortmöglichkeiten) gelegt. Dies ermöglichte die Betrachtung prozentualer Maße (vgl. Madsen et al., 2012).

Da die Voraussetzungen der Normalverteilung für den T-Test bei keinem der Eye-Tracking-Maße erfüllt sind, wurde der Mann-Whitney-U-Test für unabhängige Stichproben durchgeführt. Zur Identifikation geeigneter Maße für die Unterscheidung der Gruppen wurden zusätzlich Verfahren der Variablenselektion

Tab. 8.1 Itemanalyse: Verteilung der Rohwerte nach Dichotomisierung (richtig/falsch). r_{ji} = Trennschärfekoeffizient, P_j = Itemschwierigkeit, D_j = Diskriminationsindex

Item	richtig beantw.	falsch beantw.	r_{ji}	P_j	D_j	Item	richtig beantw.	falsch beantw.	r_{ji}	P_j	D_j
0.1	18	8	0,48	0,7	0,67	2.3	14	12	0,45	0,56	0,67
0.2	20	6	0,36	0,74	0,33	2.4	6	20	0,57	0,26	0,67
0.3	21	5	0,27	0,81	0,33	2.5	11	15	0,51	0,44	0,67
1.1	22	4	0,4	0,85	0,33	2.6	8	18	0,36	0,33	0,33
1.2	25	1	0,35	0,96	0,11	2.7	5	21	0,38	0,22	0,33
1.3	11	15	0,18	0,44	0,22	3.1	18	8	0,65	0,7	0,78
1.4	15	11	0,54	0,59	0,67	3.2	18	8	0,25	0,7	0,33
1.5	21	5	0,65	0,81	0,56	3.3	21	5	0,54	0,81	0,44
1.6	19	7	0,77	0,74	0,78	3.4	23	3	0,51	0,89	0,33
1.7	3	23	0,4	0,11	0,33	3.5	19	7	0,24	0,74	0,22
1.8	6	20	0,43	0,22	0,56	3.6	15	11	0,49	0,59	0,67
1.9	11	15	0,45	0,44	0,56	3.7	14	12	0,7	0,56	0,89
1.10	8	18	0,49	0,33	0,56	3.8	14	12	0,72	0,56	0,89
2.1	19	7	0,29	0,74	0,33	3.9	17	9	0,16	0,67	0,22
2.2	14	12	0,41	0,56	0,67						

angewendet (Pruscha, 2006). Um auch Kombinationen von Eye-Tracking-Maßen untersuchen zu können, kam eine nichtparametrische Alternative zum Hotelling’s T^2 -Test zum Einsatz, bei welchem zwei Teststatistiktypen (Wilks’ Lambda und ANOVA-Permutationstest) verwendet werden (Ellis et al., 2017).

Wird eine Vielzahl von Eye-Tracking-Maßen in die Analyse mit einbezogen, können mit der Diskriminanzanalyse Gewichte ermittelt werden, sodass eine best-mögliche Trennung der beiden Gruppen erreicht wird (Bortz & Schuster, 2010). Mit Blick auf adaptive Lernsysteme, die für einen „neuen“ Lerner entscheiden sollen, welcher Gruppe dieser angehört, wird im Anschluss ein Klassifikationsverfahren genutzt.

Für die zweite Forschungsfrage wurden zusätzlich AOIs für häufig auftretende Fehler festgelegt (Abb. 8.1). Die Schüler*innen wurden für jedes Item anhand ihres Antwortverhaltens in Gruppen (richtig, Fehler 1, Fehler 2, ..., Rest) eingeteilt. Je nach Anzahl der Gruppen bzw. Anzahl der AOI-Kategorien wurde entweder der Kruskal-Wallis-Test (nichtparametrische Alternative zur einfaktoriellen Varianzanalyse) oder der Mann-Whitney-U-Test durchgeführt. Mit Blick auf die Ergebnisse zur ersten Forschungsfrage wurde hierbei analog zu Madsen et al. (2012) nur die Percentaged Total Visit Duration (TVD%) ausgewertet.

8.5 Ergebnisse

Ergebnisse zu Forschungsfrage 1

Zur Testbewertung wurde zunächst eine Itemanalyse durchgeführt (z. B. Döring & Bortz, 2016). Die Ergebnisse sind in Tab. 8.1 dargestellt. Zwei der Items weisen Schwierigkeitsindizes außerhalb des Normbereichs auf ($P_j = 0,96$ bzw.

$P_j = 0.11$). Dies zeigt sich auch in der Verteilung der Testwerte. Wie in Abschn. 8.4 bereits erwähnt, wurde Item 1.2 aufgrund des geringen Schwierigkeitsgrades ausgeschlossen. Für ein Großteil der Items ergibt sich eine gute Trennschärfe mit $r_{jt} > 0.4$ (Moosbrugger & Kelava, 2012). Die Items 1.3 und 3.9 präsentieren das Gesamtergebnis am wenigsten gut und eignen sich nicht zur Unterscheidung zwischen Teilnehmenden mit hohem bzw. niedrigem Testwert. Neben dem ausgeschlossenen Item weisen diese und ein weiteres Item (3.5) auch einen geringen Diskriminationsindex auf, sind also weniger geeignet, zwischen leistungsstärkeren und -schwächeren Personen zu unterscheiden (Rost, 2004).

Zur Unterscheidung der Gruppen (korrekt-inkorrekt) wurden die zu den relevanten Bereichen zugehörigen Eye-Tracking-Daten mittels des Mann-Whitney-U-Tests ausgewertet. Die Ergebnisse des Tests (p -Werte korrigiert nach Benjamini-Hochberg), der für alle acht Variablen separat durchgeführt wurde, sind in Tab. 8.2 dargestellt. Es fällt auf, dass die prozentualen Maße am häufigsten zu signifikanten Unterschieden führen. Betrachtet man beispielsweise die Anzahl der Fixationen in den relevanten Bereichen im Verhältnis zu allen Fixationen im Graphbereich, so findet man für 12 der 28 Items signifikante Unterschiede für die beiden Gruppen (Signifikanzniveau: 5 %). Die Variablen TTFF und FD werden für keines der Items signifikant und daher nicht weiter betrachtet. Aus den restlichen sechs Variablen wurde mittels Verfahren der Variablenselektion für jedes Item diejenige Teilmenge identifiziert, die sich am besten zur Trennung der Gruppen eignet.

Die beiden Verfahren Stepwise Regression und Best Subset entstammen dem Feld der angewandten Regressionsanalyse und liefern für jedes Item eine Teilmenge der Variablen. In beiden Ansätzen werden TVD% und FC% am häufigsten ausgewählt. Dieses Ergebnis deckt sich mit den Ergebnissen des U-Tests (Tab. 8.2). Der Test liefert für diese beiden Eye-Tracking-Maße signifikante Ergebnisse bei elf bzw. zwölf Items. Es fällt auf, dass dies nicht immer dieselben Items sind. So wird der Test für die Variable TVD% bei Item 1.1 signi-

Tab. 8.2 Ergebnisse des Mann-Whitney-U-Tests mit Benjamini-Hochberg-Korrektur (p -Werte) für die Variablen Time To First Fixation (TTFF), Total Visit Duration (TVD), Total Fixation Duration (TFD),

Fixation Count (FC), Fixation Duration (FD), Percentaged Total Visit Duration (TVD%), Percentaged Total Fixation Duration (TFD%) und Percentaged Fixation Count (FC%); auf drei Nachkommastellen gerundet

Item	TTFF	TVD	TFD	FC	FD	TVD%	TFD%	FC%	Item	TTFF	TVD	TFD	FC	FD	TVD%	TFD%	FC%
0.1	1,000	0,670	0,720	0,575	0,876	0,615	0,439	0,355	2.3	1,000	0,511	0,639	0,529	0,876	0,018	0,008	0,005
0.2	0,727	0,670	0,894	0,575	0,541	0,036	0,121	0,036	2.4	1,000	0,309	0,477	0,159	0,876	0,128	0,103	0,071
0.3	1,000	0,500	0,528	0,602	0,410	0,018	0,018	0,030	2.5	0,814	0,006	0,019	0,018	0,876	0,003	0,004	0,004
1.1	1,000	0,500	0,265	0,575	0,541	0,027	0,053	0,088	2.6	1,000	1,000	0,934	0,956	0,876	0,615	0,723	0,681
1.3	1,000	0,670	0,645	0,575	0,876	0,616	0,370	0,564	2.7	1,000	0,385	0,477	0,575	0,876	0,111	0,110	0,088
1.4	0,771	0,085	0,098	0,111	0,543	0,067	0,087	0,108	3.1	1,000	0,500	0,528	0,575	0,911	0,444	0,297	0,371
1.5	1,000	0,918	0,706	0,956	0,410	0,067	0,087	0,355	3.2	1,000	0,500	0,629	0,529	0,876	0,085	0,110	0,037
1.6	0,727	0,670	0,645	0,575	0,876	0,025	0,037	0,030	3.3	1,000	0,345	0,477	0,159	0,876	0,615	0,554	0,578
1.7	0,771	0,039	0,039	0,054	0,426	0,022	0,022	0,030	3.4	1,000	0,039	0,039	0,038	0,138	0,032	0,037	0,035
1.8	1,000	0,006	0,006	0,007	0,138	0,003	0,004	0,004	3.5	0,727	1,000	0,894	0,880	0,876	0,348	0,439	0,379
1.9	1,000	0,621	0,645	0,575	0,876	0,979	0,938	0,775	3.6	1,000	0,352	0,528	0,162	0,876	0,645	0,443	0,355
1.10	1,000	0,250	0,235	0,168	0,541	0,231	0,110	0,096	3.7	1,000	0,500	0,629	0,575	0,876	0,070	0,069	0,035
2.1	0,813	0,006	0,006	0,007	0,568	0,003	0,004	0,005	3.8	1,000	0,385	0,477	0,159	0,876	0,616	0,446	0,172
2.2	1,000	0,794	0,684	0,646	0,876	0,008	0,004	0,004	3.9	1,000	1,000	0,894	0,880	0,876	0,616	0,697	0,578

Tab. 8.3 Ergebnisse des nichtparametrischen Tests (R-Funktion nrmv) für die beiden Variablen TVD% und FC% mit Benjamini-Hochberg-Korrektur (*p*-Werte aufgrund des Teststatistiktyps Wilks' Lambda (grau) bzw. aufgrund des ANOVA-Permutationstests (schwarz))

Item	<i>p</i> -Wert nrmv-Test	Item	<i>p</i> -Wert nrmv-Test	Item	<i>p</i> -Wert nrmv-Test	Item	<i>p</i> -Wert nrmv-Test
0.1	0,625	1.6	0.014*	2.3	<0.001*	3.3	0,836
0.2	0.017*	1.7	0.016*	2.4	0.047*	3.4	0.016*
0.3	0.016*	1.8	<0.001*	2.5	<0.001*	3.5	0.344
1.1	0.039*	1.9	0,934	2.6	0.640	3.6	0,522
1.3	0,836	1.10	0,087	2.7	0,074	3.7	0.039*
1.4	0,079	2.1	<0.001*	3.1	0,344	3.8	0,344
1.5	0.039*	2.2	<0.001*	3.2	0.044*	3.9	0.640

fikant, währenddessen er für die Variable FC% stattdessen bei den Items 3.2 und 3.7 signifikante Ergebnisse liefert. Um zu überprüfen, ob eine Kombination dieser beiden Eye-Tracking-Maße ein besseres Ergebnis erzielt, wird eine nichtparametrische Alternative zum Hotelling's T²-Test angewendet. Tab. 8.3 zeigt das Ergebnis. Werden beide Variablen berücksichtigt, kann bei 15 Items ein signifikanter Unterschied nachgewiesen werden.

Bei der Diskriminanzanalyse wird im Fall von zwei untersuchten Gruppen eine neue Achse gesucht, auf der sich die Projektionen der Punkte der beiden Gruppen möglichst deutlich voneinander unterscheiden. Diese Achse bezeichnet man als Diskriminanzfunktion oder (in Analogie zur Faktorenanalyse) als Diskriminanzfaktor (Bortz & Schuster, 2010). Zur Bestimmung dieser werden für jedes Item Diskriminanzkoeffizienten (Gewichte für die einbezogenen Variablen) ermittelt. Hieraus kann dann für jeden Teilnehmenden ein Faktorwert berechnet werden. In die Berechnung wurden alle oben beschriebenen acht Variablen mit einbezogen. Tab. 8.4 zeigt die korrigierten Ergebnisse des Mann-Whitney-U-Tests für diese Faktorwerte. Es wird deutlich, dass es für fast jedes Item (außer Item 0.1 und 1.9) möglich ist, mithilfe der acht Eye-Tracking-Maße eine signifikante Unterscheidung der beiden Gruppen zu erzielen. Um zu überprüfen, wie gut sich Personen anhand des ermittelten Diskriminanzfaktors den ursprünglichen Gruppen zuordnen lassen, können Klassifikationsverfahren angewendet werden. Bei der Leave-one-out-Methode (Spezialfall der Kreuzvalidierung) besteht die Trainingsgruppe hierbei aus N-1 Personen, wobei die jeweils ausgelassene Person auf Basis

Tab. 8.4 *p*-Werte des Mann-Whitney-U-Tests (mit Benjamini-Hochberg-Korrektur) für die Faktorwerte und Trefferquote der Leave-one-out-Methode in %

Item	<i>p</i> -Wert	Trefferquote	Item	<i>p</i> -Wert	Trefferquote	Item	<i>p</i> -Wert	Trefferquote	Item	<i>p</i> -Wert	Trefferquote
0.1	0,107	50,0%	1.6	<0.001	69,2%	2.3	<0.001	61,5%	3.3	0,010	69,2%
0.2	<0.001	69,2%	1.7	0,006	88,5%	2.4	0,001	84,6%	3.4	0,006	92,3%
0.3	0,001	80,8%	1.8	<0.001	96,2%	2.5	<0.001	80,8%	3.5	0,004	57,7%
1.1	0,002	76,9%	1.9	0,154	30,8%	2.6	0,040	57,7%	3.6	0,010	38,5%
1.3	0,001	57,7%	1.10	0,002	61,5%	2.7	0,007	69,2%	3.7	0,012	42,3%
1.4	0,001	53,8%	2.1	<0.001	76,9%	3.1	0,040	42,3%	3.8	<0.001	73,1%
1.5	0,001	80,8%	2.2	<0.001	57,7%	3.2	0,002	57,7%	3.9	0,003	46,2%

der anderen klassifiziert werden soll. Dieser Vorgang wird N-mal durchgeführt. Tab. 8.4 zeigt für jedes Item die Trefferquote in Prozent.

Für Item 1.8 ist die Trefferquote beispielsweise sehr hoch. 96.2 % aller Beobachtungen können hier aufgrund ihrer Diskriminanzwerte korrekt klassifiziert werden. Bei 9 der 28 Items liegt die Trefferquote über 75 %, für fünf Items werden Teilnehmende allerdings in weniger als der Hälfte der Fälle korrekt klassifiziert.

Ergebnisse zu Forschungsfrage 2

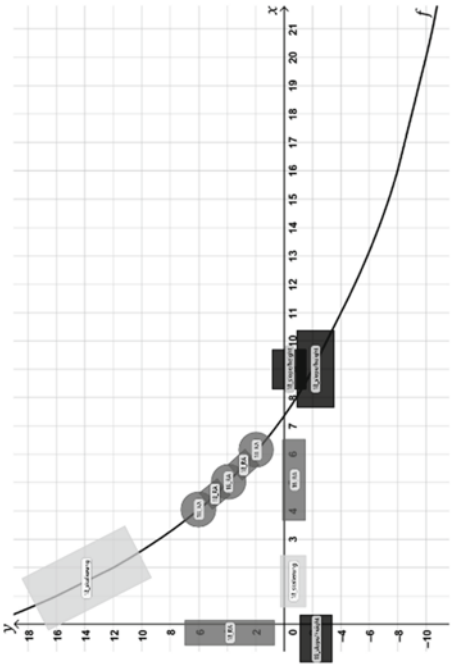
Für die zweite Forschungsfrage ist die Auswertung für jedes Item sehr individuell, da verschiedene AOI-Bereiche für die jeweils (in unterschiedlicher Anzahl) vorkommenden typischen Fehler festgelegt werden. Aus diesem Grund werden die Ergebnisse exemplarisch an den beiden Items 1.8 und 2.5 dargestellt. Wie in Abb. 8.1 ersichtlich, wurden für Item 1.8 drei AOI-Bereiche (relevante Bereiche sowie Bereiche für den Skalierungs- und den Slope-Height-Fehler) festgelegt. Ein Kruskal-Wallis-Test bestätigt, dass es in allen drei Bereichen signifikante Unterschiede zwischen den Gruppen hinsichtlich der Percentaged Total Visit Duration (TVD%) gibt (Tab. 8.5). Anschließend durchgeführte Post-hoc-Tests (Dunn-Bonferroni) zeigen, dass sich für die relevanten Bereiche (AOI RA) die Gruppe mit richtiger Lösung von den beiden anderen Gruppen signifikant unterscheiden ($z = 3.197$, $p = .004$, Effektstärke nach Cohen (1992): $r = .826$ bzw. $z = 3.135$, $p = .003$, $r = .870$). Analog ergibt sich für das zum Skalierungsfehler zugehörige Gebiet ein signifikanter Unterschied zwischen der Gruppe, die die Antwort zum Skalierungsfehler ausgewählt hat, und den jeweils anderen Gruppen ($z = 3.836$, $p < .001$, $r = .959$ bzw. $z = 2.754$, $p = .009$, $r = .711$). Gleiches gilt für den zweiten Fehlerbereich. Der Boxplot in Abb. 8.2 veranschaulicht dieses Ergebnis.

Für Item 2.5 wurde nur ein zusätzlicher Bereich für den Fehler „Slope-Height-Confusion“ festgelegt (Abb. 8.1). Für die beiden Gruppen („richtig“ bzw. „falsch_slope“) mit $n_1 = n_2 = 11$ ergeben sich mit dem exakten Mann-Whitney-U-Test für beide Gebiete (AOI RA und AOI Slope-Height) signifikante Unterschiede ($W = 9$, $p < .001$, $r = .737$ bzw. $W = 113$, $p < .001$, $r = .698$). Die Effektstärke entspricht mit $r > .50$ einem starken Effekt. Die Boxplots in Abb. 8.3 veranschaulichen das Ergebnis.

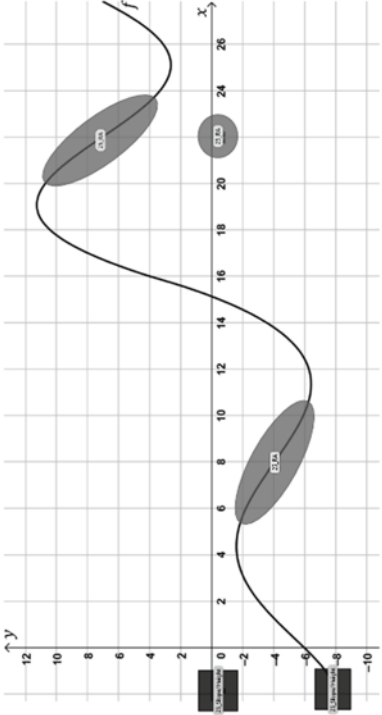
8.6 Diskussion

Die Ergebnisse zeigen, dass Blickdaten eine Vorhersage darüber ermöglichen, für welche Antwort sich eine Probandin bzw. ein Proband entscheiden wird. Somit ist denkbar, dass ein adaptives System Schüler*innen, die den irrelevanten Bereichen besonders viel Aufmerksamkeit schenken, schon im Lösungsprozess durch visuelle Hilfen unterstützen kann. Abb. 8.4 (links) zeigt den gaze plot einer Schülerin für Item 1.9 mit der Aufgabenstellung „Bestimmen Sie die Steigung der

Die Abbildung zeigt den Funktionsgraphen einer Funktion $f(x)$.
Geben Sie eine Stelle an, an der die Steigung -2 beträgt.



Die Abbildung zeigt den Funktionsgraphen einer Funktion $f(x)$.
Für welchen Wert von x nimmt die Ableitung ihren kleinsten Wert an?



- a) $x = -2$ b) $x = 8$ c) $x = 16$ d) $x = 22$

Abb. 8.1 Item 1.8 (offen gestellt) und Item 2.5 (Single Choice) mit eingezeichneten AOIs für relevante Bereiche (grau), für die Slope-Height Confusion (schwarz) und für den Skalierungsfehler (hellgrau)

Tab. 8.5 Ergebnisse des Kruskal-Wallis-Tests (χ^2 und p -Wert, Variable: TVD%) für die drei AOI-Bereiche AOI RA, AOI Skalierung und AOI Slope-Height sowie paarweiser Gruppenvergleich mit z - und p -Wert (Dunn-Bonferroni-Test) für Item 1.8

	χ^2	p -Wert	Skalierung-Slope		Skalierung-Richtig		Slope-Richtig	
			z	p	z	p	z	p
AOI RA	12,788	0,002	0,118	0,906	-3,197	0,004	-3,135	0,003
AOI Skal	16,317	<0,001	3,836	<0,001	2,754	0,009	-0,865	0,387
AOI Slope	13,999	0,001	-3,464	0,002	-0,164	0,870	2,982	0,004

Tangente an den Graphen an der Stelle $x=2$ näherungsweise“. Es wird ersichtlich, dass die Schülerin beim Lösen der Aufgabe richtig vorgeht und die Tangente scheinbar mit ihrem Blick anlegt. Trotzdem wird sie eine falsche Lösung abgeben, da sie die Skalierung der y -Achse nicht betrachtet. Rechts ist die Heatmap aller Teilnehmenden für das Item 3.4 dargestellt, welches von 88 % der Teilnehmenden richtig gelöst wurde. Hier lässt sich erkennen, dass die Schüler*innen auf der Suche nach dem passenden Graphen der Ableitung den Extremstellen besonders viel Aufmerksamkeit schenken.

Bezugnehmend zur ersten Forschungsfrage konnte bei vielen Items gezeigt werden, dass Schüler*innen, die eine Aufgabe richtig lösen, signifikant länger in relevante Bereiche schauen, als Schüler*innen, die eine Aufgabe falsch lösen. Insgesamt wurden bei 15 der 28 Items signifikante Unterschiede gefunden. Diese „gut funktionierenden“ Items zeichnen sich dadurch aus, dass relevante Bereiche des Graphen sich räumlich von Bereichen abgrenzen, die typische Fehler abbilden. Weiterhin enthalten diese Items im Aufgabentext keine direkten Verweise auf Graphbereiche und sind in ausreichender Größe dargestellt. Einige der genutzten Items erfüllen diese Kriterien nicht und müssten dementsprechend verändert werden, um sinnvoll in ein adaptives Lernsystem mit Eye-Tracking integriert werden zu können. So wird die in der Aufgabenstellung von Item 1.9 benannte Stelle $x=2$ im Graphen gleichermaßen von Probanden und Probandinnen betrachtet, die imstande sind, die Aufgabe zu lösen, als auch von allen anderen. Eine Unterscheidung der beiden Gruppen wird hierdurch erschwert. Küchemann et al. (2020) umgehen dieses Problem, indem sie alle Bereiche isolieren, die in direktem Zusammenhang mit den im Text genannten Stellen stehen. Weiter liegen bei diesem Item relevante Bereiche und typische Fehlerbereiche zu dicht aneinander. Die Wahl eines stärker gekrümmten Graphen würde dieses Problem abschwächen. Items der Kategorie 3 zeigen meist vier Graphen, aus denen der passende Graph der Ableitungsfunktion gewählt werden soll. Hieraus ergibt sich das Problem, dass die Graphen recht klein und relevante von irrelevanten Stellen schwieriger zu trennen sind.

Bezugnehmend zur zweiten Fragestellung zeigte sich für ausgewählte Items (z. B. 1.8 und 2.5), dass Schüler*innen, die einen bestimmten Fehler machen, länger in dazu passende Bereiche schauen. Diese passenden Bereiche (AOIs) sollten durch eine Vorstudie z. B. mit lautem Denken festgelegt werden. Für viele Items funktionierte die Analyse hier nicht wie bei den oben beschriebenen Items. Ein Grund hierfür stellt die kleine Stichprobe dar, welche wiederum zu sehr

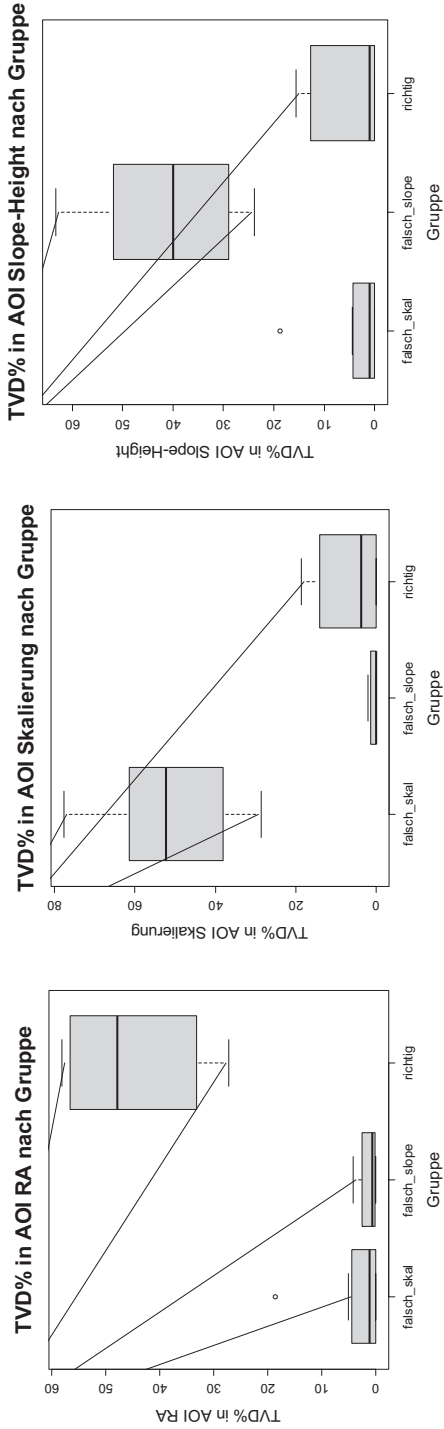


Abb. 8.2 Boxplots zu Item 1.8

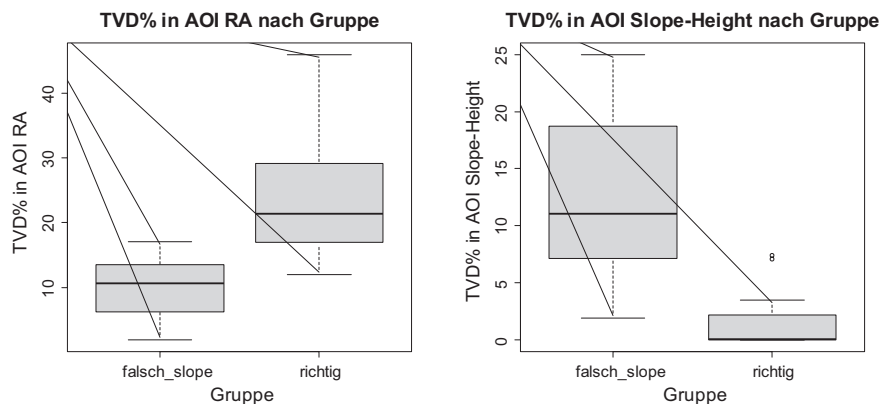


Abb. 8.3 Boxplots zu Item 2.5

kleinen Teilgruppen für die bestimmten Fehler führte. Weiter ist hier besonders wichtig, dass die Bereiche für die häufig auftretenden Fehler überschneidungsfrei sind. Die Konstruktion der Items ist maßgeblich für eine erfolgreiche Unterscheidung verschiedener Gruppen und somit für die Vorhersagbarkeit.

In der vorliegenden Studie führen die beiden Eye-Tracking-Maße Percentaged Total Visit Duration (TVD%), und Percentaged Fixation Count (FC%) am häufigsten zu einer signifikanten Unterscheidung zwischen den beiden Gruppen. Für ein adaptives System, welches Eye-Tracking-Daten in Echtzeit auswerten soll, sind diese Maße jedoch nicht gut geeignet. Da sie erst am Ende der Bearbeitung des jeweiligen Items berechnet werden können, eignen sie sich für eine Fehlerdiagnose im Anschluss an die Bearbeitung, nicht aber für (z. B. visuelle) Hilfestellungen im Lösungsprozess. Scheiter et al. (2019) nutzen beispielsweise die in Echtzeit auswertbaren Eye-Tracking-Maße FC und TFD für ihr adaptives System. Ein adaptives Lernsystem, welches Percentaged-Maße als Entscheidungsgrundlage für Hilfestellungen nutzt, müsste während der Bearbeitung zu bestimmten Zeitpunkten Berechnungen durchführen, die sich auf den jeweils bisherigen Bearbeitungszeitraum beziehen.

Das Design von Interventionen auf Basis von Eye-Tracking-Daten steckt noch „in den Kinderschuhen“ (Scheiter & Eitel, 2017). Es ist eine offene Frage, welche Eye-Tracking-Parameter gewählt werden sollten, um über die Adaption zu entscheiden. Auch stellt sich die Frage, zu welchem Zeitpunkt und ab welchen Grenzwerten eine Intervention erfolgen soll. Dies muss individuell entschieden werden. Für die Auswertung von Items der Kategorie 3 könnte die zusätzliche Auswertung von Transitionen (zwischen dem Graphen und dem Graphen der Ableitungsfunktion) Aufschluss über die Kompetenz des Nutzers geben und eine bessere Vorhersagbarkeit gewährleisten.

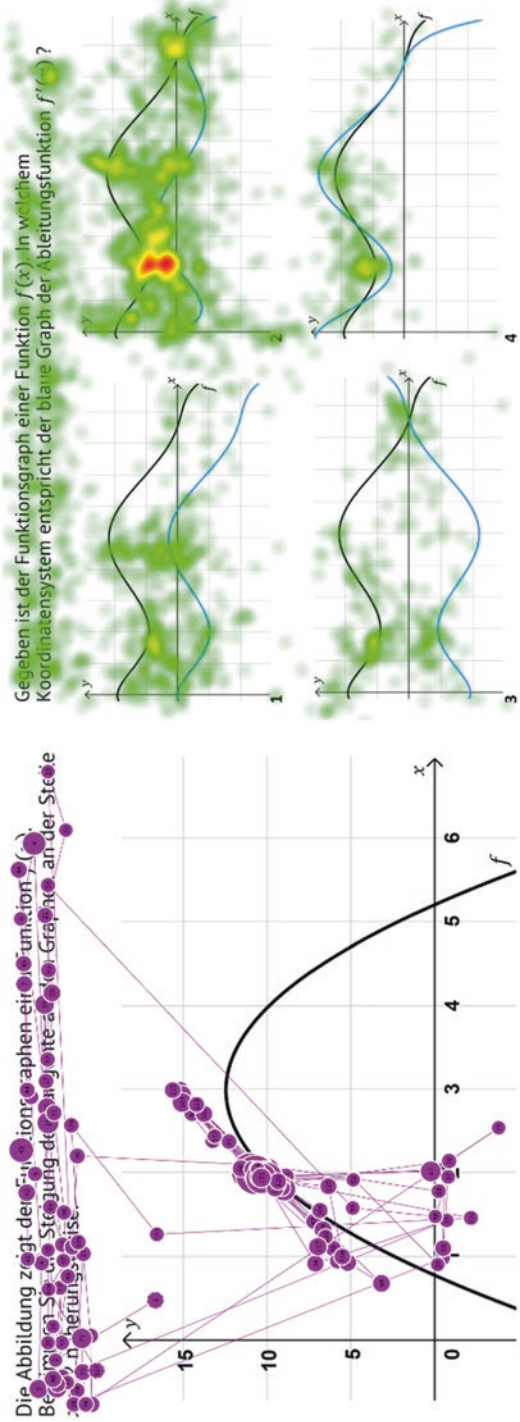


Abb. 8.4 Gaze plot einer Schülerin des Items 1.9 (links) und Heatmap aller Schüler*innen des Items 3.4 (rechts)

Um eine bessere Auswertung der Items zu ermöglichen und die Klassifikation eines Nutzers zu einer Gruppe zu verbessern, sollte die Stichprobe vergrößert werden. Außerdem sollten die Items teilweise überarbeitet werden, sodass o. g. Kriterien erfüllt sind.

Danksagung/Angaben zur finanziellen Förderung Das Projekt MAL-i als Teilprojekt des Vorhabens „U.EDU: Unified Education – Medienbildung entlang der Lehrerbildungskette“ wird im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern aus Mitteln des Bundesministeriums für Bildung und Forschung gefördert.

Literatur

- Asiala, M., Cottrill, J., Dubinsky, E., & Schwingendorf, K. E. (1997). The development of students' graphical understanding of the derivative. *The Journal of Mathematical Behavior*, 16(4), 399–431.
- Aspinwall, L., Shaw, K. L., & Presmeg, N. C. (1997). Uncontrollable mental imagery: Graphical connections between a function and its derivative. *Educational studies in mathematics*, 33(3), 301–317.
- Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, 62(8), 750–762.
- Berg, C. A., & Smith, P. (1994). Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. *Science Education*, 78(6), 527–554.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Springer.
- Brusilovsky, P. (1998, August). Adaptive educational systems on the world-wide-web: A review of available technologies. In *Proceedings of Workshop "WWW-Based Tutoring" at 4th International Conference on Intelligent Tutoring Systems (ITS'98), San Antonio, TX*.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and instruction*, 20(2), 155–166.
- Ceuppens, S., Bollen, L., Deprez, J., Dehaene, W., & De Cock, M. (2019). 9th grade students' understanding and strategies when solving $x(t)$ problems in 1D kinematics and $y(x)$ problems in mathematics. *Physical Review Physics Education Research*, 15(1), 010101.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155.
- Conati, C., & Merten, C. (2007). Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. *Knowledge-Based Systems*, 20(6), 557–574.
- D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies*, 70(5), 377–398.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation*. Springer.
- Ellis, A. R., Burchett, W. W., Harrar, S. W., & Bathke, A. C. (2017). Nonparametric inference for multivariate data: The R package npmv. *Journal of Statistical Software*, 76(4), 1–18.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552.
- Hahn, S., & Prediger, S. (2008). Bestand und Änderung – Ein Beitrag zur Didaktischen Rekonstruktion der Analysis. *Journal für Mathematik-Didaktik*, 29(3–4), 163–198.
- Hussmann, S., & Prediger, S. (2010). Vorstellungsortorientierte Analysis – auch in Klassenarbeiten und zentralen Prüfungen. *Praxis der Mathematik in der Schule*, 52(31), 35–38.

- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 172.
- Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154.
- Kerslake, D. (1981). Graphs. In K. M. Hart (Hrsg.), *Children's understanding of mathematics concepts. 11–16* (S. 120–136). London: John Murray.
- Klinger, M. (2017). *Funktionales Denken beim Übergang von der Funktionenlehre zur Analysis: Entwicklung eines Testinstruments und empirische Befunde aus der gymnasialen Oberstufe*. Springer.
- Klinger, M., & Thurm, D. (2016). Zwei Graphen aber eine Funktion? – Konzeptuelles Verständnis von Koordinatensystemen mit digitalen Werkzeugen entwickeln. *transfer Forschung – Schule*, 2(2), 225–232.
- Küchemann, S., Klein, P., Becker, S., Kumari, N., & Kuhn, J. (2020). Classification of students' conceptual understanding in STEM education using their visual attention distributions: A comparison of three machine-learning approaches. *CSEDU*, 1, 36–46.
- Leber, J., & Skuballa, I. T. (2014). Lernräume adaptiv gestalten. Ein blickbewegungsbasierter Ansatz. In: Rummler, Klaus (Hrsg.), *Lernräume gestalten – Bildungskontexte vielfältig denken* (S. 579–591). Münster u.a.: Waxmann. URN: [urn:nbn:de:0111-pedocs-101173](https://nbn-resolving.org/urn:nbn:de:0111-pedocs-101173).
- Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of educational research*, 60(1), 1–64.
- Madsen, A. M., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2012). Differences in visual attention between those who correctly and incorrectly answer physics problems. *Physical Review Special Topics-Physics Education Research*, 8(1), 010122.
- Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur Rheinland-Pfalz. (2015). Lehrplan Mathematik. Grund- und Leistungsfach in der gymnasialen Oberstufe (Mainzer Studienstufe). Anpassung an die Bildungsstandards für die allgemeine Hochschulreife. Stand März 2015.
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. Springer.
- Nemirovsky, R., & Rubin, A. (January 1992). *Students' tendency to assume resemblances between a function and its derivative*. Cambridge, MA: TERC Communications. Working Paper 2–92.
- Planinic, M., Ivanjek, L., Susac, A., & Milin-Sipus, Z. (2013). Comparison of university students' understanding of graphs in different contexts. *Physical Review Special Topics Physics Education Research*, 9, 020103c.
- Pruscha, H. (2006). *Statistisches Methodenbuch: Verfahren, Fallstudien*. Springer.
- Rost, J. (2004). *Testtheorie und Testkonstruktion [Test theory and test construction]*. Hans Huber.
- Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (S. 71–78).
- Scheiter, K., & Eitel, A. (2017). The use of eye tracking as a research and instructional tool in multimedia learning. In *Eye-tracking technology applications in educational research* (S. 143–164). IGI Global.
- Scheiter, K., Schubert, C., Schüler, A., Schmidt, H., Zimmermann, G., Wassermann, B., & Eder, T. (2019). Adaptive multimedia: Using gaze-contingent instructional guidance to provide personalized processing support. *Computers & Education*, 139, 31–47.
- Schindler, M., Schaffernicht, E., & Lilienthal, A. (2020). *Identifying student strategies through eye tracking and unsupervised learning: The case of quantity recognition*. In 44th Conference of the International Group for the Psychology of Mathematics Education, Khon Kaen University, Thailand (Virtual Meeting), July 21–22, 2020 (S. 518–527).
- Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. *Adaptive technologies for training and education*, 7(27), 1–35.

- Tobii AB. (2015). Accuracy and precision Test report: Tobii Pro X3-120 fw 1.7.1. <https://www.tobii.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-pro-x3-120-accuracy-and-precision-test-report.pdf>.
- Ubuz, B. (2007). Interpreting a graph and constructing its derivative graph: Stability and change in students' conceptions. *International Journal of Mathematical Education in Science and Technology*, 38(5), 609–637.
- Vock, M., & Gronostaj, A. (2017). *Umgang mit Heterogenität in Schule und Unterricht*. Friedrich-Ebert-Stiftung, Abt. Studienförderung.

Eye-Tracking als Methode hypothesenprüfender Forschung in den Naturwissenschaftsdidaktiken – Neue Einblicke in die Effekte von Demonstrationsexperimenten

Stefanie Lenzer und Andreas Nehring

Inhaltsverzeichnis

9.1	Einleitung	145
9.2	Planung und Durchführung einer konfirmatorischen Eye-Tracking-Studie – Von der Fragestellung bis zur Auswertung	147
9.3	Ausblick: Eye-Tracking in der konfirmatorischen Modellierung von Lernangebot, -prozessen und -produkten	154
	Literatur	156

9.1 Einleitung

In zahlreichen Studien wird die Methode des Eye-Trackings genutzt, um detaillierte Einblicke in bisher unzureichend erforschte Gebiete des Lehrens und Lernens in der Chemie zu bekommen (z. B. Baluyut & Holme, 2019; Cullipher & Seavian, 2015). Gleichzeitig fällt beim Eye-Tracking eine Vielzahl an Daten an, die in ihrem Umfang leicht über einen Datensatz aus Large-Scale-Studien, wie z. B. PISA oder TIMSS, hinausgeht. So schreibt ein Eye-Tracker, der eine Messung mit einer Samplingrate von 120 Hz umsetzt, in der x- und y-Koordinaten der Blickpunkte fixiert werden, 120 Datenzeilen pro Sekunde. Um die Dimensionen

S. Lenzer (✉) · A. Nehring
Institut für Didaktik der Naturwissenschaften, Fachgebiet Didaktik der Chemie,
Leibniz Universität Hannover, Hannover, Deutschland
E-Mail: lenzer@idn.uni-hannover.de

A. Nehring
E-Mail: nehring@idn.uni-hannover.de

zu verdeutlichen: In einer Messung von 100 Personen, die je 5 min dauert, kann damit ein Datensatz entstehen, in dem 3.600.000 Zeilen einer Analyse zugeführt werden müssen.

Möglicherweise ist das ein Grund dafür, dass eine Vielzahl aktueller Eye-Tracking-Studien in der Chemiedidaktik einen eher fallorientierten oder explorativen Charakter hat (Karch et al., 2018; Topczewski et al., 2017; Williamson et al., 2013). Diese Studien zeichnen sich meist durch geringe Stichprobenumfänge und eine deskriptive Auswertung der Daten aus.

Ein weiterer Grund für die Vielzahl qualitativer Eye-Tracking-Studien in der Chemiedidaktik ist, dass es trotz hoher Relevanz und vielversprechender Erkenntnisse für die fachdidaktische Forschung im Vergleich zu anderen Domänen, wie z. B. in der kognitiven Psychologie (Richter & Scheiter, 2019), kaum Beispiele für quantitative Eye-Tracking-Studien gibt. Insbesondere für Forschungsgegenstände, für die sich aufgrund umfassender empirischer Grundlagen präzise Forschungsfragen und Hypothesen ergeben, kann jedoch auch eine Eye-Tracking-Studie mit größeren Stichproben und Datenmengen sinnvoll sein (z. B. Nehring & Busch, 2018).

Am Beispiel einer konfirmatorischen Eye-Tracking-Studie zu Effekten des Aufbaus von chemischen Demonstrationsexperimenten¹ auf die visuelle Aufmerksamkeit von Lernenden (Nehring et al., 2018; Nehring & Busch, 2020) soll dieses Kapitel sowohl methodisch als auch inhaltlich illustrieren, wie eine quantitative Eye-Tracking-Studie geplant und umgesetzt werden kann. Dazu wird ein im Rahmen der Beispielstudie erprobtes methodisches Vorgehen zur Planung und Durchführung einer solchen Studie schrittweise beschrieben. Schließlich wird ausgehend von der Beispielstudie beschrieben, wie quantitativ erhobene Eye-Tracking-Daten in ein empirisch prüfbares Modell eingebunden werden können, um eine gemeinsame, konfirmatorische Modellierung von Lernangebot, -prozess und -produkt vorzunehmen.

Insgesamt sollen in diesem Kapitel die folgenden Fragen beantwortet werden:

1. Wie kann eine konfirmatorisch ausgerichtete Eye-Tracking Studie methodisch geplant und durchgeführt werden (Abschn. 9.2)?
2. Welche Rolle können quantitativ erhobene Eye-Tracking Daten bei der Aufklärung von Ergebnissen konfirmatorischer Studien spielen (Abschn. 9.3)?

¹Es sei angemerkt, dass in diesem Kapitel, dem typischen Sprachgebrauch folgend, ein weiterer „Experimentierbegriff“ angewendet wird, wenngleich in der Erforschung, Planung oder Reflexion von Chemieunterricht zwischen Demonstrationsversuchen und -experimenten unterschieden werden sollte.

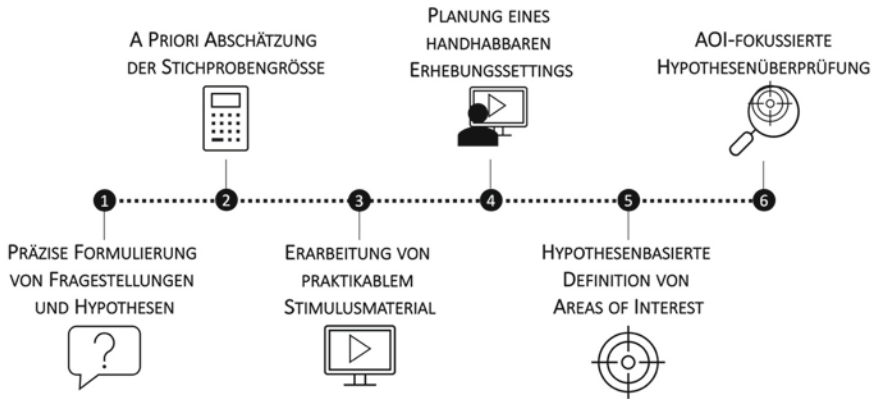


Abb. 9.1 Exemplarische Übersicht für die Planung einer konfirmatorischen Eye-Tracking-Studie

9.2 Planung und Durchführung einer konfirmatorischen Eye-Tracking-Studie – Von der Fragestellung bis zur Auswertung

Für die Planung einer konfirmatorischen Eye-Tracking-Studie gilt prinzipiell das gleiche Vorgehen wie bei der Planung jeder empirischen Studie (z. B. Döring & Bortz, 2016; Krüger et al., 2014). Im Vergleich zu anderen Studien können bei einer Eye-Tracking-Studie allerdings wesentlich größere Datenmengen aufkommen. Daher kann es hilfreich sein, bei der Planung und Durchführung einer solchen Studie Maßnahmen zur Reduktion der Daten sowie zur Fokussierung der Analyse auf konkrete Fragestellungen anzuwenden. Das vorliegende Kapitel liefert einen theoretischen Rahmen und einige bereits erprobte Maßnahmen, die bei der Planung und Durchführung einer Eye-Tracking-Studie mit großen Datenmengen helfen können. Die Maßnahmen werden am Beispiel der Untersuchung der Effekte des Gestaltgesetzes der Einfachheit für Demonstrationsexperimente auf die visuelle Aufmerksamkeit von Lernenden sowohl illustriert als auch konkretisiert (Abb. 9.1).

Präzise Formulierung von Fragestellungen und Hypothesen

Die Planung und Durchführung einer konfirmatorischen Eye-Tracking-Studie ist nur dann sinnvoll, wenn der adressierte Forschungsgegenstand bereits hinreichend empirisch erforscht ist, sodass für die Studie präzise Forschungsfragen und Hypothesen formuliert werden können (Döring et al., 2016). Ist dies nicht der Fall, sollte auf einen explorativen Untersuchungsansatz zurückgegriffen werden

(Döring et al., 2016). Es sollte insbesondere darauf geachtet werden, dass die, aus dem Stand der aktuellen Forschung, abgeleiteten Forschungsfragen und Hypothesen dem Eye-Tracking zugänglich sind. Eye-Tracking eignet sich besonders, um Fragestellungen und Phänomene zu untersuchen, die Blickhäufigkeiten und -bewegungen als mögliches Maß für die visuelle Aufmerksamkeit und deren Verlauf beinhalten (Hoffman & Subramaniam, 1995; Holmqvist et al., 2011; Kowler, 2011; Van der Stigchel et al., 2009).

Die hier präsentierte Beispielstudie hatte das Ziel, ein solches Phänomen zu untersuchen – die Effekte des Aufbaus chemischer Demonstrationsexperimente auf die visuelle Aufmerksamkeit von Lernenden. Der Stand der theoretischen Arbeiten war mit der Formulierung konkreter Gestaltgesetze, die Lehrende beim Aufbau von Demonstrationsexperimenten beachten sollten (Schmidkunz, 1983, 1992, 2002, 2004; Schmidkunz & Büttner, 1992), und aktuellen Erkenntnissen aus dem Forschungsgebiet des Multimedia-Designs (z. B. Mayer & Mayer, 2005; Mayer, 2010; Renkl & Scheiter, 2017; Ware, 2008) derartig hoch, dass eine präzise Ableitung von Forschungsfragen und Hypothesen möglich war. Basierend auf Forschungsergebnissen aus der Gestaltpsychologie (Bruce et al., 2003; Schultz & Schultz, 2015), hat der Chemiepädagoge Heinz Schmidkunz Gestaltgesetze formuliert, die Lehrenden Regeln für den Aufbau von Demonstrationsexperimenten an die Hand geben sollten (Schmidkunz, 1983, 1992, 2002, 2004; Schmidkunz et al., 1992) – darunter das *Gesetz der Einfachheit*, nach dem alle Geräte und Apparaturen, die für die Durchführung, das Verständnis und die Erklärung eines Experiments nicht relevant bzw. zentral sind, entfernt werden sollten, um die Komplexität des Experimentalaufbaus zu reduzieren (Schmidkunz et al., 1992). Der Einsatz dieses Gesetzes solle es Lehrende ermöglichen, die visuelle Aufmerksamkeit der Lernenden gezielt auf die für ein konzeptionelles Verständnis zentralen Teile des Experimentalaufbaus zu lenken (Schmidkunz et al., 1992). Eine Untersuchung der Implementation dieser Gesetze sollte daher eine Erhöhung der visuellen Aufmerksamkeit auf zentralen Apparaturen des Demonstrationsexperiments anzeigen.

Basierend auf diesen skizzierten theoretischen Überlegungen fokussierte die Beispielstudie die folgende Forschungsfrage und Hypothese:

Forschungsfrage: Inwieweit beeinflusst die Implementation des Gestaltgesetzes der Einfachheit die visuelle Aufmerksamkeit auf zentralen Apparaturen des Aufbaus eines chemischen Demonstrationsexperiments?

Hypothese: Der Aufbau eines chemischen Demonstrationsexperimentes nach dem Gestaltgesetz der Einfachheit führt zu einer höheren visuellen Aufmerksamkeit auf zentralen Apparaturen als ein Aufbau, der dieses Gesetz nicht berücksichtigt.

A-priori-Abschätzung der Stichprobengröße und Beurteilung der Umsetzbarkeit

Ebenso wie in anderen quantitativen Untersuchungen kann eine zu kleine Stichprobe zu einer Unterpowerung der Studie führen und die Forschenden vor das

Problem einer schweren Interpretierbarkeit nichtsignifikanter Ergebnisse stellen (Döring et al., 2016). Anhand der erwarteten und als substanziell eingeschätzten Effekte sowie des zu rechnenden Signifikanztests sollte zuvor eine A-priori-Stichprobenberechnung durchgeführt werden; dazu eignet sich z. B. die Freeware G*Power 3.1 (Faul et al., 2007) oder das Monte-Carlo-basierte Simulationsverfahren z. B. in MPLUS (Muthén & Muthén, 2009). Sollen neben der Methode des Eye-Trackings weitere Methoden, wie z. B. Fragebögen oder Paper-Pencil-Tests, zur Erhebung der Daten genutzt werden, ist bei der Stichprobenplanung auch die Power dieser Tests bzw. die Größe der Zusammenhänge, die man zwischen Eye-Tracking-Daten und den weiteren Instrumenten untersucht, zu beachten.

Für die Beispielstudie wurde durch eine A-priori-Stichprobenberechnung mithilfe der Freeware G*Power 3.1 die benötigte Stichprobengröße ermittelt (Faul et al., 2007). Mit dem Schwerpunkt auf großen Effekten, einem einseitigen Mann-Whitney-U-Test und einer Fehlerwahrscheinlichkeit von $p = .05$ zeigte die Software eine Mindeststichprobengröße von 106 Proband*innen an. Bei der Datenerhebung an zwei Schulen wurde schließlich eine Stichprobengröße von $n = 146$ Proband*innen erreicht. 74.1 % der Proband*innen befanden sich in der Sekundarstufe I und 25.9 % in der Sekundarstufe II, 17.6 % der Proband*innen trugen eine Brille, und 59.6 % waren weiblich (Nehring et al., 2018, 2020). Die vorherige Abschätzung der Stichprobengröße bot die Möglichkeit, signifikante Effekte zu erkennen, ohne die Studie durch ein übermäßiges Datenaufkommen zu überwältigen, d. h., auch durch diese Maßnahme erfolgte eine Reduktion der Daten. Post hoc betrachtet wurde so eine Power von $\beta - 1 = 0.90$ für mittlere Effekte (ausgedrückt in Cohens d) erreicht.

Erarbeitung von geeignetem Stimulusmaterial

Bei der Erstellung von geeignetem Stimulusmaterial sollten vor allem die Eigenschaften und Spezifika des Eye-Trackers einbezogen werden. Prinzipiell eignet sich dazu verschiedenstes Material, das den Proband*innen in live oder in Form von Bildern oder Videos vorgelegt werden kann. Je nachdem welches Material genutzt wird, kann es zu Messungenauigkeiten (z. B. bei zu überfüllten Bildern) oder zu einer erhöhten Datenmenge (z. B. durch den Einsatz von Videos) kommen (Holmqvist et al., 2011). Um einem zu hohen Datenaufkommen entgegenzuwirken, werden in Eye-Tracking-Studien daher häufig unbewegte Bilder als Stimulusmaterial verwendet (z. B. Scheiter & Eitel, 2017).

Im Rahmen der konfirmatorischen Beispielstudie wurde basierend auf der konkreten Forschungsfrage nur eine Momentaufnahme des Experiments untersucht. Daher konnte auch in dieser Studie der Einsatz von unbewegten Bildern dazu beitragen, eine potenziell hohe Datenmenge zu reduzieren. Basierend auf den theoretischen Grundlagen zu den Gestaltgesetzen (Schmidkunz et al., 1992) wurden Bilder entwickelt, die jeweils eine Fotografie des Aufbaus eines Demonstrationsexperiments zeigten (Nehring et al., 2018). Ein Bild zeigte

den Aufbau des Demonstrationsexperiments, bei dem das Gestaltgesetz der Einfachheit umgesetzt wurde (*Gestaltsetting*; Abb. 9.2 oben). Das andere Bild zeigte den gleichen Experimentalaufbau, wobei das Gestaltgesetz verletzt wurde (*Kontrollsetting*; Abb. 9.2 unten). Die Apparaturen auf den Fotos enthielten bewusst keine Chemikalien, da diese sich während des Experiments typischerweise verändern, z. B. durch einen Farbumschlag oder die Veränderung des Aggregatzustands. Diese Veränderungen hätten die visuelle Aufmerksamkeit zusätzlich beeinflussen können, ohne mit den Gestaltgesetzen selbst in Zusammenhang zu stehen. Auf diese Weise wurde versucht, die interne Validität zu erhöhen. Die externe Validität wurde durch die Verwendung von Geräten und Experimenten berücksichtigt, die traditionell auch im Chemieunterricht oder in Vorlesungen zum Einsatz kommen. Für die Studie wurde das Boyle-Experiment ausgewählt, welches das Prinzip der Massenerhaltung demonstriert. Üblicherweise wird zur Demonstration ein mit Holzkohle und Sauerstoff gefüllter Rundkolben mithilfe eines Ballons verschlossen. Die Holzkohle wird über dem Bunsenbrenner entzündet und verbrennt. Der Ballon bläst sich auf. Das Gewicht des Rundkolbens wird vor und nach dem Verbrennen der Holzkohle mithilfe einer Analysewaage ermittelt. Die Masse ist zu beiden Zeitpunkten dieselbe. Für die Durchführung des Experiments sind die Waage, der Rundkolben, der Ballon und der Bunsenbrenner relevant. Nach dem Gesetz der Einfachheit sollten alle anderen Geräte, die für die Durchführung nicht relevant sind, entfernt werden, da somit die Komplexität des Versuchsaufbaus reduziert wird. Daher zeigte die Abbildung für das Gestaltsetting nur die für die Durchführung des Experiments relevanten Geräte, während das Kontrollsetting weitere Geräte zeigte, z. B. ein Becherglas oder weitere Stopfen (Abb. 9.2).

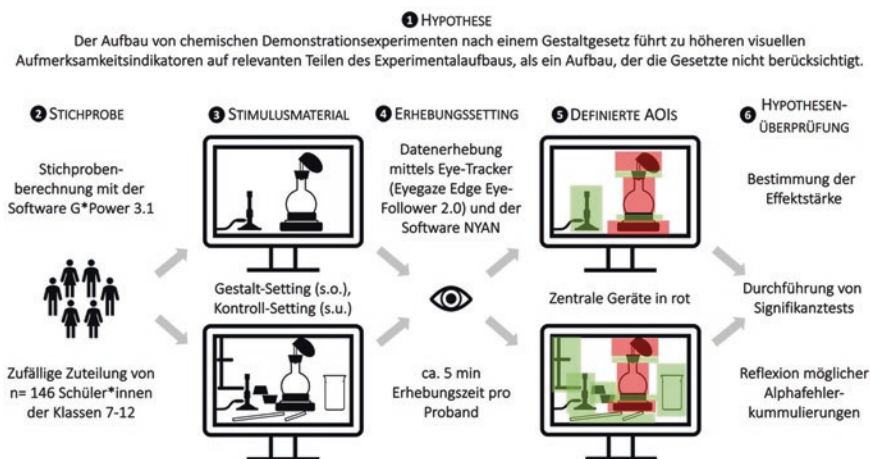


Abb. 9.2 Übersicht über das Design und die Durchführung der Beispielstudie

Planung eines handhabbaren Erhebungssettings

Da der Stichprobenumfang im Fall einer solchen quantitativen Studie relativ groß ist, sollte auch das Erhebungssetting im Vorfeld ausführlich geplant werden (Holmqvist et al., 2011). Es sollten u. a. die folgenden Aspekte beachtet werden:

- Wie viele Räume und Personen werden benötigt, um die geplante Stichprobe zu koordinieren?
- Wie müssen die Räume beschaffen sein?
- Welche Aspekte sind zu beachten, um Störfaktoren zu minimieren?
- Welche Anweisungen benötigen die Proband*innen im Vorfeld?
- Welche Eye-Tracker werden benötigt – stationäre oder mobile?
- Wie viele Eye-Tracker werden benötigt?
- Welche weitere technische Ausstattung wird benötigt – Computer und Monitore?

Die Datenerhebung für die Beispielstudie wurde an zwei Schulen in Hannover durchgeführt.

Es wurde ein stationärer Eye-Tracker genutzt, da aufgrund der Hypothesen vordefinierte AOIs untersucht wurden (s. unten). Ein mobiler Eye-Tracker wäre in diesem Fall ungeeignet gewesen, da die AOIs mit einem solchen Gerät, z. B. aufgrund von Kopfbewegungen der Proband*innen, nicht so genau bestimmt werden können. Durch diese Ungenauigkeiten kann der Einsatz eines mobilen Eye-Trackers unter Umständen auch zu einer größeren Menge an Daten führen. Der Eye-Tracker wurde jeweils in einem ruhigen und abgedunkelten Raum innerhalb der Schule aufgestellt. Eine solche Umgebung ist extrem wichtig für die Erhebung von Eye-Tracking-Daten, denn dadurch können die äußeren Einflüsse auf die Blickbewegungen der Proband*innen (Störfaktoren) minimiert und somit das gesamte Erhebungssetting optimiert werden (Holmqvist et al., 2011). Den Proband*innen wurde mitgeteilt, dass sie an einer Studie zu chemischen Demonstrationsexperimenten teilnehmen würden. Die zugrunde liegenden Gestaltungsgesetze wurden im Vorfeld nicht thematisiert. Nach der Kalibrierung wurde den Proband*innen eine offen formulierte Anweisung gezeigt: „Im Folgenden werden Ihnen sechs Bilder von chemischen Demonstrationsexperimenten gezeigt. Beobachten Sie sie sorgfältig.“ Den Proband*innen wurden zufällig Bilder von je drei Gestaltungssettings und drei Kontrollsettings gezeigt. Durch die Verwendung von zwölf Bildserien mit unterschiedlichen Sequenzen, in denen jedes Bild an einer anderen Position des Bildschirms angezeigt wurde, wurde versucht, Reihenfolgeeffekte (Rammstedt, 2004) über die Bilder hinweg zu reduzieren. Vor jedem Bild wurde ein weißer Bildschirm ohne weitere Details oder Fixierungspunkte gezeigt. Jedes Bild wurde für 20 s gezeigt; der weiße Bildschirm wurde für je 5 s gezeigt (Kowler, 2011; Nehring et al., 2018). Insgesamt dauerte die Eye-Tracking-Sitzung pro Proband*in etwa 5 min.

Als Messinstrument für die Beispielstudie wurde ein Eyegaze Edge Eye-Follower 2.0 mit NYAN-Software verwendet. Er wurde unter einem ASUS-VN248-Computermonitor (24 Zoll) befestigt und enthielt vier Kameras zur Messung von Augenfixationen und Sakkaden. Da der Eye-Tracker freie Kopfbewegungen erlaubte (51 cm hin und her, 76 cm von links nach rechts), musste der Kopf nicht fixiert werden, was für die teilnehmenden Proband*innen eine etwas unnatürliche Situation geschaffen hätte. Der Eye-Tracker hat eine Abtastrate von 120 Hz und eine automatische Augenerfassung mit Gesichtserkennung. Das Gerät verfolgt beide Augen über den gesamten Bereich der Kopfbewegung. Alle Proband*innen wurden mithilfe einer automatischen 9-Punkt-Kalibrierung kalibriert, bevor mit der Augenerfassung begonnen wurde. Es wurde eine Fixationsschwelle von 100 ms verwendet. Die Bilder, die den Proband*innen gezeigt wurden, erschienen auf einem Computerbildschirm mit einer Auflösung von 1920×1080 Pixeln (Full HD). Der Abstand zwischen den Proband*innen und dem Bildschirm wurde auf 60 cm eingestellt, indem ein weißer Streifen auf den Boden gezeichnet und der Stuhl, auf dem die Proband*innen sitzen würden, auf den Streifen gestellt wurde. Nach den Angaben des Herstellers ist eine Messungenauigkeit von 0.5° bei der Bestimmung der Blickrichtung zu berücksichtigen. Die Kalibrierdaten des Eye-Trackers zeigten eine Ungenauigkeit von 0.45° . Nimmt man 0.5° und 60 cm als Berechnungswerte, bedeutet dies, dass auf dem Bildschirm ein Kreis mit einem Radius von 0.5 cm nicht mit ausreichender Genauigkeit erkannt werden kann. Folglich wurden Objekte, die kleiner als ein Kreis von 1 cm waren, nicht in die Analyse einbezogen.

Hypothesenbasierte Definition von Areas of Interest (AOIs)

Im Vergleich zu einer explorativen Studie kann es für die Auswertung einer konfirmatorischen Eye-Tracking-Studie sinnvoll sein, dass bereits im Vorfeld Kategorien bzw. AOIs definiert werden. Sie geben vor, an welchen Punkten mit der Analyse der Daten angesetzt werden sollte, und helfen, aus den großen Datenmengen zu selektieren bzw. bereits im Vorfeld unpassende Daten für die Analyse auszuschließen (Holmqvist et al., 2011). Zusätzlich können mögliche Unterstützungen der Eye-Tracking-Software, wie z. B. Bildungen von Mittelwerten- und Standardabweichungen auf den AOIs, genutzt werden, um das Datenaufkommen an dieser Stelle zu reduzieren (Holmqvist et al., 2011).

In der Beispielstudie wurden die Bilder, die den Proband*innen gezeigt wurden, mit AOIs versehen. Die AOIs wurden durch das Gestaltgesetz vorgegeben, wobei zwischen AOIs für zentrale (rot) und für nichtzentrale (grün) Geräte und Apparaturen unterschieden wurde (Abb. 9.2). Ein bestimmtes Gerät oder eine Apparatur im Experimentalaufbau wurde dann als „zentral“ erachtet, wenn mit dessen Hilfe das demonstrierte Prinzip nachvollzogen und erklärt werden konnte. Für die Auswertung bzw. Erklärung in Bezug auf das Prinzip der Massenerhaltung ist die Masse im Kolben vor und nach der „Reaktion“ bedeutsam, d. h., von den dargestellten Geräten waren die Waage, der Rundkolben sowie

der Ballon zentral. Da der Bunsenbrenner relevant für die Durchführung, aber nicht für die Erklärung des Prinzips ist, wurde er auf dem Bild dargestellt, aber nicht als zentrale AOI definiert.

AOI-fokussierte Hypothesenprüfung

Um die Datenkomplexität an dieser Stelle zu reduzieren, sollte sich die Datenanalyse ausschließlich auf die zuvor definierten AOIs fokussieren (s. oben). Für jede dieser AOIs können dann die Gesamtzahl der Fixationen, die Gesamtfixationsdauer, die gesamte mittlere Fixationsdauer und die Zeit bis zur ersten Fixation betrachtet werden (Holmqvist et al., 2011; Scheiter & Eitel, 2017). Für die Prüfung der Hypothesen können dann die Daten der fokussierten AOIs aggregiert und Untersuchungen auf Signifikanz und Effektstärken unterworfen werden. Anschließend sollte noch eine Reflexion möglicher Alphafehlerkumulierungen erfolgen (Döring et al., 2016).

Um im Rahmen der Beispielstudie die Unterschiede an visueller Aufmerksamkeit zu ermitteln, die in einer bestimmten AOI zwischen Gestalt- oder Kontrollsetting auftraten, wurden, wie zuvor beschrieben, die Gesamtzahl der Fixationen, die Gesamtfixationsdauer, die gesamte mittlere Fixationsdauer und die Zeit bis zur ersten Fixation betrachtet. Die Zeit bis zur ersten Fixation wurde als Indikator für den Umfang der erforderlichen visuellen Suche genutzt (Rayner, 2009). Im Durchschnitt reichte eine typische Fixation von 200–300 ms.

Da die Eye-Tracking-Daten, z. B. die Zeit bis zur Fixation, nicht alle Anforderungen für t-Tests (Normalverteilung und Homogenität der Varianzen) erfüllten, wurden zur Überprüfung der statistischen Signifikanz einseitige Mann-Whitney-U-Tests verwendet. Je nachdem, wie spezifisch im Vorfeld vorausgesagt wurde, dass die Indikatoren für die visuelle Aufmerksamkeit in einem Gestaltsetting höher sein würden, wurde ein einseitiger oder zweiseitiger Test genutzt, um die Signifikanz zu überprüfen. Für die Eye-Tracking-Indikatoren (Tab. 9.1) wurden bidirektionale Tests verwendet. Dann wurden die Effektgrößen r berechnet. Nach Cohen (Cohen, 1992; Tang & Pienta, 2012) stellen r -Koeffizienten zwischen 0.10 und 0.30 kleine Effekte dar. Bereiche von 0.30 bis 0.50 repräsentieren mittlere Effekte, und r -Koeffizienten über 0.50 entsprechen großen Effekten. Für das Gesetz der Einfachheit wurden die Eye-Tracking-Daten im Hinblick auf drei zentrale Geräte des Experimentalaufbaus überprüft (Waage, Rundkolben, Ballon; in Abb. 9.2 rot hervorgehoben), wobei der Ballon aus Gründen der Messungengenauigkeit nicht in die weitere Hypothesenüberprüfung mit einging. Die Ergebnisse für die AOIs des Prinzips der Einfachheit sind in Tab. 9.1 dargestellt. Basierend auf der Hypothese, dass es den Lernenden ohne Ablenkung durch unnötige Geräte leichter fällt, sich auf zentrale Geräte zu fokussieren, wurden zur Überprüfung der Signifikanz unidirektionale Tests genutzt. Die Daten in Tab. 9.1 zeigen signifikante Unterschiede mit geringen Effekten zwischen dem Gestalt- und dem Kontrollsetting. Im Gestaltsetting wurden die Waage und der Rundkolben insgesamt häufiger und länger fixiert. Die Zeit bis zur ersten

Tab. 9.1 Daten für die Eye-Tracking-Indikatoren der visuellen Aufmerksamkeit

Eye-Tracking Indikatoren	Erprobung des Gesetzes der Einfachheit				Effektstärke/Richtung		
	Gestalt- Setting, <i>M</i>	Gestalt- Setting, <i>SD</i>	Kontroll- Setting, <i>M</i>	Kontroll- Setting, <i>SD</i>	<i>z</i>	<i>p</i>	<i>r</i>
<i>Gesamtzahl der Fixationen</i>	7,27	5,81	4,43	3,63	-3,82	<0.001	0,32
<i>Gesamtfixationsdauer in s</i>	2,54	2,06	1,35	1,10	-4,41	<0.001	0,36
<i>Gesamte mittlere Fixationsdauer in s</i>	0,38	0,37	0,32	0,21	-1,95	0.026	0,16
<i>Zeit bis zur ersten Fixation in s</i>	4,36	3,72	6,42	5,08	-3,14	0.001	0,26

Fixierung war im Gestaltsetting insgesamt kürzer. Somit konnte die zu Beginn aufgestellte Hypothese für das Gesetz der Einfachheit bestätigt werden. Der Aufbau des Experiments nach dem Gesetz der Einfachheit führte zu einer höheren visuellen Aufmerksamkeit auf zentralen Geräten des Experimentalaufbaus als ein Aufbau, der dieses Gesetz nicht berücksichtigte.

9.3 Ausblick: Eye-Tracking in der konfirmatorischen Modellierung von Lernangebot, -prozessen und -produkten

Eye-Tracking-Daten haben erst dann einen größeren Wert für das Lehren und Lernen, wenn sie an weitere Maße geknüpft werden (z. B. Richter et al., 2019). So ermöglichen die Eye-Tracking-Daten aus der zuvor beschriebenen Studie zwar Aussagen über die Effekte der Gestaltgesetze im Bereich visueller Aufmerksamkeit, jedoch keine Aussagen darüber, ob Lernende tatsächlich auch höhere Lernerfolge erzielen, wenn Gestaltgesetze, wie das Gesetz der Einfachheit, auf Demonstrationsexperimente angewendet werden (Nehring et al., 2018). Das Ziel weiterführender Forschung in diesem Bereich sollte sein, über die Prüfung einzelner Zusammenhänge oder Unterschiede hinauszugehen und Eye-Tracking-Daten in eine Modellierung von Merkmalen des Lernangebots, der Lernprozesse bzw. Angebotsnutzung und der Produkte zu integrieren. Insbesondere die Fülle von Daten, die Eye-Tracking-Studien anspruchsvoll macht, forciert eine Einbindung in konfirmatorische Modelle, die eine Prüfung von Zusammenhängen, aber auch von Interaktions-, Mediations- und Moderationseffekten ermöglichen. Daher wurde für die Anschlussforschung an die zuvor beschriebene Eye-Tracking-Studie in Anlehnung an das Angebots-Nutzungs-Modell nach Vieluf et al. (2020) ein Pfadmodell (Angebots-Prozess-Produkt-Mediationsmodell) entwickelt, das im Rahmen zukünftiger Forschungsprojekte detaillierter beforscht

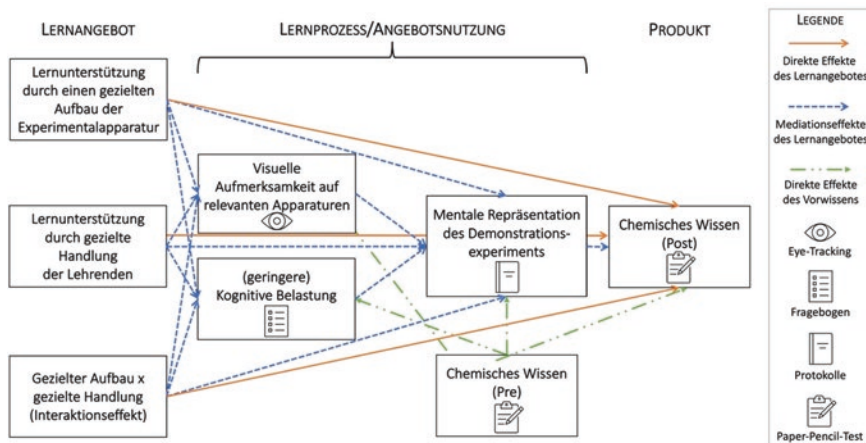


Abb. 9.3 Das als Teil der Anschlussforschung entwickelte Angebots-Prozess-Produkt-Mediationsmodell verknüpft Theorie, Studiendesign und Datenanalyse

wird (Abb. 9.3). In diesem Modell werden zusätzlich zu den Effekten der Lernunterstützung durch die Gestehtgesetze auch die Effekte der Lernunterstützung durch das Handeln von Lehrenden beim Einsatz von Demonstrationsexperimenten abgebildet. Dazu wurden die Merkmale des Lernangebotes, d. h. der unterschiedlichen Lernunterstützung, in der Funktion von unabhängigen, exogenen Variablen integriert. Basierend auf verschiedenen Vorarbeiten postuliert das Modell, dass sowohl die Lernunterstützung durch einen gezielten Aufbau der Experimentalapparatur (Nehring et al., 2018) als auch die Lernunterstützung durch gezielte Handlung der Lehrenden (Börlin, 2012) sowie die Interaktion beider Elemente zu einer Erhöhung der visuellen Aufmerksamkeit auf inhaltlich zentrale Apparaturen des Experiments und zu einer Verringerung der kognitiven Belastung beim Lernen führen (Mayer & Moreno, 2003). Die Lernunterstützungen, die gesteigerte visuelle Aufmerksamkeit und die geringere kognitive Belastung könnten schließlich auch zu adäquateren Repräsentationen über das Experiment und einem erhöhten chemischen Wissen führen. Der nachweislichen Bedeutung des bereichsspezifischen Vorwissens (Richter et al., 2018; Schauble, 1990) wird Rechnung getragen, indem das chemische Wissen der Lernenden als eigenständige exogene Variable in das Modell aufgenommen wurde. In diesem Zusammenhang besteht die Annahme, dass auch mit höherem Vorwissen eine höhere visuelle Aufmerksamkeit auf zentralen Apparaturen, eine geringere kognitive Belastung, adäquatere mentale Repräsentationen und ein höheres Wissen nach dem Lernsetting einhergehen (Amadiou et al., 2009; Cook et al., 2008).

Um die vermuteten Effekte zu überprüfen, wird der Lernprozess bzw. die Nutzung des Lernangebotes im Modell sowohl über Eye-Tracking-Daten als Indikatoren der visuellen Aufmerksamkeit als auch über Daten zur kognitiven Belastung (Mayer et al., 2003), die mithilfe von Fragebögen erhoben werden,

sowie durch Daten zur mentalen Repräsentation der Demonstrationsexperimente, die über chemietypische Protokolle der Proband*innen erhoben werden, abgebildet. Ein Pre- und Posttest zum chemischen Wissen (Papier-Pencil-Test) ermöglicht eine Quantifizierung des Standes des Vorwissens sowie des Lernerfolgs und -produkts. Die Prüfung, ob das Vorwissen moderierend auf die Zusammenhänge im Modell wirkt, ist ebenso möglich, wenngleich sich die Stichprobenanforderungen an die Modellierung durch Einführung von *random slopes* erhöhen würden (Muthén & Muthén, 2009).

Aus einer übergeordneten Perspektive stellen solche „Angebots-Prozess-Produkt-Mediationsmodelle“ (Vieluf et al., 2020) eine besondere Innovation naturwissenschaftsdidaktischer Forschung dar: Im Falle von experimentellen Interventionsdesigns werden Lernprozesse nicht mehr als Black Box behandelt, sondern gezielt beleuchtet und gar mit Merkmalen des Lernangebots und der Lernprodukte geprüft.

Danksagung Die Autor*innen danken allen Beteiligten für die Unterstützung bei der Durchführung der Studie, insbesondere Sebastian Busch sowie allen Proband*innen, die an der Studie teilgenommen haben.

Literatur

- Amadiou, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction*, 19(5), 376–386. <https://doi.org/10.1016/j.learninstruc.2009.02.005>.
- Baluyut, J. Y., & Holme, T. A. (2019). Eye tracking student strategies for solving stoichiometry problems involving particulate nature of matter diagrams. *Chemistry Teacher International*, 1(1). <https://doi.org/10.1515/cti-2018-0003>.
- Bruce, V., Green, P. R., & Georgeson, M. A. (2003). *Visual perception: Physiology, psychology, & ecology*. Psychology Press.
- Börnin, J. (2012). *Das Experiment als Lerngelegenheit: Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität* (Bd. 132). Logos Verlag.
- Cook, M., Wiebe, E. N., & Carter, G. (2008). The influence of prior knowledge on viewing and interpreting graphics with macroscopic and molecular representations. *Science Education*, 92(5), 848–867. <https://doi.org/10.1002/sce.20262>.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Cullipher, S., & Sevian, H. (2015). Atoms versus bonds: How students look at spectra. *Journal of Chemical Education*, 92(12), 1996–2005. <https://doi.org/10.1021/acs.jchemed.5b00529>.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41089-5_1.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–219. <https://doi.org/10.3758/BF03193146>.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552. <https://doi.org/10.1007/s10648-011-9174-7>.
- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, 57(6), 787–795. <https://doi.org/10.3758/BF03206794>.

- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Karch, J. M., García Valles, J. C., & Sevian, H. (2018). Looking into the black box: Using Gaze and Pupillometric data to probe how cognitive load changes with mental tasks. *Journal of Chemical Education*, 95(10), 1724–1735. <https://doi.org/10.1021/acs.jchemed.9b00014>.
- Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, 51(13), 1457–1483. <https://doi.org/10.1016/j.visres.2010.12.014>.
- Krüger, D., Parchmann, I., & Schecker, H. (2014). Methoden in der naturwissenschafts-didaktischen Forschung. Springer Spektrum, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37827-0_1.
- Mayer, R., & Mayer, R. E. (Hrsg.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6.
- Mayer, R. E. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno, & R. Brünken (Hrsg.), *Cognitive load theory* (S. 131–152). Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744.009>.
- Muthén, L. K., & Muthén, B. O. (2009). How to use a Monte Carlo study to decide on sample size and determine how to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4), 599–620. <https://doi.org/10.1207/S15328007SEM0904>.
- Nehring, A., & Busch, S. (2018). Chemistry demonstrations and visual attention: Does the setup matter? Evidence from a double-blinded eye-tracking study. *Journal of Chemical Education*, 95(10), 1724–1735. <https://doi.org/10.1021/acs.jchemed.8b00133>.
- Nehring, A., & Busch, S. (2020). Setting up chemistry demonstrations according to the left-to-right principle: An eye-movement-pattern-based analysis. *Journal of Chemical Education*, 97(1), 275–282. <https://doi.org/10.1021/acs.jchemed.9b00102>.
- Rammstedt, B. (2004). *Zur Bestimmung der Güte von Multi-Item-Skalen: Eine Einführung (GESIS-How-to, 12)*. Zentrum für Umfragen, Methoden und Analysen (ZUMA). <https://nbn-resolving.org/urn:nbn:de:0168-ssao-201443>.
- Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. <https://doi.org/10.1080/2F17470210902816461>.
- Renkl, A., & Scheiter, K. (2017). Studying visual displays: How to instructionally support learning. *Educational Psychology Review*, 29(3), 599–621. <https://doi.org/10.1007/s10648-015-9340-4>.
- Richter, J., & Scheiter, K. (2019). Studying the expertise reversal of the multimedia signaling effect at a process level: Evidence from eye tracking. *Instructional Science*, 47(6), 627–658. <https://doi.org/10.1007/s11251-019-09492-3>.
- Richter, J., Scheiter, K., & Eitel, A. (2018). Signaling text–picture relations in multimedia learning: The influence of prior knowledge. *Journal of Educational Psychology*, 110(4), 544. <https://doi.org/10.1037/edu0000220>.
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, 49(1), 31–57. [https://doi.org/10.1016/0022-0965\(90\)90048-D](https://doi.org/10.1016/0022-0965(90)90048-D).
- Scheiter, K., & Eitel, A. (2017). The use of eye tracking as a research and instructional tool in multimedia learning. In *Eye-tracking technology applications in educational research* (S. 143–164). IGI Global.
- Scheiter, K., & Van Gog, T. (2009). Introduction: Using eye tracking in applied research to study and stimulate the processing of information from multi-representational sources. *Applied Cognitive Psychology*, 23(9), 1209–1214. <https://doi.org/10.1002/acp.1524>.

- Schmidkunz, H. (1983a). Die Gestaltung chemischer Demonstrationsexperimente nach wahrnehmungspsychologischen Erkenntnissen. *Naturwissenschaften im Unterricht Chemie*, 31(10), 360–367.
- Schmidkunz, H. (1983b). Die Gestaltung chemischer Demonstrationsexperimente nach wahrnehmungspsychologischen Erkenntnissen. *Naturwissenschaften im Unterricht Physik*, 31(10), 360–367.
- Schmidkunz, H. (1992) Die Symmetrie als Prägnanz-bildender Faktor bei Demonstrationsexperimenten. *Naturwissenschaften im Unterricht Chemie*, 3, 12–16.
- Schmidkunz, H., & Büttner, D. (1992) Aufbau und Durchführung chemischer Demonstrationsexperimente. *Naturwissenschaften im Unterricht Chemie*, 14, 130–135.
- Schmidkunz, H. (2002) *Organisationsformen: Demonstrationsexperiment und Schülerexperiment*. In P. Pfeifer, K. Häusler, & B. Lutz. (Hrsg.), *Konkrete Fachdidaktik Chemie* (S. 301–307). Oldenbourg Wissenschaftsverlag.
- Schmidkunz, H. (2004). Das Gesetz der Dynamik von links nach rechts bei chemischen Demonstrationen. *Chemkon*, 11(3), 137–140.
- Schultz, D. P., & Schultz, S. E. (2015). *A history of modern psychology*. Cengage Learning.
- Tang, H., & Pienta, N. (2012). Eye-tracking study of complexity in gas law problems. *Journal of Chemical Education*, 89(8), 988–994. <https://doi.org/10.1021/ed200644k>.
- Topczewski, J. J., Topczewski, A. M., Tang, H., Kendhammer, L. K., & Pienta, N. J. (2017). NMR spectra through the eyes of a student: Eye tracking applied to NMR items. *Journal of Chemical Education*, 94(1), 29–37. <https://doi.org/10.1021/acs.jchemed.6b00528>.
- Van der Stigchel, S., Belopolsky, A. V., Peters, J. C., Wijnen, J. G., Meeter, M., & Theeuwes, J. (2009). The limits of top-down control of visual attention. *Acta psychologica*, 132(3), 201–212. <https://doi.org/10.1016/j.actpsy.2009.07.001>.
- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M., & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts: Ein kritischer Vergleich verschiedener Modellvarianten. *Zeitschrift für Pädagogik* 2, 66(1/20), 63–80.
- Ware, C. (2008). *Visual thinking: For design (Morgan Kaufmann Series in Interactive Technologies)*. Morgan Kaufmann.
- Williamson, V. M., Hegarty, M., Deslongchamps, G., Williamson, K. C., & Shultz, M. J. (2013). Identifying student use of ball-and-stick images versus electrostatic potential map images via eye tracking. *Journal of Chemical Education*, 90(2), 159–164. <https://doi.org/10.1021/ed200259j>.



Fachlich vs. instruktional: Welche Aufgabenmerkmale werden bei der Schwierigkeitseinschätzung von Bruchrechenaufgaben identifiziert und evaluiert? – Eine Studie mit Eye-Tracking Stimulated Recall Interviews

Saskia Schreiter, Markus Vogel, Markus Rehm und Tobias Dörfler

Inhaltsverzeichnis

10.1	Einleitung	159
10.2	Theoretischer Hintergrund	160
10.3	Forschungsfragen	164
10.4	Material und Methode	165
10.5	Ergebnisse	168
10.6	Diskussion	171
10.7	Fazit	173
	Literatur	174

10.1 Einleitung

Die Schwierigkeit von Bruchrechenaufgaben wird neben fachlichen Aufgabenmerkmalen (z. B. gleichnamige vs. ungleichnamige Brüche; Padberg & Wartha, 2017) auch von instruktionalen Aufgabenmerkmalen beeinflusst (z. B. Split-

S. Schreiter (✉) · M. Vogel · M. Rehm · T. Dörfler
Pädagogische Hochschule Heidelberg, Heidelberg, Deutschland
E-Mail: schreiter@ph-heidelberg.de

M. Vogel
E-Mail: vogel@ph-heidelberg.de

M. Rehm
E-Mail: rehm@ph-heidelberg.de

T. Dörfler
E-Mail: doerfler@ph-heidelberg.de

Attention vs. integriertes Aufgabendesign; Sweller et al., 2011). Studienergebnisse zeigen, dass Lehrkräfte bei der Einschätzung von Aufgabenschwierigkeiten das instruktionale Aufgabendesign häufig nicht adäquat berücksichtigen (Hellmann & Nückles, 2013; Schreiter et al., 2021). Es wird vermutet, dass instruktionale Aufgabenmerkmale weniger häufig als schwierigkeitsgenerierend identifiziert und/oder bezüglich ihrer Schwierigkeit für Schüler*innen weniger häufig korrekt (d. h. heoriekonform; s. Abschn. 10.2) evaluiert werden im Vergleich zu fachlichen Aufgabenmerkmalen. Über die Prozesse, die diagnostischen Lehrerurteilen zugrunde liegen, ist bislang jedoch nur wenig bekannt (Leuders et al., 2018).

Die Metaanalyse von Südkamp et al. (2012) zeigt, dass Lehrkräfte, die über den Vergleichsmaßstab ihres Urteils informiert werden, genauere diagnostische Urteile treffen. Aufgrund des korrelativen Charakters dieser Studie bleiben Rückschlüsse hinsichtlich zugrunde liegender Urteilsprozesse explorativ (Loibl et al., 2020). Man könnte annehmen, dass Informationen, die zu bewertende Aufgabenmerkmale eingrenzen (im Folgenden *informierte Diagnose* genannt), Lehrkräfte dazu befähigen, ihre Aufmerksamkeit zu fokussieren und folglich mehr relevante Merkmale zu identifizieren. Zur empirischen Prüfung dieser Annahme bedarf es einer systematischen Variation der Informationen, die einer Lehrkraft beim Diagnostizieren zur Verfügung stehen, und einer gezielten Untersuchung der Prozesse, die dem diagnostischen Urteil zugrunde liegen (vgl. Loibl et al., 2020). Eine Möglichkeit, um solche Informationsverarbeitungsprozesse zu untersuchen, stellt die Erhebung von Eye-Tracking-Daten gefolgt von Eye-Tracking Stimulated Recall Interviews (ET SRI; s. Abschn. 10.2) dar (Schindler & Lilienthal, 2019).

Dieses Kapitel zielt darauf ab, die Identifikation und Evaluation von fachlichen und instruktionalen Aufgabenmerkmalen bei der Schwierigkeitseinschätzung von Bruchrechenaufgaben zu untersuchen. Hierbei wird ein potenzieller Einflussfaktor durch die Unterscheidung zweier Urteilsmodi (informierte vs. uninformierte Diagnose; s. Abschn. 10.4) adressiert. Die Blickbewegungen beim Diagnostizieren werden zusammen mit den zugehörigen ET SRI als Indikatoren für Identifikations- und Evaluationsprozesse herangezogen und zwischen den Experimentalbedingungen verglichen.

10.2 Theoretischer Hintergrund

Aufgabenmerkmale mit Relevanz für die Aufgabenschwierigkeit

Schwierigkeitsgenerierende Aufgabenmerkmale (Leuders & Prediger, 2016) umfassen neben inhaltspezifischen fachlichen auch instruktionale Aufgabenmerkmale.

Fachliche Aufgabenmerkmale

Brüche stellen einen zentralen Lerninhalt der Sekundarstufe I dar, der Schüler*innen häufig Probleme bereitet (z. B. Lortie-Forgues et al., 2015). Die Untersuchung typischer Schülerfehler und deren Ursachen stehen im Zentrum zahlreicher Studien (z. B. Eichelmann et al., 2012; Padberg & Wartha, 2017). Padberg (1986) gelang es, anhand von empirischen Lösungsraten eine Abfolge im Schwierigkeitsgrad für die Addition von Brüchen zu erstellen: Während die Mehrheit der Schüler*innen (85 %) Aufgaben zur Addition gleichnamiger Brüche korrekt löst, sinken die Lösungsraten für Aufgaben zur Addition ungleichnamiger Brüche (70 %) und Aufgaben, die die Addition einer natürlichen Zahl und einem Bruch verlangen (55 %). Eine häufige Fehlerart ist die getrennte Addition der Zähler und Nenner ($\frac{a}{b} + \frac{c}{d} + \frac{a+c}{b+d}$), die bei ungleichnamigen Brüchen öfter auftritt als bei gleichnamigen Brüchen (Eichelmann et al., 2012). Bei der Addition einer natürlichen Zahl und eines Bruches wird häufig fehlerhaft die natürliche Zahl zum Zähler des Bruches addiert ($n + \frac{a}{b} = \frac{n+a}{b}$), oder die Schüler*innen wählen über die Umrechnung der natürlichen Zahl in einen Bruch einen umständlichen und fehleranfälligen Rechenweg (Eichelmann et al., 2012). Padberg und Wartha (2017) begründen den hohen Schwierigkeitsgrad der Addition von Bruch und natürlicher Zahl auch damit, dass Lehrkräfte die Schwierigkeit unterschätzen und daher die Behandlung des Aufgabentyps im Unterricht vernachlässigen. Neben der Kombination von Brüchen und natürlichen Zahlen stellt auch die Präsentation der Brüche als gemischte Zahlen ein schwierigkeitsgenerierendes Merkmal dar (Padberg & Wartha, 2017).

Instruktionale Aufgabenmerkmale

Die Cognitive Load Theory (CLT; z. B. Sweller et al., 2011) unterscheidet verschiedene Arten kognitiver Belastung. Da die Kapazität des Arbeitsgedächtnisses beschränkt ist, kann ein hoher *intrinsic cognitive load* und/oder ein hoher *extraneous cognitive load* zu einer Überlastung führen (Sweller et al., 2011). Dies verhindert ein erfolgreiches Lernen (Ayres, 2006). Anders als *intrinsic cognitive load*, der sich aus der Komplexität der Aufgabe ergibt (z. B. gleichnamige vs. ungleichnamige Brüche; s. oben) und dem Lernmaterial inhärent ist, kann *extraneous cognitive load* durch Modifikationen im Instruktionsdesign reduziert werden (Sweller et al., 2019). Im Folgenden werden zwei CLT-Gestaltungsempfehlungen beschrieben, die sich in zahlreichen Studien als sehr effektiv erwiesen haben, um den *extraneous cognitive load* zu reduzieren:

Der *Split-Attention-Effekt* tritt auf, wenn Lernende ihre Aufmerksamkeit zwischen mindestens zwei aufeinander bezogene Informationsquellen teilen müssen, die räumlich getrennt präsentiert werden (Sweller et al., 2011). In der Beispielaufgabe (Abb. 10.1), sind verschiedene Informationsquellen (Wanderoute, Zeitangaben) getrennt dargeboten. Beim Lösen der Aufgabe muss ein Lernender zwischen den Informationsquellen hin und her springen und die Informationen mental integrieren. Es wird daher ein integriertes Aufgabendesign

Die Klasse 6a plant einen Wanderausflug.

Wie lange dauert die Tour bis zur Pause?

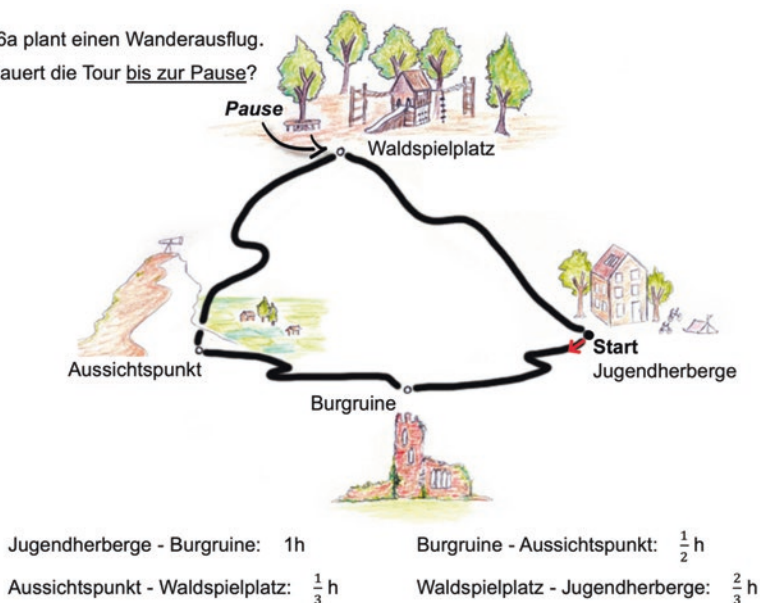


Abb. 10.1 Beispiel einer Bruchrechenaufgabe mit spezifischen schwierigkeitsgenerierenden Aufgabenmerkmalen im fachlichen (ungleichnamige Brüche, Mischung von natürlicher Zahl und Brüchen) und instruktionalen Bereich (Split-Attention, Redundanz)

empfohlen, bei dem relevante Informationen nah beieinander präsentiert werden (z. B. Kester et al., 2005). Der *Redundanzeffekt* (Sweller et al., 2011) beschreibt eine lernbeeinträchtigende Bedingung, die zustande kommt, wenn zusätzliche und für die Lösung irrelevante Informationen enthalten sind (Redundanz 1; vgl. Strecke Waldspielplatz–Jugendherberge der Beispielaufgabe in Abb. 10.1) und/oder dieselbe Information durch verschiedene Informationsquellen präsentiert wird (Redundanz 2). Zahlreiche Studien zeigen, dass die Eliminierung redundanter Informationen zu höheren Lernleistungen führt (z. B. Mayer & Moreno, 2003).

Die Effektivität dieser Gestaltungsempfehlungen hängt vom Wissensstand des Lernenden ab und ist besonders wirksam bei keinem oder geringem Vorwissen (Kalyuga et al., 2003). Für Lehrkräfte oder erfahrenere Lernende kann sich der positive Effekt der Gestaltungsempfehlungen verringern, verschwinden oder gar umkehren, was als *Expertise-Umkehr-Effekt* bezeichnet wird (Kalyuga et al., 2003). Daher ist es denkbar, dass Lehrkräfte bei der Diagnose von Aufgabenschwierigkeiten den lernförderlichen oder -hinderlichen Effekt vom Aufgabendesign nicht als solchen wahrnehmen und ihre Aufmerksamkeit auf andere Merkmale richten.

Diagnostische Lehrerurteile zur Schwierigkeit von Mathematikaufgaben

Studien belegen, dass Lehrkräfte im Allgemeinen Aufgabenschwierigkeiten häufig unzureichend einschätzen (z. B. Karing & Artelt, 2013) und im Speziellen das Instruktionsdesign von Aufgaben nicht adäquat berücksichtigen (Hellmann & Nückles, 2013). Die Metaanalyse von Südkamp et al. (2012) belegt, dass Lehrkräfte, die über den Vergleichsmaßstab ihres Urteils informiert werden, genauere diagnostische Urteile treffen. Die Urteilsgenauigkeit in dieser Metaanalyse ist als Korrelation von Lehrereinschätzung und tatsächlicher Merkmalsausprägung definiert. Aufgrund des korrelativen Charakters bleiben Rückschlüsse hinsichtlich zugrunde liegender Urteilsprozesse explorativ (Loibl et al., 2020). Informationen, die zu bewertende Aufgabenmerkmale eingrenzen, befähigen Lehrkräfte vermutlich dazu, ihre Aufmerksamkeit zu fokussieren und folglich mehr relevante Merkmale zu identifizieren. Diese Annahme wird gestützt durch Befunde von Prediger und Zindel (2017), die zeigen, dass mittels *focus questions* die Aufmerksamkeit beim Diagnostizieren auf spezifische Merkmale gelenkt werden kann. Zur empirischen Prüfung der Annahme bedarf es einer systematischen Variation der Informationen, die einer Lehrkraft beim Diagnostizieren zur Verfügung stehen, und einer gezielten Untersuchung der Informationsverarbeitungsprozesse beim Diagnostizieren (vgl. Loibl et al., 2020).

In Anlehnung an Loibl et al. (2020) kann der Prozess, der einem diagnostischen Urteil zugrunde liegt, als ein Informationsverarbeitungsprozess modelliert werden, der aus den folgenden Komponenten besteht: Schwierigkeitsgenerierende Aufgabenmerkmale können als Hinweisreize *identifiziert* oder auch *ignoriert* werden. Aufgabenmerkmale können nur dann als Hinweisreize identifiziert werden, wenn sie zuvor *wahrgenommen* wurden und ihnen ausreichend *Aufmerksamkeit* gewidmet wurde (vgl. Bless & Greifender, 2017). Bereits identifizierte Aufgabenmerkmale können hinsichtlich der Schwierigkeit für Schüler*innen als leicht oder schwer *evaluiert* werden. Dies erfordert spezifisches Wissen darüber, was Schüler*innen typischerweise leicht- oder schwerfällt. Zahlreiche Forschungsarbeiten heben die Relevanz von spezifischem Wissen über schwierigkeitsgenerierende Aufgabenmerkmale für diagnostische Lehrerurteile hervor (z. B. Ostermann et al., 2017).

Im Bereich der Bruchrechnung hat sich gezeigt, dass instruktionale Aufgabenmerkmale weniger häufig wahrgenommen und verarbeitet werden im Vergleich zu fachlichen Aufgabenmerkmalen (Schreiter et al., 2021). Unklar bleibt jedoch, wieso es zu diesen Unterschieden kommt: Werden instruktionale Aufgabenmerkmale seltener wahrgenommen, im weiteren Verarbeitungsprozess weniger häufig als schwierigkeitsgenerierend identifiziert oder bezüglich ihrer Schwierigkeit weniger häufig korrekt evaluiert?

Eye-Tracking und Stimulated Recall Interviews zur Untersuchung diagnostischer Urteilsprozesse

Die Eye-Tracking-Technologie hat sich als effektives Werkzeug erwiesen, um kognitive Prozesse zu untersuchen (z. B. Strohmaier et al., 2020). Die Analyse von Eye-Tracking-Daten basiert auf der *Eye-Mind-Hypothese* (Just & Carpenter, 1976). Diese postuliert, dass das Sehen und die kognitive Verarbeitung des Gesehenen eng miteinander verknüpft sind. Von dieser Annahme ausgehend können etwa die *Anzahl der Fixationen* und die *Fixationsdauer* spezifische Indikatoren dafür sein, wie häufig und wie lange Aufmerksamkeit auf bestimmte Areas of Interest (AOIs) gerichtet wurde (Holmqvist & Andersson, 2017). Die Anzahl an *Transitionen* zwischen zwei Objekten kann Aufschluss darüber geben, wie oft zwischen zwei Objekten hin und her geschaut wurde, etwa um die Informationen zweier Informationsquellen zu integrieren (Holmqvist & Andersson, 2017). Allerdings sind Eye-Tracking-Daten keinesfalls selbsterklärend und häufig nicht eindeutig interpretierbar (z. B. Schindler & Lilienthal, 2019; Strohmaier et al., 2020). Um diese Unsicherheiten bei der Interpretation von Eye-Tracking-Daten zu adressieren, wird eine Triangulation von Eye-Tracking mit anderen Forschungsmethoden wie ET SRI empfohlen (Wyss et al., 2020). Bei einem ET SRI werden Proband*innen aufgefordert, anhand einer Videosequenz ihrer Blickverläufe die eigenen Gedanken retrospektiv zu beschreiben (Lyle, 2003). Indem eigene Blickbewegungen sichtbar gemacht werden, kann eine tiefere Reflexionsebene der eigenen internen Prozesse angeregt werden (Stickler & Shi, 2017). Unter Berücksichtigung bestimmter methodischer Vorgaben (z. B. einer kurzen Zeitspanne zwischen Eye-Tracking und ET SRI zur Vorbeugung etwaiger Erinnerungsverluste) haben sich ET SRI als effektive Methode erwiesen, um Einblicke in die Informationsverarbeitung zu erhalten (z. B. Schindler & Lilienthal, 2019).

10.3 Forschungsfragen

Das Ziel dieser Studie ist die Untersuchung der Identifikation und Evaluation von fachlichen und instruktionalen Aufgabenmerkmalen bei der Schwierigkeitseinschätzung von Bruchrechenaufgaben. Die Schwerpunkte der Untersuchung lassen sich durch die folgenden Forschungsfragen illustrieren:

1. Welche schwierigkeitsgenerierenden Aufgabenmerkmale (fachlich vs. instruktional) werden von angehenden Lehrkräften identifiziert und evaluiert?
2. Beeinflussen zusätzliche Informationen, die zu bewertende Aufgabenmerkmale eingrenzen (informierte vs. uninformierte Diagnose) die Identifikations- und Evaluationsprozesse beim Diagnostizieren?
3. Zeigen sich Unterschiede in der visuellen Aufmerksamkeit beim Diagnostizieren zwischen informierten und uninformierten Personen?

Auf Basis der theoretischen Überlegungen (s. oben) werden die folgenden Hypothesen aufgestellt:

- Hypothese 1: Fachliche Aufgabenmerkmale werden häufiger identifiziert und/oder korrekt evaluiert im Vergleich zu instruktionalen Aufgabenmerkmalen.
- Hypothese 2a: Zusätzliche Informationen, die zu bewertende Aufgabenmerkmale eingrenzen, führen dazu, dass mehr schwierigkeitsgenerierende Aufgabenmerkmale – fachliche wie auch instruktionale – identifiziert werden.
- Hypothese 2b: Zusätzliche Informationen, die zu bewertende Aufgabenmerkmale eingrenzen, haben keinen Einfluss auf die Evaluation der identifizierten Aufgabenmerkmale, da dieser Prozess stärker Wissen darüber voraussetzt, was Schüler*innen typischerweise leicht- und/oder schwerfällt.
- Hypothese 3: Auf Basis der Annahme, dass bei einer informierten Diagnose die Aufmerksamkeit auf die zu bewertenden Aufgabenmerkmale gelenkt wird, wird bei informierten Personen eine größere Anzahl an Fixationen, längere Fixationsdauern und eine größere Anzahl an Fixationen auf vorab festgelegte fachliche und instruktionale AOIs erwartet.

10.4 Material und Methode

Forschungsdesign und Stichprobe

Den berichteten Ergebnissen liegen Daten von $N=31$ Lehramtsstudierenden des Faches Mathematik zugrunde. Die Teilnehmenden wurden randomisiert einer Experimentalgruppe ($n=15$) oder einer Kontrollgruppe zugeordnet. Die Diagnoseaufgabe bestand darin, vier Bruchrechenaufgaben möglichst genau hinsichtlich der Frage „Was macht die Aufgabe leicht/schwer für Schüler*innen?“ zu beurteilen. Die Experimentalgruppe (informierte Diagnose) erhielt zusätzliche Informationen, die zu bewertende Aufgabenmerkmale eingrenzen: „Bitte achten Sie auf die folgenden zwei Bereiche: 1. Fachliche Schwierigkeit bzgl. der Bruchrechnung, 2. Instruktionale Schwierigkeit bzgl. der Darstellung und Gestaltung der Aufgabe.“

Zur Vergleichbarkeit der Gruppen, und um mögliche Einflussfaktoren auf diagnostische Urteilsprozesse zu kontrollieren, wurden soziodemografische Daten erhoben (Tab. 10.1). Hierbei ergaben sich keine signifikanten Unterschiede zwischen den Gruppen. Da das bereichsspezifische Wissen einen potenziellen Einflussfaktor auf diagnostische Urteilsprozesse darstellt (s. Abschn. 10.2), wurde ein Vorwissenstest entwickelt (s. unten). Da dieser Vorwissenstest gleichzeitig eine Lern Gelegenheit darstellen könnte, wurde er von beiden Studiengruppen erst unmittelbar nach der Diagnoseaufgabe am Eye-Tracker bearbeitet.

Tab. 10.1 Vergleich der Studiengruppen: Soziodemografische Daten und spezifisches Vorwissen

	Kontrollgruppe <i>n</i> = 15	Experimental- gruppe <i>n</i> = 16	<i>t</i> (29)	<i>p</i>	Cohens <i>d</i>
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)			
Alter	22,0 Jahre (1,7)	22,8 Jahre (2,0)	−1,09	.288	−0,42
Praxiserfahrung	8,0 Wochen (10,5)	15,8 Wochen (13,4)	−1,80	.083	−0,65
Nachhilfeeerfahrung	59,7 h (71,7)	40,1 h (54,2)	0,86	.397	0,31
Studium Mathematik	4,8 Semester (2,7)	5,7 Semester (2,5)	−1,00	.329	−0,36
Abiturnote Mathematik	10,7 Punkte (3,0)	11,8 Punkte (2,5)	−1,09	.285	−0,39
Score Vorwissenstest*					
Fachlich	77,8 % (30,0)	66,7 % (29,8)	1,03	.310	0,37
Instruktional	58,7 % (29,7)	65,0 % (23,6)	−0,66	.515	−0,24

*Test zur Erhebung des bereichsspezifischen Vorwissens.

Da die Teilnehmenden während der Diagnoseaufgabe keinerlei Wissensinput erhalten, ist davon auszugehen, dass der Vorwissenstest den Wissensstand widerspiegelt, den die Teilnehmenden zum Zeitpunkt der Diagnoseaufgabe hatten. Die Studiengruppen unterscheiden sich hinsichtlich ihres spezifischen Vorwissens zu schwierigkeitsgenerierenden fachlichen und instruktionalen Aufgabenmerkmalen nicht signifikant voneinander (Tab. 10.1).

Diagnoseaufgabe am Eye-Tracker

Für die Diagnoseaufgabe am Eye-Tracker wurden vier Bruchrechenaufgaben in Anlehnung an typische Schulbuchaufgaben der Klassenstufe 6 erstellt. Zwischen den Aufgaben wurden schwierigkeitsgenerierende fachliche und instruktionale Merkmale gemäß der theoretischen Überlegungen in Abschn. 10.2 systematisch variiert: Die fachliche Schwierigkeit wurde durch Modifikation der Nenner (gleichnamig vs. ungleichnamig), durch die Mischung von natürlichen Zahlen und Brüchen und durch die Verwendung von gemischten Brüchen variiert. Im instruktionalen Bereich ist die Schwierigkeit auf Basis des Split-Attention-Effekts und des Redundanzeffekts variiert worden.

Bereichsspezifischer Vorwissenstest

Um das Vorwissen zu schwierigkeitsgenerierenden fachlichen und instruktionalen Aufgabenmerkmalen zu erheben, wurde ein Vorwissenstest entwickelt. Im ersten Bereich werden jeweils zwei Aufgaben, die sich nur in der fachlichen Schwierigkeit voneinander unterscheiden, gegenübergestellt präsentiert. Im zweiten Bereich unterscheiden sich gegenübergestellte Aufgaben nur in der instruktionalen Schwierigkeit. Die Teilnehmenden werden aufgefordert, die Aufgabe anzukreuzen,

die sie für Schüler*innen schwieriger einschätzen. Eine mögliche Ankreuzoption in der Mitte steht für eine Einschätzung gleicher Schwere.

Ablauf: Eye-Tracking und ET SRI

Der Studienablauf wiederholte sich für jede Aufgabe und bestand aus drei aufeinanderfolgenden Teilen: Eye-Tracking, schriftliche Notizen und ET SRI. Die Aufgaben wurden einzeln und in randomisierter Reihenfolge präsentiert. Allen Teilnehmenden wurden der Studienablauf sowie die Aufgabenstellung anhand eines Aufgabenbeispiels erläutert.

Zur Erhebung der Eye-Tracking-Daten wurde mit einem monitorbasierten Eye-Tracker (Tobii Pro Fusion) gearbeitet, der binokulare Blickbewegungen mit Abtastraten von bis zu 120 Hz und einer durchschnittlichen Genauigkeit von 0.55° ($SD=0.4^\circ$) erfasste. Durch das stereoskopische System mit zwei Eye-Tracking-Kameras blieb eine hohe Qualität der Blickdaten auch bei natürlichen Kopfbewegungen stabil. Die Teilnehmenden wurden so positioniert, dass der mittlere Abstand zwischen dem Bildschirm (Fujitsu B24T-7 LED, Größe: 24 Zoll, Auflösung: 1920×1080 Pixel) und dem Probanden ca. 60 cm betrug. Vor jeder einzelnen Aufgabe wurde eine 9-Punkt-Kalibrierung durchgeführt, um optimale Eye-Tracking-Ergebnisse zu erzielen (Holmqvist & Andersson, 2017).

Nach der Diagnoseaufgabe am Eye-Tracker schrieben die Teilnehmenden stichpunktartig auf, welche Aufgabenmerkmale ihnen als leicht oder schwer aufgefallen sind. Die zu beurteilenden Aufgaben waren hierbei nicht mehr sichtbar. Die schriftlichen Notizen dienten den Teilnehmenden während der anschließenden ET SRI als zusätzliche Erinnerungshilfe und ermöglichten eine unmittelbare Prüfung, ob den Teilnehmenden beim Reflektieren der eigenen Blickbewegungen zuvor noch nicht genannte, weitere Aufgabenmerkmale auffielen.

Die Zeitspanne zwischen der Diagnoseaufgabe am Eye-Tracker und dem dazugehörigen ET SRI betrug zwischen 1 und 3 min. Während des Interviews beschrieben die Probanden anhand ihrer gezeigten Blickverläufe, was sie beim Diagnostizieren getan und gedacht haben. Für die Aufnahme der ET SRI wurde mit der Software OBS gearbeitet, die Bildschirminhalte inklusive Ton aufzeichnete, sodass für die spätere Auswertung die Videos der Blickverläufe mit den dazugehörigen Kommentaren der Probanden zur Verfügung standen.

Analyse der Daten

Zur Auswertung der Eye-Tracking-Daten wurde die Software Tobii Pro Lab genutzt. In jeder Aufgabe wurden spezifische AOIs definiert, die die jeweils variierten fachlichen und instruktionalen Aufgabenmerkmale umfassen. Im Bereich der fachlichen Aufgabenmerkmale wurden jeweils gleich große AOIs um die in den Aufgaben enthaltenen Bruchangaben festgelegt und die Fixationsdauer sowie die Anzahl an Fixationen in diesen Bereichen analysiert. Bei den

instruktionalen Aufgabenmerkmalen wurden unterschiedliche Maße betrachtet: Bei Aufgaben, die unnötige Informationen enthalten (s. Abschn. 10.2), wurde um die irrelevante Information ein AOI definiert, und die Fixationsdauer sowie die Anzahl an Fixationen wurden in diesem Bereich analysiert. Wurde in einer Aufgabe ein und dieselbe Information durch verschiedene Informationsquellen präsentiert (s. Abschn. 10.2) oder wurden die relevanten Informationen voneinander entfernt präsentiert (Split-Attention; s. Abschn. 10.2), wurden die doppelten bzw. voneinander entfernten Informationen durch jeweils ein AOI umfasst und analysiert, wie viele Transitionen zwischen den beiden AOIs stattgefunden haben. Die Anzahl an Fixationen und Fixationsdauern wurden mit dem Tobii-I-VT-Fixation-Filter (Velocity-Threshold Identification; z. B. Salvucci & Goldberg, 2000) ermittelt. Zur Bestimmung der Anzahl an Transitionen wurden die Videos der Blickbewegungen visuell inspiziert.

Für die Auswertung der ET-SRI-Daten wurde ein Mixed-Methods-Ansatz genutzt: Die ET SRI wurden transkribiert und mittels qualitativer Inhaltsanalyse nach Mayring (2015) deduktiv codiert. Die vergebenen Codes wurden anschließend in einen quantitativen Datensatz integriert, um potenzielle Unterschiede zwischen den Experimentalbedingungen varianzanalytisch zu untersuchen. Bei der qualitativen Inhaltsanalyse wurde die Analysetechnik der Strukturierung (vgl. Mayring, 2015) gewählt. Demnach wurde das folgende Kategoriensystem festgelegt und binär codiert (in Klammern): Schwierigkeitsgenerierende Aufgabenmerkmale (fachlich oder instruktional) können beim Diagnostizieren einer Aufgabe *identifiziert* (1), oder *nicht identifiziert* (0) werden. Es zeigte sich, dass manche Aufgabenmerkmale erst beim Reflektieren der eigenen Blickbewegungen während des SRI identifiziert wurden (subsummiert unter (0)). Identifizierte Aufgabenmerkmale können hinsichtlich der Schwierigkeit für Schüler*innen *korrekt evaluiert* (1) oder *inkorrekt evaluiert* (0) werden. Es zeigte sich, dass Aufgabenmerkmale teilweise zwar identifiziert, aber hinsichtlich ihrer Schwierigkeit für Schüler*innen *nicht weiter evaluiert* (0) wurden. Aufgabenmerkmale, die erst während des SRI evaluiert wurden, wurden subsummiert unter (0). Die Transkripte wurden von der Erstautorin und einer wissenschaftlichen Hilfskraft codiert mit hoher Interraterreliabilität (Cohens Kappa = .88).

10.5 Ergebnisse

Um zu untersuchen, welche schwierigkeitsgenerierenden Aufgabenmerkmale (fachlich vs. instruktional) beim Diagnostizieren identifiziert bzw. evaluiert wurden, und um einen potenziellen Einfluss von den zur Verfügung stehenden Informationen (informierte vs. uninformierte Diagnose) zu adressieren, wurden zwei ANOVAs mit Messwiederholung berechnet (getrennt für den Prozess *Identifizieren* und *Evaluiieren*). Hierbei wurde jeweils der Innersubjektfaktor *Aufgaben-*

merkmale (*fachlich/instruktional*) und der Zwischensubjektfaktor *Bedingung* (*informiert/uninformiert*) definiert.

Identifikation von fachlichen und instruktionalen Aufgabenmerkmalen

Die Ergebnisse zeigten einen signifikanten Unterschied mit hoher Effektstärke zwischen der Identifikation von fachlichen und instruktionalen Aufgabenmerkmalen ($F(1,29)=17,76$, $p<.001$, $\eta^2=.38$), der unabhängig von der Experimentalbedingung (informierte vs. uninformierte Diagnose) bestand. Bonferroni-korrigierte Post-hoc-Tests zeigten, dass sowohl in der Kontrollgruppe als auch in der Experimentalgruppe signifikant mehr fachliche als instruktionale Aufgabenmerkmale identifiziert wurden (Hypothese 1; Abb. 10.2). Instruktionale Aufgabenmerkmale wurden zu einem großen Teil nicht identifiziert. Bei der uninformierten Diagnose wurde durchschnittlich nur etwa die Hälfte ($M=.55$, $SD=.24$) der instruktionalen Aufgabenmerkmale identifiziert (Abb. 10.2). Darüber hinaus gab es einen signifikanten Unterschied mit hoher Effektstärke zwischen den Experimentalbedingungen ($F(1,29)=11,5$, $p=.002$, $\eta^2=.29$). Bonferroni-korrigierte Post-hoc-Tests zeigten, dass die Teilnehmenden der Experimentalgruppe sowohl mehr instruktionale Aufgabenmerkmale als auch mehr fachliche Aufgabenmerkmale identifizierten im Vergleich zur Kontrollgruppe (Hypothese 2a; Abb. 10.2). Zusätzliche Informationen, die zu bewertende Aufgabenmerkmale eingrenzen, hatten demnach dazu geführt, dass beim Diagnostizieren mehr schwierigkeitsgenerierende Aufgabenmerkmale identifiziert wurden.

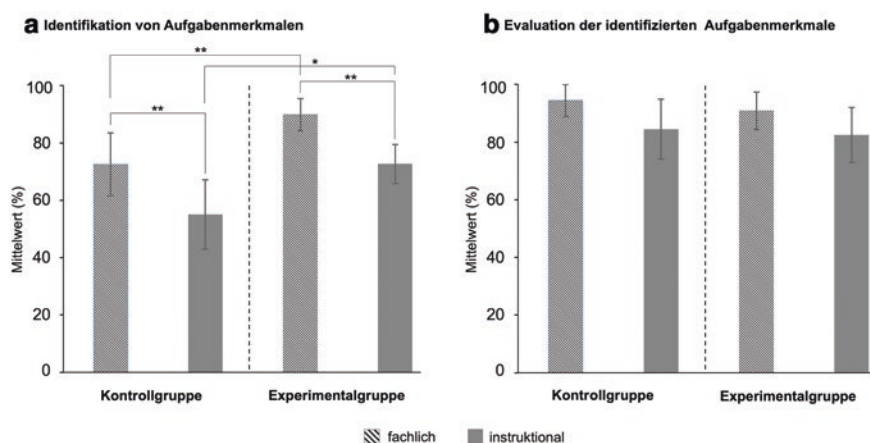


Abb. 10.2 Mittelwerte und Standardfehler für die Identifikation von Aufgabenmerkmalen (fachlich und instruktional) (a) und die Evaluation der identifizierten Aufgabenmerkmale (fachlich und instruktional) (b), jeweils getrennt für die Kontrollgruppe und Experimentalgruppe (* $p<.05$, ** $p<.01$).

Evaluation der identifizierten Aufgabenmerkmale hinsichtlich der Schwierigkeit für Schüler*innen

Identifizierte Aufgabenmerkmale konnten hinsichtlich der Schwierigkeit für Schüler*innen entweder korrekt oder inkorrekt evaluiert oder nicht weiterverarbeitet werden. Die Ergebnisse zeigten keine signifikanten Unterschiede zwischen der Evaluation von fachlichen und instruktionalen Aufgabenmerkmalen ($F(1,29)=3.5$, $p=.071$, $\eta^2=.11$). Ebenso konnten keine signifikanten Unterschiede zwischen den Studiengruppen festgestellt werden ($F(1,29)=.43$, $p=.519$, $\eta^2=.01$). Sowohl fachliche als auch instruktionale Aufgabenmerkmale wurden von beiden Gruppen überwiegend korrekt evaluiert (Abb. 10.2). Die Evaluation von Aufgabenmerkmalen hinsichtlich der Schwierigkeit für Schüler*innen blieb somit von der informierten Diagnose unbeeinflusst (Hypothese 2b).

Visuelle Aufmerksamkeit beim Diagnostizieren

Die Anzahl an Fixationen, die Fixationsdauer sowie die Anzahl an Transitionen wurden als Indikatoren herangezogen, um zu untersuchen, ob, wie oft und wie lange Aufmerksamkeit auf spezifische Aufgabenmerkmale gerichtet wurde. Um Gruppenunterschiede hinsichtlich der visuellen Aufmerksamkeit zu adressieren (Hypothese 3), wurden einseitige t -Tests berechnet (Tab. 10.2). Im fachlichen Bereich konnten keine signifikanten Gruppenunterschiede bzgl. der Anzahl an

Tab. 10.2 Übersicht der Eye-Tracking Maße

		Kontroll- gruppe ^a	Experimental- gruppe ^b	$t(29)$	p	Cohens d
		$M (SD)$	$M (SD)$			
<i>Fachliche AOIs</i>						
Bruchzahlen	Anzahl Fixationen	27,78 (12,77)	24,96 (8,11)	0,74	,234	0,26
	Fixationsdauer (s)	7,97 (4,39)	6,74 (2,12)	0,99	,168	0,36
<i>Instruktionale AOIs</i>						
Redundanz 1	Anzahl Fixationen	33,66 (7,21)	40,70 (6,58)	-2,84	,004	-1,02
	Fixationsdauer (s)	6,78 (1,44)	8,85 (2,26)	-3,03	,003	-1,09
Redundanz 2	Anzahl Transitionen	3,93 (2,15)	6,50 (4,89)	-1,87	,036	-0,68
Split-Attention	Anzahl Transitionen	13,93 (6,70)	16,31 (6,43)	-1,01	,161	-0,36

^a $n=15$, ^b $n=16$

Fixationen und der Fixationsdauer auf die in den Aufgaben enthaltenen Bereiche mit Bruchzahlen festgestellt werden. Im instruktionalen Bereich wurden unterschiedliche Eye-Tracking-Maße ausgewertet. Kommen in einer Aufgabe irrelevante Informationen vor (Redundanz 1), so zeigte die Experimentalgruppe durchschnittlich eine signifikant größere Anzahl an Fixationen und signifikant längere Fixationsdauern auf diese Bereiche (Tab. 10.2). Bei Aufgaben, in denen dieselbe Information durch verschiedene Informationsquellen präsentiert wird (Redundanz 2), zeigte die Experimentalgruppe durchschnittlich eine signifikant größere Anzahl an Transitionen zwischen diesen Informationsquellen. Die beschriebenen Unterschiede wiesen durchweg hohe Effektstärken auf (Tab. 10.2). Bei einer getrennten Präsentation der relevanten Informationsquellen (Split-Attention) konnten keine signifikanten Unterschiede zwischen Experimental- und Kontrollgruppe festgestellt werden.

10.6 Diskussion

Interpretation der Ergebnisse

In der Studie wurde untersucht, welche Aufgabenmerkmale (fachlich vs. instruktional) von angehenden Lehrkräften beim Diagnostizieren der Schwierigkeit von Bruchrechenaufgaben identifiziert und hinsichtlich der Schwierigkeit für Schüler*innen evaluiert werden. Dabei war der Einfluss zweier Urteilsmodi (informierte vs. uninformierte Diagnose) von besonderem Forschungsinteresse. Die Blickbewegungen beim Diagnostizieren wurden zusammen mit den zugehörigen ET SRI als Indikatoren für Identifikations- und Evaluationsprozesse herangezogen und zwischen den Experimentalbedingungen verglichen.

Die Ergebnisse zeigten, dass fachliche im Vergleich zu instruktionalen Aufgabenmerkmalen häufiger identifiziert, aber nicht häufiger korrekt evaluiert werden. Instruktionale Aufgabenmerkmale wurden häufig nicht als schwierigkeitsgenerierend identifiziert. Diese Ergebnisse stützen bestehende Forschungsergebnisse (Hellmann & Nückles, 2013; Schreiter et al., 2021), die belegen, dass instruktionale Aufgabenmerkmale von Lehrkräften nur unzureichend berücksichtigt werden. Zusätzlich liefert die vorliegende Studie Erkenntnisse über Informationsverarbeitungsprozesse, die diagnostischen Urteilen zugrunde liegen. Durch die Erhebung direkter Prozessindikatoren konnte gezeigt werden, dass schwierigkeitsgenerierende instruktionale Aufgabenmerkmale zu einem großen Teil gar nicht erst identifiziert werden, während bereits identifizierte Aufgabenmerkmale mehrheitlich korrekt evaluiert werden. Wenn Lehrkräfte instruktionale Aufgabenmerkmale nicht berücksichtigen, riskieren sie, dass es im Lernprozess der Schüler*innen zu einer kognitiven Überlastung kommt, was ein erfolgreiches Lernen verhindert (z. B. Sweller et al., 2011). Die Studienergebnisse weisen daher auf den Bedarf hin, die Identifikation von schwierigkeitsgenerierenden instruktionalen Aufgabenmerkmalen in der Lehrkräfteausbildung zu fördern.

In einem experimentellen Design wurde weiterhin untersucht, ob zusätzliche Informationen, die zu bewertende Aufgabenmerkmale eingrenzen, die Identifikations- und Evaluationsprozesse beeinflussen (informierte vs. uninformierte Diagnose). Erwartungskonform zeigte sich, dass bei der informierten Diagnose mehr schwierigkeitsgenerierende Aufgabenmerkmale – fachlich wie auch instruktional – identifiziert werden. Demnach kann die Fähigkeit zur Identifikation von schwierigkeitsgenerierenden Aufgabemerkmale bei angehenden Lehrkräften verbessert werden, indem sie über vorab eingegrenzte, relevante Kategorien von schwierigkeitsgenerierenden Aufgabemerkmale informiert werden. Damit liefert die Studie konkrete Impulse für die Aus- und Fortbildung von Lehrkräften. Die Evaluation von Aufgabenmerkmalen hinsichtlich der Schwierigkeit für Schüler*innen blieb von der informierten Diagnose unbeeinflusst. Dies könnte damit zusammenhängen, dass Evaluationsprozesse stärker vom Wissen darüber, was Schüler*innen leicht und/oder schwerfällt, abhängig sind. Im mittelmäßig ausfallenden bereichsspezifischen Vorwissenstest dieser Studie (s. Tab. 10.1) unterschieden sich beide Studiengruppen nicht signifikant voneinander. Um den aufgeführten Erklärungsansatz empirisch zu untermauern, müssten das Vorwissen experimentell variiert und potenzielle Effekte auf Evaluationsprozesse gezielt untersucht werden. Zu berücksichtigen ist ebenfalls, dass nur bereits identifizierte Aufgabenmerkmale hinsichtlich ihrer Schwierigkeit evaluiert werden können. Da zahlreiche schwierigkeitsgenerierende Aufgabenmerkmale nicht identifiziert wurden, bilden die Ergebnisse zu den Evaluationsprozessen nur einen Ausschnitt der Aufgabenmerkmale ab.

Spezifische Blickbewegungsmaße wurden als Indikatoren herangezogen, um zu untersuchen, ob, wie oft und wie lange Aufmerksamkeit auf fachliche und instruktionale Aufgabenmerkmalen gerichtet wurde. Dabei war die Untersuchung von Gruppenunterschieden (informierte vs. uninformierte Diagnose) von besonderem Forschungsinteresse. Einerseits zeigte sich erwartungskonform, dass informierte Personen eine größere Anzahl an Fixationen, längere Fixationsdauern und eine größere Anzahl an Transitionen auf bzw. zwischen den meisten instruktionalen Aufgabenmerkmalen aufweisen. Dies stützt die Annahme, dass bei einer informierten Diagnose die Aufmerksamkeit auf spezifische zu bewertende Aufgabenmerkmale gelenkt werden kann und infolgedessen mehr Aufgabenmerkmale identifiziert werden. Andererseits zeigten sich entgegen der Erwartung keine signifikanten Gruppenunterschiede hinsichtlich der analysierten Blickbewegungsmaße auf fachliche Aufgabenmerkmale. Dies könnte damit zusammenhängen, dass der Fokus der Aufmerksamkeit, auch bei der uninformierten Diagnose, möglicherweise ohnehin schon auf fachlichen Aufgabenmerkmalen liegt. Diese Annahme wird dadurch gestützt, dass bei beiden Gruppen mehr fachliche als instruktionale Aufgabenmerkmale identifiziert wurden. Diese Ergebnisse ergänzen bestehende Forschung, die Zusammenhänge zwischen den Informationen, die beim Diagnostizieren zur Verfügung stehen, und der Genauigkeit von diagnostischen Lehrerurteilen untersucht haben (vgl. Südkamp et al., 2012).

Limitationen und Implikationen für zukünftige Forschung

Eine Studienlimitation stellt der enge Fokus auf zwei ausgewählte Bereiche schwierigkeitsgenerierender Aufgabenmerkmale im Bereich der Bruchrechnung dar. Die Schwierigkeit von Bruchrechenaufgaben ist von weiteren Aufgabenmerkmalen beeinflusst, etwa der sprachlichen Komplexität (vgl. Leuders & Prediger, 2016), die hier nicht variiert wurden. Ebenfalls ist einschränkend zu bemerken, dass die Ergebnisse aufgrund der kleinen Stichprobe lediglich erste Indizien hinsichtlich diagnostischer Urteilsprozesse bei der Schwierigkeitseinschätzung von Bruchrechenaufgaben darstellen, die einer Replikation mit einer größeren Stichprobe erfordern.

Mit der experimentellen Manipulation der zur Verfügung stehenden Informationen konnte ein Faktor festgestellt werden, der die Identifikation von schwierigkeitsgenerierenden Aufgabenmerkmalen positiv beeinflusst. Dennoch werden, auch bei der informierten Diagnose, zahlreiche Aufgabenmerkmale nicht identifiziert und adäquat evaluiert. Dies wirft die Frage auf, wie die Identifikation/Evaluation von schwierigkeitsgenerierenden Aufgabenmerkmalen weiterhin verbessert werden kann. Das spezifische Wissen über schwierigkeitsgenerierende Aufgabenmerkmale wird als relevanter Aspekt für diagnostische Lehrerurteile angesehen (s. Abschn. 10.2). Zukünftige Studien sollten untersuchen, ob die Identifikation und Evaluation von fachlichen und instruktionalen Aufgabenmerkmalen durch die Vermittlung von spezifischem Wissen positiv beeinflusst werden kann.

10.7 Fazit

Die Studienergebnisse ermöglichen erste Einblicke in interne Informationsverarbeitungsprozesse, die diagnostischen Lehrerurteilen zugrunde liegen. Über die Erhebung von Prozessindikatoren mittels Eye-Tracking und ET SRI konnten Erkenntnisse gewonnen werden, welche schwierigkeitsgenerierenden Aufgabenmerkmale von angehenden Lehrkräften identifiziert und hinsichtlich der Schwierigkeit für Schüler*innen evaluiert werden. Hierbei konnte ein Förderbedarf hinsichtlich der Identifikation von instruktionalen Aufgabenmerkmalen herausgestellt werden. Durch die Unterscheidung zweier Urteilsmodi (informierte vs. uninformierte Diagnose) konnten zudem konkrete Impulse für die Lehrkräfteausbildung hervorgebracht werden: Die Fähigkeit zur Identifikation von schwierigkeitsgenerierenden Aufgabemerkmalen kann bei angehenden Lehrkräften verbessert werden, indem sie über vorab eingegrenzte, relevante Kategorien von schwierigkeitsgenerierenden Aufgabemerkmalen informiert werden.

Danksagung Die Studie wurde im Rahmen des Forschungs- und Nachwuchskollegs „Diagnostische Kompetenzen von Lehrkräften: Einflüsse, Struktur und Forderung (DiaKom)“ vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert.

Literatur

- Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology*, 20(3), 287–298. <https://doi.org/10.1002/acp.1245>.
- Bless, H., & Greifender, R. (2017). General framework of social cognitive processing. In R. Greifender, H. Bless & K. Fiedler (Hrsg.), *Social Cognition. How individuals construct social reality* (S. 16–37). Psychology Press.
- Eichmann, A., Narciss, S., Schnaubert, L., & Melis, E. (2012). Typische Fehler bei der Addition und Subtraktion von Brüchen – Ein Review zu empirischen Fehleranalysen. *Journal für Mathematik-Didaktik*, 33(1), 29–57. <https://doi.org/10.1007/s13138-011-0031-5>.
- Hellmann, K., & Nückles, M. (2013). Expert blind spot in pre-service and in-service mathematics teachers: task design moderates overestimation of novices' performance. In M. Knauff, M. Pauen, N. Sebanz & I. Wachsmuth (Hrsg.), *Proceedings of the 35th annual meeting of the cognitive science society* (S. 2518–2523).
- Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms, and measures* (2. Aufl.). CreateSpace.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4), 441–480. [https://doi.org/10.1016/0010-0285\(76\)90015-3](https://doi.org/10.1016/0010-0285(76)90015-3).
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31. https://doi.org/10.1207/s15326985EP3801_4.
- Karing, C., & Artelt, C. (2013). Genauigkeit von Lehrpersonenurteilen und Ansätze ihrer Förderung in der Aus- und Weiterbildung von Lehrkräften. *Beiträge Zur Lehrerbildung*, 31(2), 166–173.
- Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer-simulated problem solving. *The British Journal of Educational Psychology*, 75, 71–85. <https://doi.org/10.1348/000709904X19254>.
- Leuders, T., Dörfler, T., Leuders, J., & Philipp, K. (2018). Diagnostic competence of mathematics teachers: Unpacking a complex construct. In T. Leuders, K. Philipp, & J. Leuders (Hrsg.), *Diagnostic competence of mathematics teachers* (Vol. 3, S. 3–31). Springer International Publishing. https://doi.org/10.1007/978-3-319-66327-2_1.
- Leuders, T., & Prediger, S. (2016). *Flexibel differenzieren und fokussiert fördern im Mathematikunterricht. Sekundarstufe I + II*. Cornelsen.
- Loibl, K., Leuders, T., & Dörfler, T. (2020). A framework for explaining teachers' diagnostic Judgements by cognitive modeling (Diacom). *Teaching and Teacher Education*, 91, 1–10. <https://doi.org/10.1016/j.tate.2020.103059>.
- Lortie-Forgues, H., Tian, J., & Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult? *Developmental Review*, 38, 201–221. <https://doi.org/10.1016/j.dr.2015.07.008>.
- Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal*, 29(6), 861–878. <https://doi.org/10.1080/0141192032000137349>.
- Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. Aufl.). Beltz. (Beltz Pädagogik).
- Ostermann, A., Leuders, T., & Nückles, M. (2017). Improving the judgment of task difficulties: Prospective teachers' diagnostic competence in the area of functions and graphs. *Journal of Mathematics Teacher Education*, 21(6), 579–605. <https://doi.org/10.1007/s10857-017-9369-z>.
- Padberg, F. (1986). Über typische Schülerschwierigkeiten in der Bruchrechnung – Bestandsaufnahme und Konsequenzen. *Der Mathematikunterricht*, 32(3), 58–77.
- Padberg, F., & Wartha, S. (2017). *Didaktik der Bruchrechnung. Mathematik Primarstufe und Sekundarstufe I + II*. Springer Spektrum. <https://doi.org/10.1007/978-3-662-52969-0..>

- Prediger, S., & Zindel, C. (2017). Deepening prospective mathematics teachers' diagnostic judgments: Interplay of videos, focus questions and didactic categories. *European Journal of Science and Mathematics Education*, 5(3), 222–242.
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. T. Duchowski (Hrsg.), *Proceedings of the symposium on Eye tracking research & applications – ETRA '00* (S. 71–78). ACM Press. <https://doi.org/10.1145/355017.355028..>
- Schindler, M., & Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: Towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics*, 101(1), 123–139. <https://doi.org/10.1007/s10649-019-9878-z>.
- Schreiter, S., Vogel, M., Rehm, M., & Dörfler, T. (2021). Teachers' diagnostic judgment regarding the difficulty of fraction tasks: A reconstruction of perceived and processed task characteristics [Special Issue]. *RISTAL*, 4, 127–146. <https://doi.org/10.23770/RT1846>.
- Stickler, U., & Shi, L. (2017). Eyetracking methodology in SCMC: A tool for empowering learning and teaching. *ReCALL*, 29(2), 160–177. <https://doi.org/10.1017/S0958344017000040>.
- Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, 104(2), 147–200. <https://doi.org/10.1007/s10649-020-09948-1>.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>.
- Sweller, J., Ayres, P., & Kalyuga, S. (Hrsg.). (2011). *Cognitive Load Theory*. Springer New York. <https://doi.org/10.1007/978-1-4419-8126-4>.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>.
- Wyss, C., Rosenberger, K., & Bühner, W. (2021). Student Teachers' and Teacher Educators' Professional Vision: Findings from an Eye Tracking Study. *Educational Psychology Review*, 33(1), 91–107. <https://doi.org/10.1007/s10648-020-09535-z>.



Blickverhalten beim Lernen und Problemlösen mit Graphen – Ein Literaturüberblick bis 2020

11

Stefan Küchemann, Nils Cullmann, Sarah Kovac, Sebastian Becker, Pascal Klein, Kathrin Kennel, Stefan Ruzika und Jochen Kuhn

Inhaltsverzeichnis

11.1	Einleitung	178
11.2	Theorie	178
11.3	Forschungsfragen	180
11.4	Ergebnisse	181
11.5	Diskussion	188
	Literatur	190

S. Küchemann (✉) · N. Cullmann · S. Kovac · S. Becker · J. Kuhn
Didaktik der Physik, Fachbereich Physik, Technische Universität Kaiserslautern,
Kaiserslautern, Deutschland
E-Mail: s.kuechemann@lmu.de

S. Becker
E-Mail: sbeckerg@uni-koeln.de

J. Kuhn
E-Mail: jochen.kuhn@physik.uni-muenchen.de

P. Klein
Physik und ihre Didaktik, Fakultät für Physik, Georg-August-Universität Göttingen,
Göttingen, Deutschland
E-Mail: pascal.klein@uni-goettingen.de

K. Kennel · S. Ruzika
Wirtschafts- und Schulmathematik, Fachbereich Mathematik, Technische Universität
Kaiserslautern, Kaiserslautern, Deutschland
E-Mail: kkennel@mathematik.uni-kl.de

S. Ruzika
E-Mail: ruzika@mathematik.uni-kl.de

© Der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil
von Springer Nature 2022

P. Klein et al. (Hrsg.), *Eye-Tracking in der Mathematik- und
Naturwissenschaftsdidaktik*, https://doi.org/10.1007/978-3-662-63214-7_11

177

11.1 Einleitung

Graphen spielen eine zentrale Rolle für den mathematisch-naturwissenschaftlichen Unterricht (Glazer, 2011; Leinhardt et al., 1990). Beispielsweise sind sie essenziell für die Visualisierung von funktionalen Zusammenhängen und für die Darstellung von Daten. Verschiedene Autor*innen (Gal, 2002; Murray, 2003) sowie der National Council of Teachers of Mathematics (2000) in den USA heben die Dateninterpretation sowie die Arbeit mit grafischen Repräsentationen als Meilensteine in den Fähigkeiten des 21. Jahrhunderts hervor. Zusätzlich ist das Verständnis von Graphen eines der zentralen Elemente von disziplinübergreifenden Fähigkeiten und damit relevant für die Argumentation mit Internetinhalten („online reasoning“; Wineburg et al., 2018), das Verständnis von Daten (Cowie & Cooper, 2017) und das Problemlösen (Brand-Gruwel et al., 2009). Dieser Literaturüberblick konzentriert sich auf Blickverhalten bei Lern- und Problemlöseprozessen, da Graphen hierbei eine zentrale Informationsquelle darstellen und Lernende durch die enge Verknüpfung von Graphen mit algebraischen Darstellungen von Funktionen die Möglichkeit haben, Werte abzulesen, mathematische Prozeduren wie das Differenzieren und Integrieren visuell durchzuführen. Abgesehen davon spielen Graphen natürlich auch eine wichtige Rolle in anderen Bereichen der Lehr-/Lernforschung, wie beispielsweise in Lehreinheiten und Experimentierprozessen, jedoch bestanden zu diesen Themen bis zum Ende der Literaturrecherche keine Blickverhaltensuntersuchungen.

Zunächst stellt dieser Literaturüberblick in einem theoretischen Abschnitt die relevanten Definitionen und frühere Arbeiten im Bereich des Graphenverständnisses dar. Dieses schafft die Grundlage und Motivation von drei Forschungsfragen, die das Ziel dieses Literaturüberblicks bilden. Die methodische Vorgehensweise der Literaturrecherche wird im Anschluss an die Forschungsfragen dargelegt, und daraufhin werden die Ergebnisse geordnet nach den Forschungsfragen als Hauptteil dieser Arbeit aufgeführt. Das Kapitel endet mit einer abschließenden Diskussion der Ergebnisse.

11.2 Theorie

Im Allgemeinen umfasst der Begriff „Repräsentation“ neben den „Organisationsformen individuellen Wissens“ auch „die Prozesse der Veränderung dieses Wissens“, die „Ableitung neuen Wissens mittels bewusster oder unbewusster Schlussfolgerungsprozesse“ und die „Generierung von Handlungsplänen“ (Denken, Handlungsplanung, analoges Schließen, logisches Schließen) (Dorsch, 2021). In diesem Kapitel wird der Begriff „Repräsentation“ synonym für die erstgenannte Definition der „Organisationsformen individuellen Wissens“ und die Aufteilung der Repräsentationstypen nach Lemke (1998) in symbolhaft-mathematische, verbal-textbasierte, visuell-grafische und aktional-operationale Repräsentationen verwendet. Gemäß der Kategorisierung von Lemke (1998) gehören Diagramme und Graphen zu der Kategorie der visuell-grafischen Repräsentationen. Diagramme

sind „grafische Darstellungen von Größenverhältnissen, bzw. Zahlenwerten in anschaulicher leicht überblickbarer Form“ (Duden, 2021). Die in dieser Arbeit behandelten Graphen sind Diagramme, mit denen die funktionale Abhängigkeit $f(x)$ zweier Werte skizziert wird, also eine Darstellung der Menge aller Tupel $(x, f(x))$. Je nach Kontext können diese Tupel z. B. einer explizit gegebenen Funktion entstammen oder bspw. Messdaten zu einer gegebenen Eingangsgröße sein oder etwa Simulationsergebnisse zu einem gegebenen Parameter repräsentieren.

Frühere Arbeiten haben gezeigt, dass Studierende sowie Schüler*innen spezifische Schwierigkeiten bei der Arbeit mit Graphen haben (Glazer, 2011; Leinhardt et al., 1990; McDermott et al., 1987). Dazu zählen beispielsweise die Verwechslung von Steigung und Höhe in einem Graphen, die Interpretation von Höhenänderungen und Änderungen der Steigung ebenso wie die Fläche unter einer Kurve (McDermott et al., 1987). Darüber hinaus beobachtete Beichner (1993) auch eine Verwechslung von abhängigen Variablen und wies auf die Komplexität hin, wenn die mathematischen Verfahren in einem physikalischen Kontext angewendet werden. Um diese Schwierigkeiten zu untersuchen, entwickelte er den Test of Understanding Graphs in Kinematics (TUG-K) (Beichner, 1994).

Neben der Vielzahl an Vorteilen dieses Tests hat der TUG-K den Nachteil, dass er nicht zwischen dem konzeptuellen Verständnis und der Repräsentationskompetenz, also der Fähigkeit, mit Graphen zu arbeiten, sie zu interpretieren oder diese in andere Darstellungen überzuführen, differenziert. Hier bietet die Analyse des Blickverhaltens die Möglichkeit, Expert*innen und Noviz*innen nicht nur anhand des Konzeptwissens, sondern zusätzlich auf der Basis der visuellen Strategie bei der Arbeit mit Repräsentationen zu unterscheiden (Küchemann et al., 2020a). Beispielsweise konnten Gegenfurtner et al. (2011) im Rahmen einer Metaanalyse von Eye-Tracking-Studien zeigen, dass Expert*innen schneller einen relevanten Bereich in einer Repräsentation identifizieren als Noviz*innen. Ebenso weist Rau (2017) ohne Belege durch Blickdaten darauf hin, dass Expert*innen bei der Arbeit mit visuellen Repräsentationen perzeptuelle Prozesse optimiert und damit eine gewisse visuelle Gewandtheit erreicht haben, die es ihnen ermöglicht, Informationen zu Chunks zusammenzufassen und mit wenig Aufwand visuelle Repräsentationen zu verarbeiten.

Im Zusammenhang mit verschiedenen multiplen Repräsentationen im physikalischen Kontext konnten Kozhevnikov et al. (2007) in einer Eye-Tracking-Studie herausstellen, dass Studierende mit hohem räumlichem Vorstellungsvermögen eine deutlich gezieltere visuelle Lösungsstrategie anwenden als Studierende mit geringem räumlichem Vorstellungsvermögen. Beim Lösen von repräsentationsbasierten Fragen in dem sog. Force Concept Inventory (FCI), einem validierten Konzepttest zur Newton'schen Mechanik (Hestenes et al., 1992), sowie weiteren diagrammbasierten Fragen wurde in der Eye-Tracking-Analyse allerdings auch deutlich, dass die visuellen Strategien bei einer korrekten Lösung stark repräsentationsabhängig sind (Kozhevnikov et al., 2007). Eine erfolgreiche Lösungsstrategie bei einer repräsentationsbasierten Frage kann nicht unmittelbar

auf eine Frage mit einer anderen Repräsentation übertragen werden. Entsprechend ist auch das Blickverhalten bei Graphen stark von der Darstellungsform abhängig und nicht übertragbar.

11.3 Forschungsfragen

Der o. g. Überblicksartikel zu Eye-Tracking von Gegenfurtner et al. (2011) ist nicht spezifisch für Graphen und erlaubt daher keine Aussage über visuelle Strategien, die es ermöglichen würde, ineffektive Verarbeitungen oder Schwächen beim Lernen und Problemlösen mit Graphen zu identifizieren.

Frühere Überblicksartikel adressieren die Einsatzmöglichkeiten von Eye-Tracking in der Lehr-/Lernforschung (Lai et al., 2013), untersuchen den Bezug zu kognitiven Prozessen beim multimedialen Lernen (Alemdag & Cagiltay, 2018), stellen zusammen, wie Eye-Tracking in der Mathematikdidaktik eingesetzt wird (Strohmaier et al., 2020), oder schaffen einen Überblick über die Veröffentlichungen von Eye-Tracking im Rahmen der Psychology-of-Mathematics-Education-Konferenz (Lilienthal & Schindler, 2019). Weiterhin berichtete Tippet (2016), dass 7.5 % von insgesamt 79 Studien zwischen 2000 und 2014 im Zusammenhang mit Diagrammen Eye Tracking nutzen, um Lernen *von* oder *mit* Diagrammen zu untersuchen. Die Autorin weist ebenfalls auf die steigende Popularität von Eye-Tracking-Untersuchungen im Kontext von Diagrammen aufgrund verbesserter Technologien in diesem Zeitraum hin.

Dieses Kapitel fügt sich in die Lücke der visuellen Verarbeitung einer spezifischen Repräsentation, indem es einen Überblick über Eye-Tracking-Studien mit Graphen zusammenstellt. Aufgrund der erwähnten Abhängigkeit des Blickverhaltens von der Darstellungsform eines Graphen und des Kontexts fokussiert Forschungsfrage 1 auf diese Aspekte:

Forschungsfrage 1: Welchen Einfluss hat der Graphentyp und der Kontext auf das Blickverhalten?

Da weitere Repräsentationen ebenso verarbeitet werden müssen, stellt sich die Frage, ob die visuelle Verarbeitung von Graphen auch durch simultan angebotene Repräsentationen beeinflusst wird:

Forschungsfrage 2: Welchen Einfluss haben weitere simultan angebotene Repräsentationen auf das Blickverhalten?

Neben dem Einfluss der Expertise (Gegenfurtner et al., 2011) adressiert Forschungsfrage 3 den Einfluss weiterer Personencharakteristika auf das Blickverhalten bei der Verarbeitung von Graphen.

Forschungsfrage 3: Welche weiteren Personencharakteristika neben der Expertise beeinflussen das Blickverhalten?

Literaturrecherche

Das Ziel der Literaturrecherche war die Suche nach Artikeln, die das Blickverhalten bei Graphen im Kontext von Lernen und Problemlösen im MINT-Bereich untersuchen. Um relevante Artikel zu identifizieren, die in Zeitschriften und Büchern veröffentlicht wurden, durchsuchten wir Forschungsdatenbanken wie ERIC, SCOPUS, Pedocs, SpringerLink und Google Scholar. Dort haben wir die Schlüsselwörter „eye tracking“, „graph“, „diagram“ sowie „function“, „viewing behavior“ und „visual attention“ in verschiedenen Kombinationen gekoppelt. Die Stichwortsuche wurde auf Titel und Abstract beschränkt. Da nicht jede Datenbank die Möglichkeit der Paraphrasierung oder Trunkierung bot, wurden Schlüsselwörter teilweise durch ihre adjektivische oder adverbiale Form ersetzt, um alternative Formulierungen in den Suchbereich einzubeziehen. Zusätzlich wurden die Schlagwörter ins Deutsche übersetzt und die Suchprozedur wiederholt, da es eine Vielzahl von Artikeln gibt, die den Kontext von Darstellungen in der deutschsprachigen Bildungsforschung untersuchen. Die Suche wurde im Mai und Juni 2020 durchgeführt. Einsendeschluss für relevante Publikationen war der 1. April 2020. Es gab keine weiteren formalen Einschränkungen, weder nach Jahr noch nach Publikationstyp. In allen Arbeiten handelt es sich um empirische Studien. Im Rahmen der Suche in den o. g. Datenbanken wurden 13 Artikel identifiziert, die das Blickverhalten beim Lernen oder Problemlösen mit Graphen untersucht haben.

Zusätzlich zu der Suche in o. g. Datenbanken wurde mithilfe der Schneeballsuche in den identifizierten Artikeln gesucht. Diese Suche ergab weitere 14 Artikel, sodass dieser Überblicksartikel insgesamt 27 Artikel berücksichtigt (Tab. 11.1).

11.4 Ergebnisse

Zeitraum der Veröffentlichungen

Den ersten Artikel zum Thema Eye-Tracking beim Lernen oder Problemlösen mit Graphen konnten wir im Jahr 2003 identifizieren (Abb. 11.1). Mit Ausnahme von jeweils einem Artikel in den Jahren 2006 und 2007 konnten in der Literaturrecherche keine Artikel zu dem Thema im Zeitraum von 2003 und 2011 gefunden werden. Ab 2011 gibt es einen leichten Zuwachs mit einem Maximum im Jahr 2018. Anschließend folgte ein Abfall. Im letzten Jahr ist zu berücksichtigen, dass nur relevante Publikationen in den ersten drei Monaten berücksichtigt wurden.

Im Vergleich zu anderen aktuellen, domänenspezifischen Reviews in der Mathematik stellt das Histogramm einen sehr ähnlichen Verlauf dar. Strohmaier et al. (2020) stellten fest, dass nur vereinzelt Eye-Tracking-Studien vor 2006 in der Mathematikdidaktik veröffentlicht wurden und es danach einen deutlichen Zuwachs bis zum Jahr 2018 gab. Die Autor*innen führten diesen Zuwachs auf

Tab. 11.1 Titel, Autor*innen, Jahr der Veröffentlichung und Schlüsselwörter bei der Suche der identifizierten Artikel

Autor*innen	Jahr	Suche	Graphtyp &-stil
Andrá, Lindström, Arzarello, Holmqvist, Robutti, & Sabena	2015	„eye tracking“ und „graphs“	Linear; Streu-, Linien-, Säulendiagramm
Atkins, & McNeal	2018	Schneeballsuche	Linear; Linien- und Säulendiagramm
Brückner, Zlatkin-Troitschanskaia, Küchemann, Klein, & Kuhn	2020	„visual attention“ und „graphs“	Linear; Liniendiagramm
Goldberg, & Helfman	2011	„eye tracking“ und „diagram“ und viewing behavior	Linear und Radial; Streu-, Linien-, Flächen-, Säulendiagramm
Harsh, Campillo, Murray, Myers, Nguyen, & Maltese	2019	„eye tracking“ und „graphs“	Linear; Linien-, Säulendiagramm
Ho, Tsai, Wang, & Tsai	2013	„eye tracking“ und „diagram“ und „viewing behavior“	Linear; Liniendiagramm
Kekule	2014	Schneeballsuche	Linear; Liniendiagramm
Kim, Lombardino, Cowles, & Altmann	2014	„eye tracking“ und „line graph“	Linear; Linien-, Säulendiagramm
Klein, Lichtenberger, Küchemann, Becker, Kekule, Viiri, ... & Kuhn	2020	Schneeballsuche	Linear; Liniendiagramm
Klein, Küchemann, Brückner, Zlatkin-Troitschanskaia, & Kuhn	2019	„eye tracking“ und „graphs“	Linear; Liniendiagramm
Kozhevnikov, Motes, & Hegarty	2007	Schneeballsuche	Linear; Liniendiagramm
Küchemann, Klein, Becker, Kumari, & Kuhn	2020a	Schneeballsuche	Linear; Liniendiagramm
Madsen, Larson, Loschky, & Rebello	2012	Schneeballsuche	Linear; Liniendiagramm
Madsen, Rouinfar, Larson, Loschky, & Rebello	2013	Schneeballsuche	Linear; Liniendiagramm
Matzen, Haass, Divis, Wang, & Wilson	2017	Schneeballsuche	Linear; Linien-, Säulendiagramm
Okan, Galesic, & Garcia-Retamero	2016	„eye tracking“ und „graphs“	Linear; Linien-, Säulendiagramm
Peebles, & Cheng	2003	Schneeballsuche	Linear; Streu-, Liniendiagramm
Polatsek, Waldner, Viola, Kapec, & Benesova	2018	„eye tracking“ und „line chart“ und „representation“ und „visual attention“	Linear; Streu-, Linien-, Flächen-, Säulendiagramm
Rouinfar, Agra, Larson, Rebello, & Loschky	2014	Schneeballsuche	Linear; Liniendiagramm
Strobel, Lindner, Saß, & Köller	2018	Schneeballsuche	Linear; Säulendiagramm
Strobel, Grund, & Lindner	2019	Schneeballsuche	Linear; Säulendiagramm
Strobel	2018	„eye tracking“ und „Graphen“	Linear; Linien-, Säulendiagramm
Strobel, Sass, Lindner, & Köller	2016	„eye tracking“	Linear; Linien-, Säulendiagramm
Susac, Bubic, Kazotti, Planinic, & Palmovic	2018	„eye tracking“ und „line graph“	Linear; Liniendiagramm
Tai, Loehr, & Brigham	2006	Schneeballsuche	Linear; Liniendiagramm
Toker, Conati, Steichen, & Carenini	2013	Schneeballsuche	Linear, Radial; Linien-, Säulendiagramm
Vila, & Gomez	2016	Schneeballsuche	Linear; Säulendiagramm

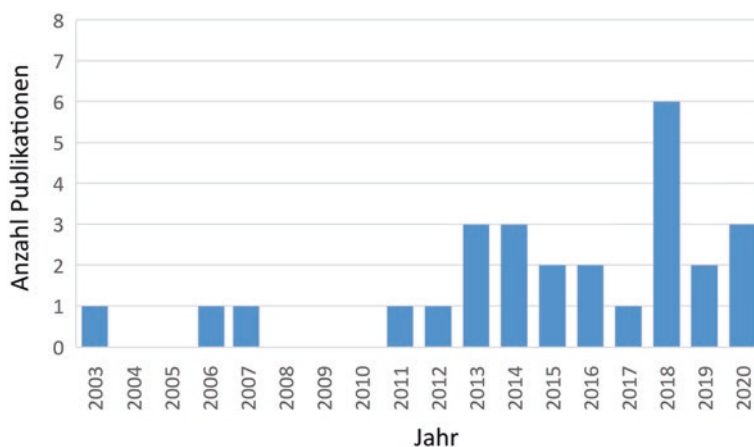


Abb. 11.1 Histogramm der Anzahl der Publikationen im Zusammenhang mit Graphen in dem Zeitraum von 2003 bis 2020

technische Fortschritte bei Eye-Tracking-Geräten und die damit verbundene einfachere Umsetzbarkeit von Studien zurück. Parallel dazu beobachteten Lilienthal und Schindler (2019), dass ab 2012 Begriffe, die im engen Zusammenhang mit Eye-Tracking stehen, zunehmend in den Tagungsbandbeiträgen der International Group for the Psychology of Mathematics Education auftraten, was auf eine wachsende Relevanz von Eye-Tracking-Studien in dieser bildungswissenschaftlichen Disziplin hindeutet.

Einfluss des Graphentyps und des Kontexts (Forschungsfrage 1)

Bei Beantwortung der Forschungsfragen 1 bis 3 ist zu beachten, dass die fachliche Expertise einen Einfluss auf das Blickverhalten hat und damit auch teilweise in Wechselwirkung mit den in den Forschungsfragen genannten Einflüssen steht. Sofern dies der Fall ist und berichtet wurde, wird an entsprechender Stelle im Text darauf hingewiesen. Neben dem rein mathematischen Kontext (Andrá et al., 2015) wurde das Blickverhalten bei Graphen bislang im Bereich von Klimadaten (Atkins und McNeal, 2018; Harsch et al., 2019), der Kinematik (Kekule, 2014; Madsen et al., 2012; Madsen et al., 2013; Rouinfar et al., 2014; Klein et al., 2020; Küchemann et al., 2020a), Gesundheitsdaten (Okan et al., 2016) und Wirtschaftswissenschaften (Vila & Gomez, 2016) untersucht. Zusätzlich vergleichen einige Studien den Kontext (Susac et al., 2018; Klein et al., 2019; Brückner et al., 2020) und andere Studien beschränken sich auf keinen spezifischen Kontext (Matzen et al., 2017; Polatsek et al., 2018; Strobel et al., 2016; Strobel, 2018; Strobel et al., 2018; Strobel et al., 2019). In der Mathematik stellen Graphen eine Repräsentationsform dar, die erlaubt, die Abhängigkeit einer Variablen von einer unabhängigen Variablen zu visualisieren und damit funktionale Abhängigkeiten

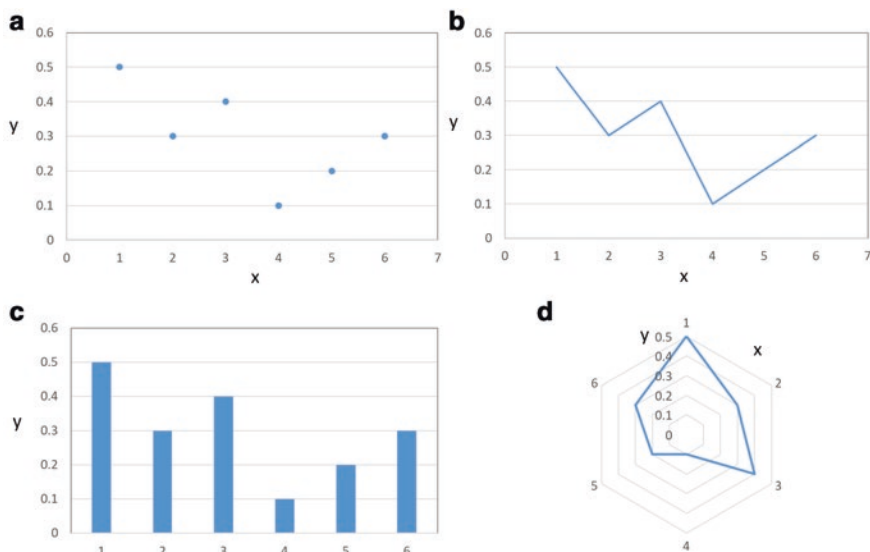


Abb. 11.2 Künstlich erzeugter Datensatz, dargestellt in unterschiedlichen Typen (**a–c** linear, **d** radial) und unterschiedlichen Stilen (**a** Streudiagramm; **b**, **d** Liniendiagramm; **c** Säulendiagramm), die in den hier identifizierten Artikeln untersucht wurden. Die Stile in **a** und **c** sind ebenfalls als radialer Graph darstellbar

visuell zu unterscheiden. Durch ihre hohe inhaltliche Nähe zu Funktionen bilden sie somit eine wertvolle Ergänzung zu formelbasierten Darstellungen. Durch diese Möglichkeit spielen sie eine zentrale Rolle in der Konstruktion von mentalen Modellen (Schnotz & Bannert, 2003). Im Kontrast zu Formeln oder einem Text sind nach Roth (2003) Graphen nicht selbstevident, und es ist notwendig, dass sie in dem eingebetteten Kontext interpretiert werden. Ebenso ist ein Graph eine Projektion aus der Perspektive des Wissenschaftlers oder der Wissenschaftlerin sowie der Lehrkraft, die ihn erzeugt hat, und verlangt neben den Konventionen somit auch ein Verständnis des Kontexts, in dem der Graph entstanden ist (Roth, 2003).

Bei der Auftragung von Abhängigkeiten sind verschiedene Graphentypen und -stile denkbar. Gemäß der Einteilung von Goldberg und Helfman (2011)¹ kann zwischen linearen und radialen Graphentypen unterschieden werden.

In den hier betrachteten Artikeln werden die Repräsentationen in Abb. 11.2 sowohl als Diagramme („chart“, „diagram“) als auch als Graphen („graph“)

¹Neben den untersuchten Formen von Goldberg und Helfman (2011) sind natürlich noch weitere Darstellungsformen von Diagrammen denkbar. Dennoch beziehen wir uns hier auf diese eingeschränkte Auswahl, da sie in der Eye-Tracking Community untersucht wurden.

bezeichnet. Die in Abb. 11.2 präsentierten Stile bilden eine Auswahl der analysierten Graphen von Goldberg und Helfman (2011). Neben den dort dargestellten Typen und Stilen gibt es noch weitere Darstellungsformen, die aber mit Ausnahme der Arbeit von Goldberg und Helfman (2011) nicht in den aufgeführten Artikeln untersucht wurden. In den 27 in diesem Literaturüberblick identifizierten Artikeln untersuchten alle 27 lineare Graphendarstellungstypen, und nur zwei Artikel betrachteten auch den radialen Darstellungstyp. Zudem verwendeten 24 Artikel Liniendiagramme, ungefähr die Hälfte der Artikel ($N=14$) untersuchten Säulendiagramme, und vier Arbeiten betrachteten auch Streudiagramme. Letzterer Darstellungsstil wurde jedoch nie ausschließlich betrachtet.

Die erhöhte Effizienz in der Verarbeitung von Information von Graphen ist allerdings spezifisch für die Darstellungsform eines Graphen. In einer Studie im Within-Subject Design mit 32 Expert*innen im Ablesen von Informationen aus Graphen im wirtschaftswissenschaftlichen Kontext stellten Goldberg und Helfman (2011) fest, dass die schnelle Informationsextraktion auf lineare Darstellungen von Graphen (das sind Graphen mit zwei rechtwinklig zueinander ausgerichteten, geradlinigen Koordinatenachsen) beschränkt ist. Die Expert*innen brauchten signifikant länger, relevante Informationen in radialen Darstellungen von Graphen zu identifizieren (gemessen anhand der Zeit bis zur ersten Fixation), zu extrahieren (gemessen anhand der mittleren Fixationsdauer) und zu verarbeiten, als in linearen Graphendarstellungen. Damit konnten sie zeigen, dass die visuelle Expertise und die damit verbundene effiziente Verarbeitung nicht vom domänenspezifischen Vorwissen herrührt. Die Autor*innen vermuteten, dass die Expert*innen geübt in der Verarbeitung von Informationen in linear präsentierten Graphen waren.

Der Zusammenhang von Expertise und dem Blickverhalten in verschiedenen Kontexten wurde erstmalig von Tai et al. (2006) untersucht. Die Autor*innen fanden in einer qualitativen Analyse des Blickverhaltens von sechs Lehramtsstudierenden bei standardisierten Aufgaben in den Kontexten Biologie, Chemie und Physik, dass sich das Blickverhalten hinsichtlich der Fixationsdauer (*fixation duration*) und Anzahl der Transitionen je nach domänenspezifischer Expertise unterscheidet. Studierende mit erhöhtem domänenspezifischem Vorwissen führen kaum visuelle Transitionen zwischen dem abgebildeten Graphen und den Antwortmöglichkeiten durch, diese treten jedoch in erhöhter Anzahl bei Studierenden mit geringerem Vorwissen auf (Tai et al., 2006).

Im Zusammenhang mit der Expertise bei verschiedenen Kontexten konnten Susac et al. (2018) in einer Analyse zeigen, dass 23 angehende Physiklehrkräfte und 22 Psychologiestudierende Probleme damit haben, eine korrekt angewendete Strategie von einem bekannten Kontext (bei Physikern ist das ein physikalischer Kontext, bei Psychologiestudierenden ein Finanzkontext) auf einen fremden Kontext zu übertragen. Sie stellten bei der Untersuchung des Blickverhaltens beim Problemlösen mit Graphen fest, dass korrekte Lösungen (hauptsächlich von Physikstudierenden) mit einer erhöhten Aufmerksamkeit (*dwell time*) auf dem Graphenbereich einhergingen. In einer Replikationsstudie von Klein et al. (2019), bei der dieselben Aufgaben bei 29 Physikstudierenden und 40 Wirtschaftswissenschaftsstudierenden getestet wurden, stellte sich heraus, dass beide

Kohorten trotz signifikant unterschiedlicher Lösungswahrscheinlichkeit keinen Unterschied in der Verweildauer auf dem Graphenbereich hatten. Stattdessen schauten Physikstudierende signifikant länger (*dwell time*) auf die Achsenwerte. Zusätzlich blickten Physikstudierende auch länger auf Bereiche entlang des Verlaufs des Graphen, wenn die Aufgabe darin bestand, die Steigung eines Graphen zu bestimmen (Klein et al., 2019). Dieses Verhalten wurde auch in den Sakkadenrichtungen deutlich, welche bei Physikstudierenden ausgeprägter entlang des Graphen deuteten. Diese Studie wurde zu Semesterbeginn durchgeführt.

Als Ergänzung dieser Ergebnisse führte Brückner et al. (2020) mit einer Teilgruppe derselben Kohorte dieselbe Studie wie Klein et al. (2019) erneut zum Semesterende durch. Die Autor*innen konnten nachweisen, dass die o. g. Strategien bei der grafischen Bestimmung der Steigung eines Graphen auch über den Verlauf eines Semesters stabil sind (Brückner et al., 2020).

Einfluss weiterer simultan angebotener Repräsentationen (Forschungsfrage 2)

Andrá et al. (2015) untersuchten, wie 46 Universitätsstudierende mit unterschiedlichen Mathematikvorkenntnissen je zwei aus insgesamt drei unterschiedlichen Repräsentationsformen (Graph, Text, Formel) in 43 Multiple-Choice-Fragen einander zuordnen. Die Autor*innen konnten feststellen, dass eine Zuordnung von textbasierten Beschreibungen zu einem vorgegebenen Graphen mit der höchsten Lösungswahrscheinlichkeit bei den Studierenden belegt war. Dabei stellte sich heraus, dass die Aufmerksamkeitsdauer (*dwell time*) auf einem Graphen in der Aufgabenstellung deutlich höher als auf einer Formel oder einem Text war. Diese erhöhte Aufmerksamkeit manifestierte sich in einer höheren Anzahl Fixationen auf den Graphen, wohingegen eine signifikant erhöhte mittlere Fixationsdauer bei Formeln als Ausgangsrepräsentation gemessen wurde, was auf eine erhöhte Komplexität in der Verarbeitung von Formeln gegenüber Graphen hindeutete. Zudem fanden die Autor*innen, dass bei Graphen (als Ausgangsrepräsentation) und Texten (als Zielrepräsentation) deutlich mehr Transitionen innerhalb von Graphen (70–75 % aller Transitionen) als der paarweise Vergleich der beiden Repräsentationsformen (10 % aller Transitionen) stattfinden. Damit ließ sich schlussfolgern, dass für Studierende die Informationsverarbeitung im Graphenbereich wichtiger als die Integration der beiden Repräsentationsformen ist (Andrá et al., 2015).

Bei einem Lernprozess stellten Ho et al. (2014) fest, dass 13 Studierende unterschiedlicher Fachrichtungen mehr Zeit mit dem Lesen des Texts verbrachten. Bei dem Stimulus handelte es sich um eine Webseite zur Darstellung des Wirkzusammenhangs des Klimawandels und des Treibhauseffekts mit zwei Graphen und einem erklärenden Text. Bei einer Unterteilung in Studierende mit hohem und niedrigem Vorwissen beobachteten die Autor*innen, dass Studierende mit hohem Vorwissen mehr Zeit als Studierende mit geringem Vorwissen auf

den Graphen verbrachten und Transitionen zwischen den beiden Graphen durchführten.

Aus dem Vergleich dieser beiden Arbeiten lässt sich schließen, dass während des Problemlösens im Vergleich zum Lernprozess eine deutlich intensivere Auseinandersetzung mit Graphen erfolgt. Ein möglicher Grund für diese Diskrepanz ist, dass die Inhalte bei dem hier untersuchten Lernprozess inhaltsgleich waren und daher eine Auseinandersetzung mit den Graphen im Gegensatz zum Problemlöseprozess nicht zwingend erfolgen muss.

Abgesehen von der Verknüpfung von Graphen mit nebenstehendem Text, haben Vila und Gomez (2016) mit 30 Studierenden untersucht, welchen Einfluss redundante Zahlenwerte in einem Säulendiagramm haben. Sie beobachteten, dass redundante Zahlen keinen Einfluss auf die Lösungswahrscheinlichkeit haben, aber die Bearbeitungszeit bei korrekter Antwort durch eine erhöhte Aufmerksamkeit auf den Zahlen verlängerte.

Einfluss weiterer Personencharakteristika auf das Blickverhalten (Forschungsfrage 3)

Neben dem Einfluss der Expertise im Umgang mit Graphen auf das Blickverhalten haben verschiedene Autor*innen auch den Einfluss weiterer Personenmerkmale untersucht. Die Verweildauer auf den Antwortoptionen konnte Klein et al. (2020) dadurch noch spezifizieren. In einer Eye-Tracking-Analyse des TUG-K mit 115 Oberstufenschüler*innen stellten die Autor*innen fest, dass die Verweildauer (*absolute dwell time*) auf den Optionen sowohl von der Expertise als auch von der Popularität (Häufigkeit, mit der ein Distraktor gewählt wurde) abhängt. Schüler*innen mit einer hohen Expertise verbrachten demnach sogar mehr Zeit auf der richtigen Antwort, aber verhältnismäßig wenig Zeit auf allen anderen Optionen. Im Vergleich dazu verbrachten Studierende, die nur wenige Items korrekt lösten, vergleichbar viel Zeit auf allen Antwortmöglichkeiten. Darüber hinaus stellten die Autor*innen fest, dass neben dem Expertiselevel auch die Selbstwahrnehmung im Zusammenhang mit dem Blickverhalten steht. Unabhängig von der Korrektheit der Antwort verbrachten Studierende, die sich unsicher mit ihrer Antwort fühlten, signifikant mehr Zeit auf den Antwortoptionen und der Frage als Studierende, die sich sicher fühlten (Klein et al., 2020). In dieser Studie wurde die Expertise anhand der Gesamtpunktzahl im TUG-K eingeordnet. Für verschiedene visuelle Repräsentationen wurde ebenfalls beobachtet, dass Studierende beim Problemlösen signifikant mehr visuelle Aufmerksamkeit auf relevante Bereiche richten, wenn sie sich unsicher mit ihrer Antwort fühlen (Küchemann et al., 2020b).

Abgesehen davon wurde festgestellt, dass die Lesekompetenz ebenfalls mit dem Ablesen von Informationen aus Graphen in Zusammenhang steht. Kim et al. (2014) untersuchten das Blickverhalten von 16 Studierenden mit einer Lese- und Rechtschreibstörung und von 23 Studierenden mit durchschnittlicher Lesekompetenz beim Problemlösen mit verschiedenen Graphentypen. Die

Autor*innen fanden heraus, dass Studierende mit einer Leseschwäche nicht nur signifikant länger auf die Aufgabenstellung und die Antworten, sondern auch auf den Graphenbereich schauten. Die Betrachtungsdauer der beiden Achsen und der Legende zeigte dagegen keinen signifikanten Gruppenunterschied.

Den Einfluss der verbalen Arbeitsgedächtnisspanne auf das Blickverhalten wurde u. a. von Toker et al. (2013) mit 35 Universitätsstudierenden bei radialen und Säulendiagrammen untersucht. Neben der verbalen Arbeitsgedächtnisspanne untersuchten die Autor*innen auch noch einen Zusammenhang mit dem räumlichen Vorstellungsvermögen, der Aufgabenschwierigkeit, dem Graphenverständnis (getrennt für die beiden Graphentypen) und den wahrnehmenden Fähigkeiten (*perceptual skills*). Sie fanden heraus, dass die Studierenden mit steigender Aufgabenschwierigkeit mehr Transitionen zu der Legende, aber weniger Transitionen zu den Achsenbeschriftungen durchführten. Die Autor*innen vermuteten, dass Studierende bei komplexeren Aufgaben die kognitive Belastung zu einem Vergessen der Legendeninformation führte und daher mehr Transitionen dorthin durchgeführt werden müssen (Toker et al., 2013). Das Vergessen von Informationen wurde auch bereits früher von Peebles und Cheng (2003) erfolgreich in der Blickverhaltensmodellierung beim Problemlösen von Graphenaufgaben berücksichtigt. Die Transitionen zurück zur Variablendefinition werden von Studierenden mit geringen wahrnehmenden Fähigkeiten häufiger durchgeführt. Zusätzlich wurde beobachtet, dass Studierende mit niedrigen wahrnehmenden Fähigkeiten deutlichere Unterschiede in den Transitionen zu den Datenpunkten zwischen den beiden Graphentypen zeigen. Das deutet darauf hin, dass Studierende mit geringen wahrnehmenden Fähigkeiten Schwierigkeiten beim Wechsel zwischen Darstellungstypen Probleme haben (Toker et al., 2013). Abgesehen davon zeigt sich der Einfluss der verbalen Arbeitsgedächtnisspanne hingegen in der Konsistenz der Informationsextraktion und -verarbeitung. Studierende mit geringer verbaler Arbeitsgedächtnisspanne zeigen häufiger inkonsistente Sakkadenrichtungen beim Problemlösen (gemessen an Standardabweichung der Sakkadenrichtungen). Allgemein ist zu berücksichtigen, dass es klare Deckeneffekte in der Studie von Toker et al. (2013) gab und ausbleibende signifikante Effekte des räumlichen Vorstellungsvermögens wahrscheinlich darauf zurückzuführen sind.

Abgesehen davon beobachteten Huang und Chen (2016) Geschlechterunterschiede beim Problemlösen mit Graphen und stellten heraus, dass Schüler*innen der 8. und 9. Klasse tendenziell weniger Aufmerksamkeit auf die Graphen und signifikant mehr Aufmerksamkeit auf den Fragetext und die Antwortoptionen richteten als gleichaltrige Schüler. Gleichzeitig führten Schüler*innen signifikant mehr Transitionen zwischen dem Text und dem Graphen durch als Schüler.

11.5 Diskussion

Dieses Kapitel stellt die publizierten Ergebnisse zum Blickverhalten im Falle einer Verknüpfung mit anderen Repräsentationen, den Einfluss des Darstellungstyps sowie des Kontexts und die Ergebnisse zum Einfluss weiterer Faktoren auf das

Blickverhalten dar. Dabei wurde beobachtet, dass ein Graph während eines Lernprozesses im Vergleich zu einem Problemlöseprozess anscheinend weniger Aufmerksamkeit bekommt. Diese Beobachtung wurde jedoch nur durch den Vergleich zweier Studien festgemacht und muss unter verschiedenen Kontrollbedingungen, wie beispielsweise der inhaltlichen Redundanz der Informationen im Text und im Graphen beim Lernprozess, noch bestätigt werden.

Die Ergebnisse von Goldberg und Helfman (2011) implizieren außerdem, dass bei einer Test- und Lernmaterialentwicklung insbesondere auf einen einheitlichen Graphentyp und -stil geachtet werden sollte. Eine abrupte Änderung des Darstellungstyps oder -stils kann die Lernleistung oder Antwortkorrektheit sowie die Effizienz in der visuellen Verarbeitung beeinträchtigen. Diese Leistungsänderung ist jedoch wahrscheinlich geringer für Lernende mit hohen perzeptuellen Fähigkeiten.

Abgesehen davon wurde in vier Studien direkt der Einfluss des Kontexts (Tai et al., 2006; Susac et al., 2018; Klein et al., 2019; Brückner et al., 2020), in dem eine Aufgabe mit Graphen präsentiert wird, untersucht. Zudem erlaubt eine weitere Studie eine Aussage über den Einfluss des domänenspezifischen Vorwissens auf die Expertise in der visuellen Verarbeitung (Goldberg & Helfman, 2011). Diese Studien ergaben, dass 1) die visuelle Expertise im Umgang mit Graphen an die Domäne gekoppelt ist, aber 2) die visuelle Expertise nicht vom domänenspezifischen Vorwissen ausgelöst wird. Stattdessen wird eine visuelle Expertise bei Graphen durch den häufigen Umgang gesteigert. Dennoch zeigt sich, dass 3) eine geringere domänenspezifische Expertise beispielsweise mit mehr visuellen Transitionen zwischen den Antwortmöglichkeiten und dem Graphenbereich beim Problemlösen einhergeht.

Darüber hinaus stellte sich heraus, dass neben der Expertise bei den kognitiven Maßen auch das räumliche Vorstellungsvermögen ebenso wie die verbale Arbeitsgedächtnisspanne von Bedeutung sind (Toker et al., 2013). Diese kognitiven Maße wirken sich beide auf die Effizienz in der Verarbeitung von Graphen aus und haben somit auch potenziell einen Einfluss auf die kognitive Belastung beim Lernen und Problemlösen mit Graphen. Es ist allerdings notwendig, letzteren Zusammenhang gezielt zu studieren. Neben den kognitiven Fähigkeiten erfüllen auch die perzeptuellen Fähigkeiten eine wesentliche Funktion. Studierende mit niedrigen wahrnehmenden Fähigkeiten zeigten deutlichere Unterschiede in den Transitionen zu den Datenpunkten zwischen linear und radial dargestellten Graphen. Dieses Ergebnis deutet daraufhin, dass Studierende mit hohen wahrnehmenden Fähigkeiten eine höhere Flexibilität bei Repräsentationswechseln haben und möglicherweise auch in der Verknüpfung von visuellen Repräsentationen, welches eine zentrale Bedeutung beim Lernen mit multiplen Repräsentationen hat (Küchemann et al., 2019; Rau, 2017).

Zusätzlich zeigt die Analyse von metakognitiven Fähigkeiten in eine vielversprechende Richtung. Klein et al. (2020) konnten zeigen, dass zwischen der Antwortsicherheit und der Aufmerksamkeitsdauer auf den Antwortoptionen ein starker Zusammenhang besteht. Im Allgemeinen haben metakognitive Prozesse eine zentrale Bedeutung für Lern- und Problemlöseprozesse (Scheiter et al., 2017).

Sie sind jedoch in Bezug auf das Blickverhalten beim Lernen und Problemlösen mit Graphen weitgehend unerforscht und bilden daher eine interessante Richtung für zukünftige Forschung der Aufmerksamkeitsverteilung beim Lernen und Problemlösen mit multiplen Repräsentationen.

Literatur

- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, 125, 413–428.
- Andrá, C., Lindström, P., Arzarello, F., Holmqvist, K., Robutti, O., & Sabena, C. (2015). Reading mathematics representations: An eye-tracking study. *International Journal of Science and Mathematics Education*, 13(2), 237–259.
- Atkins, R. M., & McNeal, K. S. (2018). Exploring differences among student populations during climate graph reading tasks: An eye tracking study. *Journal of Astronomy & Earth Sciences Education (JAESE)*, 5(2), 85–114.
- Beichner, R. J. (1993). Misunderstandings of Kinematics Graphs. In *The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics*, Misconceptions Trust: Ithaca, NY.
- Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, 62(8), 750–762.
- Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education*, 53(4), 1207–1217.
- Brückner, S., Zlatkin-Troitschanskaia, O., Küchemann, S., Klein, P., & Kuhn, J. (2020). Changes in students' understanding of and visual attention on digitally represented graphs across two domains in higher education: A postreplication study. *Frontiers in Psychology*, 11, 2090.
- Cowie, B., & Cooper, B. (2017). Exploring the challenge of developing student teacher data literacy. *Assessment in Education: Principles, Policy & Practice*, 24(2), 147–163.
- Dorsch, Lexikon der Psychologie. (2021). *Repräsentation*. Abgerufen von <https://dorsch.hogrefe.com/stichwort/repraesentation>.
- Duden. (2021). *Diagramm*. Abgerufen von <https://www.duden.de/rechtschreibung/Diagramm>.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1–25.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552.
- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, 47(2), 183–210.
- Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization*, 10(3), 182–195.
- Harsh, J. A., Campillo, M., Murray, C., Myers, C., Nguyen, J., & Maltese, A. V. (2019). “Seeing” Data like an expert: An eye-tracking study using graphical data representations. *CBE – Life Sciences Education*, 18(3), ar32.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141–158.
- Ho, H. N. J., Tsai, M. J., Wang, C. Y., & Tsai, C. C. (2014). Prior knowledge and online inquiry-based science reading: Evidence from eye tracking. *International Journal of Science and Mathematics Education*, 12(3), 525–554.
- Huang, P. S., & Chen, H. C. (2016). Gender differences in eye movements in solving text-and-diagram science problems. *International Journal of Science and Mathematics Education*, 14(2), 327–346.

- Kekule, M. (2014). Students' approaches when dealing with kinematics graphs explored by eye-tracking research method. In *Proceedings of the Frontiers in Mathematics and Science Education Research Conference, FISER* (S. 108–117).
- Kim, S., Lombardino, L. J., Cowles, W., & Altmann, L. J. (2014). Investigating graph comprehension in students with dyslexia: An eye tracking study. *Research in Developmental Disabilities, 35*(7), 1609–1622.
- Klein, P., Küchemann, S., Brückner, S., Zlatkin-Troitschanskaia, O., & Kuhn, J. (2019). Student understanding of graph slope and area under a curve: A replication study comparing first-year physics and economics students. *Physical Review Physics Education Research, 15*(2), 020116.
- Klein, P., Lichtenberger, A., Küchemann, S., Becker, S., Kekule, M., Viiri, J., & Kuhn, J. (2020). Visual attention while solving the test of understanding graphs in kinematics: An eye-tracking analysis. *European Journal of Physics, 41*(2), 025701.
- Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science, 31*(4), 549–579.
- Küchemann, S., Klein, P. & Kuhn, J. (2019). Best of Germany: VorleXung: Cross-linking Recitation Sessions and Physics Lectures using eXperiment-based Video-Analysis Tasks. In J. Theo Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning* (pp. 152-157). Amsterdam, Netherlands: Association for the Advancement of Computing in Education (AACE).
- Küchemann, S., Klein, P., Becker, S., Kumari, N., & Kuhn, J. (2020a). Classification of Students' conceptual understanding in STEM education using their visual attention distributions: A comparison of three machine-learning approaches. *CSEDU, 1*, 36–46. <https://doi.org/10.5220/0009359400360046>.
- Küchemann, S., Klein, P., Fouckhardt, H., Gröber, S., & Kuhn, J. (2020b). Students' understanding of non-inertial frames of reference. *Physical Review Physics Education Research, 16*(1), 010112.
- Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10*, 90–115.
- Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research, 60*(1), 1–64.
- Lemke, J. L. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In Martin, J. R. & R. Veel (Hrsg.), *Reading Science* (S. 87–113). Routledge, New York, NY.
- Lilienthal, A. J., & Schindler, M. (2019). Current trends in eye tracking research in mathematics education: A PME literature review: A PME survey. In *Annual Meeting of the International Group for the Psychology of Mathematics Education (PME-43)*, Pretoria, South Africa, July 7–12, 2019 (Bd. 4, S. 62–72).
- Madsen, A. M., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2012). Differences in visual attention between those who correctly and incorrectly answer physics problems. *Physical Review Special Topics-Physics Education Research, 8*(1), 010122.
- Madsen, A., Rouinfar, A., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2013). Can short duration visual cues influence students' reasoning and eye movements in physics problems? *Physical Review Special Topics – Physics Education Research, 9*(2), 020104.
- Matzen, L. E., Haass, M. J., Divis, K. M., Wang, Z., & Wilson, A. T. (2017). Data visualization saliency model: A tool for evaluating abstract data visualizations. *IEEE Transactions on Visualization and Computer Graphics, 24*(1), 563–573.
- McDermott, L. C., Rosenquist, M. L., & Van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics, 55*(6), 503–513.
- Murray, J. (2003). Contemporary literacy: Essential. *Multimedia Schools, 10*(2), 15–18.
- National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: Author.

- Okan, Y., Galesic, M., & Garcia-Retamero, R. (2016). How people with low and high graph literacy process health graphs: Evidence from eye-tracking. *Journal of Behavioral Decision Making*, 29(2–3), 271–294.
- Peebles, D., & Cheng, P. C. H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors*, 45(1), 28–46.
- Polatsek, P., Waldner, M., Viola, I., Kapec, P., & Benesova, W. (2018). Exploring visual attention and saliency modeling for task-based visual analysis. *Computers & Graphics*, 72, 26–38.
- Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, 29(4), 717–761.
- Roth, W. M. (2003). *Toward an anthropology of graphing*. In *Toward an Anthropology of Graphing* (S. 1–21). Dordrecht: Springer.
- Rouinfar, A., Agra, E., Larson, A. M., Rebello, N. S., & Loschky, L. C. (2014). Linking attentional processes and conceptual problem solving: Visual cues facilitate the automaticity of extracting relevant information from diagrams. *Frontiers in Psychology*, 5, 1094.
- Scheiter, K., Fillisch, B., Krebs, M. C., Leber, J., Ploetzner, R., Renkl, A., ... & Zimmermann, G. (2017). How to design adaptive information environments to support self-regulated learning with Multimedia. In *Informational Environments* (S. 203–223). Springer.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156.
- Strobel, B. (2018). *Der Einfluss visueller Gestaltungsmerkmale auf kognitive Prozesse bei der Bearbeitung von Aufgaben mit Diagrammen*, Dissertation. Entnommen aus dem Open Access Repository MACAU (urn:nbn:de:gbv:8-diss-242652).
- Strobel, B., Sass, S., Lindner, M. A., & Köller, O. (2016). Do graph readers prefer the graph type most suited to a given task? Insights from eye tracking. *Journal of Eye Movement Research*, 9(4), 1–15.
- Strobel, B., Lindner, M. A., Saß, S., & Köller, O. (2018). Task-irrelevant data impair processing of graph reading tasks: An eye tracking study. *Learning and Instruction*, 55, 139–147.
- Strobel, B., Grund, S., & Lindner, M. A. (2019). Do seductive details do their damage in the context of graph comprehension? Insights from eye movements. *Applied Cognitive Psychology*, 33(1), 95–108.
- Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, 104, 147–200.
- Susac, A., Bubic, A., Kazotti, E., Planinic, M., & Palmovic, M. (2018). Student understanding of graph slope and area under a graph: A comparison of physics and nonphysics students. *Physical Review Physics Education Research*, 14(2), 020109.
- Tai, R. H., Loehr, J. F., & Brigham, F. J. (2006). An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *International Journal of Research & Method in Education*, 29(2), 185–208.
- Toker, D., Conati, C., Steichen, B., & Carenini, G. (2013, April). Individual user characteristics and information visualization: Connecting the dots through eye tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (S. 295–304).
- Vila, J., & Gomez, Y. (2016). Extracting business information from graphs: An eye tracking experiment. *Journal of Business Research*, 69(5), 1741–1746.
- Wineburg, S., Breakstone, J., McGrew, S., and Ortega, T. (2018). Why google can't save us: the challenges of our post-Gutenberg moment. In O. Zlatkin-Troitschanskaia, G. Wittum, & A. Dengel (Hrsg.), *Positive Learning in the Age of Information (PLATO): A Blessing or a Curse?* (S. 221–228). Springer. https://doi.org/10.1007/978-3-658-19567-0_13.



Wechsel zwischen Diagramm und Formel im Kontext von Vektorfeldern – Einfluss der Aufgabenkomplexität auf Indikatoren visueller Aufmerksamkeit

12

Larissa Hahn, Stefan Halverscheid, Jochen Kuhn und Pascal Klein

Inhaltsverzeichnis

12.1	Einleitung	194
12.2	Theoretischer Hintergrund	195
12.3	Forschungsfrage	198
12.4	Material und Methode	198
12.5	Ergebnisse	201
12.6	Diskussion	203
	Literatur	206

L. Hahn (✉) · P. Klein

Fachbereich Physik, Georg-August-Universität Göttingen, Göttingen, Deutschland

E-Mail: larissa.hahn@uni-goettingen.de

P. Klein

E-Mail: pascal.klein@uni-goettingen.de

S. Halverscheid

Fachbereich Mathematik und Informatik, Georg-August-Universität Göttingen,
Göttingen, Deutschland

E-Mail: stefan.halverscheid@uni-goettingen.de

J. Kuhn

Fachbereich Physik, Technische Universität Kaiserslautern, Kaiserslautern, Deutschland

E-Mail: jochen.kuhn@physik.uni-muenchen.de

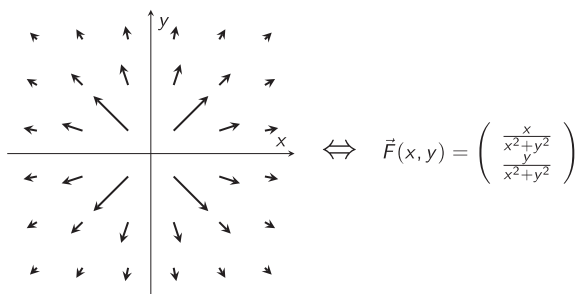
12.1 Einleitung

Vektoren und Vektorfelder sind zentrale Konzepte in der Mathematik und finden Anwendung in verschiedenen Teilgebieten der Naturwissenschaften. Insbesondere in der Physik sind sie häufig vertreten; so dienen sie z. B. der Beschreibung von Gravitationsfeldern in der Mechanik oder Strömungsfeldern in der Elektrodynamik (z. B. Papula, 2008). Bei Vektorfeldern handelt es sich um mathematische Objekte, die jedem Raumpunkt einen Vektor zuordnen (Königsberger, 2004). Sie verknüpfen vektorielle Objekteigenschaften mit dem Feldbegriff. Als Spezialisierung der in physikalischen Anwendungen typischen dreidimensionalen Felder verwendet die Physik zumeist ebene, zeitunabhängige Vektorfelder (Papula, 2008). Sind beide Komponenten eines Feldes unabhängig von den kartesischen Koordinaten, besitzt der Feldvektor in jedem Punkt dieselbe Richtung und denselben Betrag; es handelt sich um ein homogenes Vektorfeld (Königsberger, 2004). Im Nachfolgenden wird ein Vektorfeld, bei dem eine Komponente konstant null ist, als eindimensional und ein Feld mit zwei von null verschiedenen Komponenten als zweidimensional bezeichnet. Die beschriebenen Eigenschaften bilden die Grundlage für eine Klassifizierung von Vektorfeldern entsprechend ihrer Aufgabenkomplexität (Abschn. 12.4).

Im mathematisch-naturwissenschaftlichen Unterricht werden Kräfte schon in der Sekundarstufe I mithilfe gerichteter Pfeile dargestellt; im Mathematik- und Physikunterricht ist die Analytische Geometrie (im Rahmen elektromagnetischer Felder) Kernthema der Sekundarstufe II (z. B. Niedersächsisches Kultusministerium, 2015, 2018). Eine erste tiefere Auseinandersetzung mit Vektorfeldern und damit zusammenhängenden mathematischen Konzepten, wie der Divergenz und Rotation, erfolgt erst zu Beginn eines naturwissenschaftlichen Studiums. Zu den dort vorherrschenden Lernschwierigkeiten in Bezug auf verschiedene vektorielle Konzepte wurden bereits zahlreiche Studien durchgeführt (z. B. Barniol & Zavala, 2010; Knight, 1995); über die Probleme im Zusammenhang mit Vektorfeldern hingegen ist nur wenig bekannt.

Vektoren werden typischerweise bildhaft (piktoriell) oder algebraisch dargestellt. Dies überträgt sich auf Vektorfelder und in der höheren Mathematik- bzw. Physikausbildung wird vor allem die kontextabhängige Verwendung verschiedener Darstellungsformen von Vektorfeldern thematisiert (z. B. Bollen et al., 2015). Abb. 12.1 zeigt ein bekanntes Beispiel aus physikalischen Einführungsveranstaltungen, in denen das dargestellte Vektorfeld u. a. als Coulomb-Feld einer positiven Punktladung interpretiert wird. In der Repräsentation als Vektorfelddiagramm sind die Radialsymmetrie des Feldes und seine Abstandsabhängigkeit vom Ursprung einfach zu erkennen. Diese Informationen anhand der algebraischen Darstellung zu gewinnen, erfordert weitaus mehr Aufwand. Mithilfe der Formel-darstellung kann hingegen schnell nachgerechnet werden, dass es sich um ein konservatives Feld handelt. Insofern besitzen beide Repräsentationen kontext- und zielabhängige Vorteile.

Abb. 12.1 Verwendung verschiedener Repräsentationsformen von Vektorfeldern



Zahlreiche Studien haben gezeigt, dass ein flexibler Umgang mit verschiedenen Repräsentationsformen, der u. a. einen reibungslosen Wechsel zwischen verschiedenen Darstellungen beinhaltet, essenziell für die Entwicklung fachspezifischer Expertise ist (z. B. Bollen et al., 2017; van Heuvelen, 1991). Umso wichtiger ist demnach die Analyse von Lernschwierigkeiten, die beim Wechsel mit den genannten Repräsentationsformen auftreten, da die Koordination ebendieser eine zusätzliche Herausforderung im Umgang mit Vektoren und Vektorfeldern darstellt. Um grundlegende Erkenntnisse zum reinen Repräsentationswechsel zu erhalten und einen Bias durch die zusätzliche Schwierigkeit eines physikalischen Kontexts zu vermeiden, thematisiert dieses Kapitel ausschließlich ihre Darstellung via Formel und Diagramm. Hierbei wird Eye-Tracking als Forschungsmethode zur Untersuchung des prozeduralen Vorgehens beim Repräsentationswechsel eingesetzt, da es aufgrund der starken Relation zwischen dem Blickverhalten und der kognitiven Verarbeitung detaillierte Einblicke in Lern- und Problemlöseprozesse ermöglicht (Just & Carpenter, 1980; Rakoczi, 2012).

Anknüpfend an bisherige Forschungsergebnisse (z. B. Chen et al., 2011; Bollen et al., 2017; Klein et al., 2018) fokussiert diese Studie den Einfluss der Aufgabenkomplexität von Vektorfeldern auf die kognitive Verarbeitung bei Repräsentationswechseln. Dieser Zusammenhang wird unter Verwendung blickbasierter Positions- und Zeitmaße quantitativ untersucht.

12.2 Theoretischer Hintergrund

Probleme im Umgang mit vektoriellen Konzepten stellen einen zentralen Aspekt der (physik-)didaktischen Forschung dar. Zahlreiche Studien (z. B. Barniol & Zavala, 2010) haben gezeigt, dass StudienanfängerInnen sowohl im rein mathematischen als auch im physikalischen (Anwendungs-)Kontext mit vielfältigen Schwierigkeiten im Hinblick auf unterschiedliche vektorielle Konzepte konfrontiert werden. Umfassende Kenntnisse dieser sind Voraussetzung für das Verständnis von Vektorfeldern, weshalb sich Probleme im Umgang mit Vektoren unmittelbar auswirken (Knight, 1995). Zusätzlich ist es möglich, dass Lernende zwar die Grundlagen vektorieller Konzepte verstehen, aber Schwierigkeiten haben, diese auf Vektorfelder zu übertragen (Bollen et al., 2017). Über konkrete

Lernschwierigkeiten im Umgang mit Vektorfeldern ist bisher noch wenig bekannt; erste Einblicke geben Untersuchungen von Singh und Maries (2013) und Bollen et al. (2017) sowie die Eye-Tracking-Studien von Klein et al. (2018, 2019). Letztere beschäftigten sich mit der qualitativen Interpretation der Rotation und Divergenz grafischer Vektorfeldrepräsentationen und entwickelten visuelle Instruktionen und Hilfestellungen zur qualitativen Interpretation dieser Konzepte in Vektorfelddiagrammen.

Da es für viele mathematische und vor allem physikalische Konzepte zweckmäßig ist, werden diese meist unter Verwendung einer oder mehrerer Repräsentationen dargestellt (De Cock, 2012). Nach Schnotz et al. (2010) können hierbei zwei grundlegende Formen differenziert werden: Deskriptive Repräsentationen, die ein Referenzobjekt anhand von Symbolen darstellen, und depiktive Repräsentationen, welche strukturelle Gemeinsamkeiten zwischen verschiedenen Referenzobjekten hervorheben. Gängige Repräsentationen von Vektorfeldern sind sowohl algebraische Formelausdrücke mithilfe von Einheitsvektoren (deskriptiv) als auch grafische Darstellungen, u. a. als Vektorfelddiagramm (depiktiv; Bollen et al., 2017).

Werden verschiedene Repräsentationen in einer (Lern-)Umgebung zusammengeführt, wird von der Verwendung multipler Repräsentationen (MRs) gesprochen (Ainsworth, 2006). Ein zentraler Gegenstand der naturwissenschaftlich-didaktischen Forschung ist die Analyse der (Lern-)Prozesse in diesem Zuge, bei der kognitive Lerntheorien wie die *Cognitive Theory of Multimedia Learning* (CTML) nach Mayer (2005) eine zentrale Rolle spielen. Bezüglich der Rolle von multiplen Repräsentationen in einer Lernumgebung formulierte Ainsworth (1999) drei zentrale Funktionen:

1. MRs enthalten ergänzende Informationen oder unterstützen komplementäre kognitive Prozesse (Seufert, 2003).
2. MRs können Lernenden helfen, ein besseres Verständnis einer Thematik zu entwickeln, indem eine Repräsentation verwendet wird, um mögliche (Fehl-)Interpretationen einer anderen einzuschränken.
3. MRs ermöglichen ein tieferes Verständnis einer Situation oder eines Konzepts.

Darauf aufbauend berichten zahlreiche Untersuchungen von einem positiven Effekt der Verwendung multipler Repräsentationen auf den Wissenserwerb und die Problemlösefähigkeiten (z. B. Dufresne et al., 1997). So ermittelten Nieminen et al. (2012) eine starke Korrelation zwischen der Fähigkeit von Lernenden, multiple Repräsentationen konsistent zu interpretieren, und ihrem Lernzuwachs. Außerdem ist das Verständnis, das Lernende durch die Nutzung multipler Repräsentationen gewinnen, breiter, tiefer (Even, 1998), flexibler und robuster (Ainsworth et al., 2002) im Vergleich zu dem, was sie unter Verwendung einzelner Repräsentationen generieren, denn insbesondere grafische Repräsentationen helfen, die Kluft zwischen der verbalen und der abstrakt-mathematischen Darstellung zu überwinden (van Heuvelen & Zou, 2001). Insofern kann die Verwendung multipler

Repräsentationen synergetische Effekte auf die Konstruktion kohärenter Wissensstrukturen haben (Seufert, 2003). Studien von De Leone und Gire (2006) sowie Rosengrant et al. (2007, 2009) zeigen zudem, dass Studierende, deren Studieninhalte den Einsatz multipler Repräsentationen fokussieren, dazu neigen, multiple Darstellungen zur selbstständigen und flexiblen Problemlösung heranzuziehen, statt rein prozedural vorzugehen (Dufresne et al., 1997).

Trotz der beschriebenen potenziellen Vorteile gibt es zahlreiche Studien, die offenlegen, dass die Verwendung multipler Repräsentationen nicht per se zu einem höheren Lernerfolg führt (z. B. Ainsworth, 2008; Yerushalmy, 1991). Um von den Vorteilen, die multiple Repräsentationen beim Lernen und Problemlösen mit sich bringen, profitieren zu können, ist vielmehr ein tieferes Verständnis der Repräsentationen, „representational competence“ (De Cock, 2012; Nistal et al., 2012; Rau, 2017), notwendig, welches mithilfe zweier Kompetenzen beschrieben wird:

1. *Representational fluency* befähigt dazu, korrekt und schnell zwischen verschiedenen Repräsentationsformen zu wechseln und zu übersetzen (Hill, 2015).
2. *Representational flexibility* ermöglicht die Wahl einer angemessenen Repräsentationsform in einer gegebenen Problem- oder Lernsituation.

Zur Systematisierung der Probleme von Studierenden, die bei Repräsentationswechseln zwischen Vektorfelddiagrammen, Feldliniendiagrammen und algebraischen Ausdrücken auftreten, entwickelten Bollen et al. (2017) mithilfe eines *Vector Field Representations Test* (VFR) eine entsprechende Taxonomie. Bei der Konstruktion eines Vektorfelddiagramms anhand eines algebraischen Formelausdrucks zeigen sich hier vor allem Fehler, die auf Probleme bei der Vektoraddition zurückzuführen sind. Beim umgekehrten Repräsentationswechsel stellen die Wahl eines geeigneten Koordinatensystems und die Verwendung von Einheitsvektoren ein Problem dar. Gire und Price (2012), die sich mit Repräsentationswechseln im Kontext von elektromagnetischen Feldern beschäftigten, stellten zudem fest, dass Studierende häufig Probleme haben, zwischen Koordinaten und Komponenten zu differenzieren; insbesondere gilt dies bei gemischten Abhängigkeiten der Komponenten von den Koordinaten.

Die CTML nach Mayer (2005) beschreibt den Einfluss der durch multiple Repräsentationen erzeugten multimedialen Umgebung auf die kognitiven Vorgänge beim Lernprozess. Dabei konzentriert sie sich auf Prozesse im Zusammenhang mit Wörtern und Bildern, wobei Formeln bezüglich ihrer Verarbeitung als Text betrachtet werden können (Schnotz & Bannert, 2003), und beschreibt drei mentale Prozesse (SOI-Modell). Bei der Informationsentnahme aus multimedialen Stimuli werden Inhalte für einen kurzen Zeitraum im sensorischen Speicher präsent gehalten. Anschließend findet eine Auswahl relevanter Wörter und Bildinhalte für die Verarbeitung im verbalen bzw. visuellen Arbeitsgedächtnis statt (Selektion). Hier laufen bewusste Organisationsprozesse ab, in deren Zuge die selektierten verbalen und visuellen Informationsbausteine zu kohärenten verbalen

und piktorialen mentalen Modellen überführt werden (Organisation). Zur Wissenskonstruktion im Arbeitsgedächtnis findet zusätzlich eine Aktivierung des bereits erworbenen Wissens des oder der Lernenden aus dem Langzeitgedächtnis statt, welches mit den neuen Modellen verknüpft wird (Integration).

In Anlehnung an die *eye-mind assumption* nach Just und Carpenter (1980) wird häufig der Zusammenhang verschiedener Eye-Tracking-Metriken mit kognitiven Prozessen herausgestellt. So gibt nach Tsai et al. (2012) die Position der Fixation Auskunft darüber, was kognitiv verarbeitet wird, und die Fixation Duration kann als Maß für die Aufmerksamkeits- bzw. Verarbeitungsdauer und somit für die Verarbeitungsschwierigkeit verstanden werden – je länger eine Information fixiert wird, desto komplexer ist sie oder desto tiefer wird sie verarbeitet. Außerdem berichten Chen et al. (2011) sowie Zagermann et al. (2016), dass Fixation Duration und Fixation Count Indikatoren für eine erhöhte Aufmerksamkeit darstellen, die erforderlich ist, wenn die Komplexität einer Aufgabe zunimmt. In Bezug auf das SOI-Modell der CTML gelten Fixation Duration und Fixation Count als Indikatoren für ablaufende Organisationsprozesse (z. B. Scheiter & Eitel, 2017).

12.3 Forschungsfrage

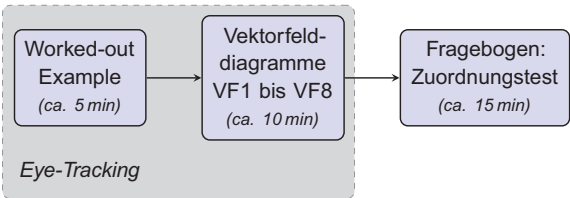
Dieses Kapitel thematisiert den Einfluss der Komplexität von Vektorfeldern auf die visuelle Aufmerksamkeit beim Repräsentationswechsel. Hierfür wird der Komplexitätsgrad über sachspezifische Eigenschaften von Vektorfeldern definiert und der Zusammenhang mit der kognitiven Verarbeitung untersucht. Im Sinne der CTML stehen dabei Organisationsprozesse im Vordergrund, sodass, aufbauend auf bisherigen Untersuchungen, die Beziehung zwischen inhärenter Aufgabenkomplexität und Eye-Tracking-Maßen auf Vektorfelder angewendet werden soll (Chen et al., 2011; Zagermann et al., 2016). Es ergibt sich folgende Forschungsfrage:

Wie beeinflusst der Komplexitätsgrad zeitunabhängiger Vektorfelder die visuelle Aufmerksamkeit – Visit Duration (VD), Fixation Count (FC) und Fixation Duration (FD) – beim Repräsentationswechsel?

12.4 Material und Methode

Die Daten wurden im Wintersemester 2019/2020 mit Studierenden verschiedener naturwissenschaftlicher Ingenieurstudiengänge an der TU Kaiserslautern im Rahmen der Vorlesung „Experimentalphysik I für angehende Ingenieure“ erhoben. Insgesamt nahmen 147 Studierende, überwiegend aus dem ersten und zweiten Fachsemester, an der Studie teil, die eine normale Sehschärfe oder korrigierte Fehlsichtigkeit aufwiesen. Da jene Studierende, deren Muttersprache nicht Deutsch ist, zumeist schwächere Leistungen aufwiesen und ein zu Beginn des

Abb. 12.2 Studienablauf



Eye-Tracking-Experiments präsentierte Worked-out Example zudem stark text- und sprachfokussiert ist, wurden nur MuttersprachlerInnen in der Analyse berücksichtigt ($N = 114$, 82 % männlich, Durchschnittsalter 20.1 ± 2.3 Jahre).

Die Eye-Tracking-Erhebung wurde mit einer individuellen Kalibrierung der Augenbewegungen am Computer eingeleitet (Abb. 12.2). Anschließend erhielten die Studierenden ein Worked-out Example (im Sinne von Renkl, 1999) zur Darstellung von Vektorfeldern, das sich aus einer textbasierten Einführung in die grafische und formale Darstellung von Vektorfeldern sowie zwei anschaulichen Beispielen zum Repräsentationswechsel zusammensetzt. Hierbei werden insbesondere das Konzept sowie die Notation der Komponentenzerlegung thematisiert.

Im Anschluss sollten die Studierenden zu acht Vektorfelddiagrammen (VF1 bis VF8) aus jeweils vier Antwortmöglichkeiten die korrekte Formel in Koordinatendarstellung auswählen (Abb. 12.4 links). Diese wurden in ähnlicher Form bereits in anderen Studien eingesetzt und genügen bestimmten Anforderungen (Klein et al., 2018; Singh & Maries, 2013). So wurde berücksichtigt, dass die Vektorfelder keiner physikalischen Realität entsprechen, sodass keine Beeinflussung der Repräsentationskompetenz durch physikalisches Fachwissen stattfindet. Außerdem sind alle Längenskalen arbiträr zu interpretieren und die Abhängigkeiten der Vektorfeldkomponenten sind höchstens linear; auf komplexere Zusammenhänge wurde zur besseren Analyse der Augenbewegungen verzichtet. Alle Diagramme wurden als zweidimensionale kartesische Koordinatensysteme mit etwa 25 repräsentativen Vektorpfeilen dargestellt.

Um die Untersuchung der beschriebenen Vektorfelder zu systematisieren, ist in Tab. 12.1 ein A-priori-Kategoriensystem verschiedener Komplexitätsgrade dargestellt, das anhand sachspezifischer Kriterien unabhängig von den empirischen

Tab. 12.1 Klassifizierung der Vektorfelder nach Dimensionalität und Komplexität

1. Komponente	2. Komponente	Dimension	Komplexitätsgrad	Vektorfeld
0	0	–	0	–
Konstant	0	1	1	VF3
Konstant	Konstant	2	2	VF1
Variabel	0	1	3	VF2, VF7
Variabel	Konstant	2	4	VF5, VF6
Variabel	Variabel	2	5	VF4, VF8

Daten gebildet wurde. Dabei sind die Kategorien auf Grundlage der Dimensionalität und Komponentenabhängigkeit zweidimensionaler, zeitunabhängiger Vektorfelder disjunkt und erschöpfend definiert. Das Schema unterscheidet sechs Komplexitätsgrade, wobei das triviale Vektorfeld zum Zweck der Vollständigkeit in die Klassifikation aufgenommen wurde (Beispiele in Abb. 12.3). Für alle weiteren Komplexitätsgrade ist die Bezeichnung der ersten und der zweiten Komponente willkürlich und vertauschbar.

Alle Stimuli im Eye-Tracking-Test wurden auf einem 22-Zoll-Computerbildschirm mit einer Auflösung von 1920×1080 Pixeln und einer Bildwiederholrate von 75 Hz präsentiert. Für die Messung der Augenbewegungen wurde der monitorbasierte Eye-Tracker Tobii X3-120 verwendet, der mit einer Genauigkeit von $< 0.40^\circ$ Sehwinkel, einer Abtastfrequenz von 120 Hz sowie mithilfe von Infrarot arbeitet. Der Abstand zwischen Bildschirm und ProbandIn betrug dabei etwa 60 cm. Zur Unterscheidung zwischen Fixationen und Sakkaden wurden für die Velocity Threshold Identification ein Geschwindigkeitsschwellenwert von $30^\circ/s$ sowie ein Beschleunigungsschwellenwert von $8500^\circ/s^2$ verwendet. Die Visualisierung und Analyse der Blickdaten mithilfe der Software Tobii Studio 3.4.8 erfolgten durch die Definition von kongruenten AOIs auf allen acht Vektorfeldditems (Abb. 12.4 rechts). Da der Fokus dieses Kapitels auf komplexitätsbedingten Innersubjekteffekten zwischen den Vektorfeldditems liegt, wurde hierbei ausschließlich das Vektorfelddiagramm betrachtet. Der Einfluss der Distraktoren soll im Rahmen dieses Kapitels nicht untersucht werden.

Nach Wahl einer der vier Antwortoptionen gelangten die Studierenden per Tastendruck zur Antworteingabe, wobei der selbstbestimmte Seitenwechsel eine klare Separation der Blickdaten des Problemlöseprozesses und derer der Antworteingabe sicherstellte. Während der gesamten Erhebung konnten die Studierenden weder zurückkehren noch inhaltliche Rückfragen stellen oder Notizen anfertigen. Außerdem gab es weder zeitliche Beschränkungen, noch erhielten die Studierenden Feedback zu ihren Leistungen. Nach Bearbeitung des achten Vektorfelddiagramms wurde die Blickdatenaufnahme beendet; abschließend wurden die soziodemografischen Daten der Studierenden erhoben.

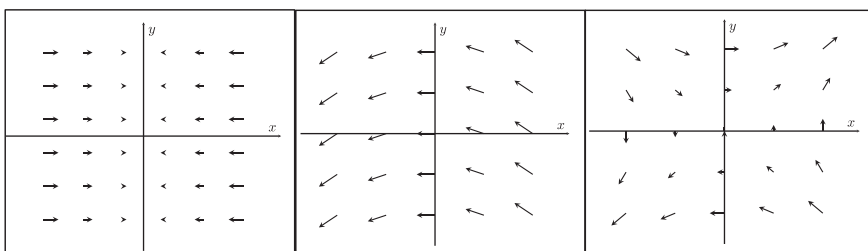


Abb. 12.3 Vektorfelder der Komplexitätsgrade 2 (links, VF2), 4 (Mitte, VF5) und 5 (rechts, VF4)

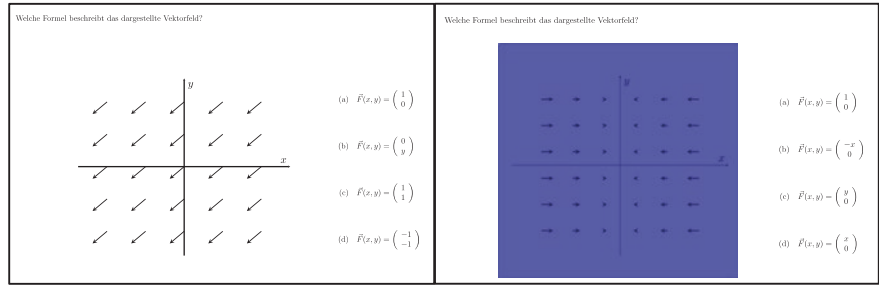


Abb. 12.4 Aufgabe 1 des Eye-Tracking-Testbogens (links; Weiterentwicklung von Klein et al., 2018) und Definition von AOIs in Tobii Studio am Beispiel von VF2 (rechts)

12.5 Ergebnisse

Für die Wahl der korrekten Antwort im Eye-Tracking-Test erhielten die Studierenden einen Punkt. So erreichten sie insgesamt einen mittleren Score von 6.5 ± 1.9 von acht möglichen Punkten bzw. $(81 \pm 24) \%$. Dieser ist als hoch zu bewerten, da die Wahrscheinlichkeit einer korrekt geratenen Antwort bei einer Auswahl aus vier Distraktoren bei 25 % liegt. Das einfachste Item war VF1, das von 96.5 % der Studierenden korrekt gelöst wurde, und die größten Schwierigkeiten bereitete Item VF4 (70.2 % korrekt gelöst). Tab. 12.2 zeigt die Mittelwerte und Standardabweichungen für den FC mit Angabe des jeweiligen Komplexitätsgrades vergleichend für die acht Vektorfelder. Aufgrund der hohen Korrelation zwischen FC und VD sowie FD ($r = 0.982, p < 0.001^{***}$; multiple Korrelationsanalyse) ergeben sich sehr ähnliche Tendenzen aller drei Metriken, weshalb alle Ergebnisse im Nachfolgenden stellvertretend anhand des FC erläutert werden. Hier zeichnen sich drei Itemgruppen ab: VF1 ($K = 2$) und VF3 ($K = 1$) weisen die geringsten, VF4 und VF8 (beide $K = 5$) die insgesamt höchsten Mittelwerte auf. Außerdem ergeben sich nur geringe Unterschiede zwischen den Werten von VF2 und VF7 (beide $K = 3$) sowie VF5 und VF6 (beide $K = 4$).

Die Beobachtungen suggerieren einen Einfluss des Komplexitätsgrades auf alle drei betrachteten Metriken. Dies wurde mithilfe von drei einfaktoriellen ANOVA-RM untersucht, wobei die acht Vektorfelditems den messwiederholenden Innersubjektfaktor darstellten. Die Ergebnisse der Analysen sind in Tab. 12.3 dargestellt. Es zeigten sich statistisch hochsignifikante Innersubjekteffekte des FC, der VD und der FD für verschiedene Komplexitätsgrade.

Für eine genauere Analyse der Unterschiede wurden Post-hoc-Tests mit Bonferroni-Korrektur für die drei Metriken durchgeführt (Tab. 12.2). Es zeigte sich ein ähnliches Bild wie das oben beschriebene für alle drei Eye-Tracking-Maße, das somit den Eindruck der Dreiteilung bestätigte. Zum einen ergab sich ein hochsignifikanter Unterschied von VF1 ($K = 2$) und VF3 ($K = 1$) gegenüber allen weiteren Vektorfeldern mit jeweils negativen Mittelwertdifferenzen,

Tab. 12.2 Ergebnisse der paarweisen Vergleiche für die FC aller acht Vektorfelder mit Angabe ihres Komplexitätsgrades *K*

	<i>K</i>	<i>M</i>	<i>SD</i>	Bonferroni-korrigierte paarweise Vergleiche							
				VF1	VF2	VF3	VF4	VF5	VF6	VF7	VF8
VF1	2	28.9	17.0		...	...	...	...	...	...	...
VF2	3	64.7	39.8	<0.001***		...	...	...	...	...	...
VF3	1	30.0	20.1	1.000	<0.001**		...	...	...	...	...
VF4	5	179.9	161.7	<0.001***	<0.001***	<0.001***		...	...	...	...
VF5	4	86.7	63.1	<0.001***	0.019*	<0.001***	<0.001***		...	...	...
VF6	4	72.3	43.9	<0.001***	1.000	<0.001***	<0.001***	0.314		...	...
VF7	3	71.3	42.3	<0.001***	1.000*	<0.001***	<0.001***	0.342	1.000		...
VF8	5	98.7	69.9	<0.001***	<0.001***	<0.001***	<0.001***	1.000	0.003**	0.001**	

Anmerkungen: Post-hoc-Tests mit Bonferroni-Korrektur. Angegeben sind der Mittelwert *M*, die Standardabweichung *SD* und die *p*-Werte (***) *p* < 0.001, ** *p* < 0.01, * *p* < 0.05). Der Stichprobenumfang beträgt *N* = 114

Tab. 12.3 Ergebnisse der ANOVA-RM zur Untersuchung des Einflusses des Komplexitätsgrades auf verschiedene Eye-Tracking-Maße

Metrik	F	p	η_p^2	f
FC	60.4	$p < 0.001^{***}$	0.35	0.73
FD	59.6	$p < 0.001^{***}$	0.35	0.73
VD	62.3	$p < 0.001^{***}$	0.36	0.74

Anmerkungen: Drei einfaktorielle ANOVA-RM mit Greenhouse-Geisser-Korrektur. Angegeben sind die Greenhouse-Geisser-korrigierten F -Werte, die Signifikanz p , die Effektgröße η_p^2 und die Effektstärke f .

allerdings keine statistisch relevante Differenz der beiden Vektorfelder untereinander. Zum anderen zeigte sich ein hochsignifikanter Unterschied mit positiven Mittelwertdifferenzen zwischen VF4 ($K = 5$) und den restlichen Vektorfeldern. VF8 unterschied sich ebenfalls statistisch relevant von den übrigen Vektorfeldern, ausgenommen VF5. Dabei war die Mittelwertdifferenz gegenüber VF4 negativ, im Vergleich zu allen anderen Vektorfeldern hingegen positiv. Für alle weiteren Vektorfelder resultierten kaum signifikante Mittelwertunterschiede; so zeigte sich auch hier wieder ein breites Mittelfeld der Vektorfelder VF2, VF5, VF6 und VF7.

12.6 Diskussion

Das Ziel dieser Studie galt der Untersuchung des Zusammenhangs zwischen der (sachlogisch definierten) Komplexität von Vektorfeldern und den Blickbewegungen bei einer Aufgabe, die eine Zuordnung zwischen Vektorfeld und Formel erfordert. Hierbei haben statistische Untersuchungen einen hochsignifikanten Einfluss des Komplexitätsgrades auf die Blickbewegungen ergeben, und es haben sich Veränderungen in den okulomotorischen Variablen insofern gezeigt, als mit steigender Komplexität des Vektorfeldes sowohl FC als auch VD und FD auf den Vektorfeld-AOIs zugenommen hat. Dies ist konsistent mit den Ergebnissen von Chen et al. (2011) sowie Zagermann et al. (2016), die ebenfalls einen Zusammenhang zwischen FD und FC sowie Aufgabenkomplexität ermittelten.

Die eben genannten Studien erklären ihre Ergebnisse vor dem Hintergrund der Cognitive Load Theory (CLT; Chandler & Sweller, 1991). Dabei wird argumentiert, dass das Niveau der intrinsischen kognitiven Belastung für eine bestimmte Aufgabe durch die Elementinteraktivität bestimmt wird (Sweller, 2010). Im Kontext von Vektorfeldern handelt es sich bei der Dimensionalität und den (gemischten) Komponentenabhängigkeiten um Aspekte, die die Elementinteraktivität bestimmen. Da die Darstellung und Notation im Verlauf des Eye-Tracking-Tests nicht verändert wurde, gilt der Extraneous Cognitive Load im Sinne der CLT als konstant. Da der Germane Cognitive Load nur indirekt über Intrinsic und Extraneous Cognitive Load beeinflussbar ist (Sweller, 2010), führt somit eine Veränderung der Elementinteraktivität, die sich direkt auf den Intrinsic

Cognitive Load auswirkt, unmittelbar zu einer veränderten kognitiven Belastung. Umgekehrt ist eine erhöhte kognitive Belastung bei gleichbleibender Gestaltung des Lernmaterials direkt auf eine höhere Elementinteraktivität der zu verarbeitenden Information zurückzuführen. Eine Interpretation der Ergebnisse vor dem Hintergrund der CLT impliziert somit, dass eine höhere Vektorfeldkomplexität mit einer höheren mentalen Belastung einhergeht, welche sich in den Blickbewegungen widerspiegelt. Diese Interpretation ist insbesondere konsistent mit Ergebnissen von Jaeger et al. (2017), die einen hochsignifikanten Zusammenhang zwischen der Aufgabenkomplexität und der subjektiven kognitiven Belastung ermittelten. Zur Prüfung dieser Hypothese wäre eine zusätzliche Erhebung des Cognitive Load mittels Ratingskalen vielversprechend (Leppink et al., 2013), um eine Verbindung zwischen den objektiven Indikatoren und der subjektiven Wahrnehmung von kognitiver Belastung herstellen zu können. Hier zeigt sich Potenzial für weiterführende Studien (Abb. 12.5).

Im Zuge dieser Interpretation ist außerdem zu bedenken, dass sowohl der Germane Cognitive Load als auch der Intrinsic Cognitive Load maßgeblich vom Wissensstand der Lernenden bestimmt werden (Sweller, 2010). So können mehrere interagierende Elemente, die für Lernende mit niedrigem Wissensstand eine hohe Belastung bedeuten, von Lernenden mit höherem Wissensstand als ein Element betrachtet und somit ressourcenschonender verarbeitet werden. Es handelt sich bei dem Vorwissen demzufolge um einen wichtigen Einflussfaktor, der im Zuge weiterführender Überlegungen berücksichtigt werden sollte.

Die Ergebnisse der paarweisen Vergleiche unterstützen ebenfalls die oben formulierten Zusammenhänge insofern, als sich die Mittelwerte der Eye-Tracking-Maße für die Vektorfelder mit den geringsten und höchsten Komplexitätsgraden statistisch signifikant von allen anderen Vektorfeldern unterscheiden. Für Vektorfelder mit $K = 3$ und $K = 4$ ergeben sich ähnliche Werte aller Metriken mit nicht-signifikanten Mittelwertdifferenzen, sodass über die Unterschiede der kognitiven Verarbeitung zwischen diesen Komplexitätsgraden durch die betrachteten Maße

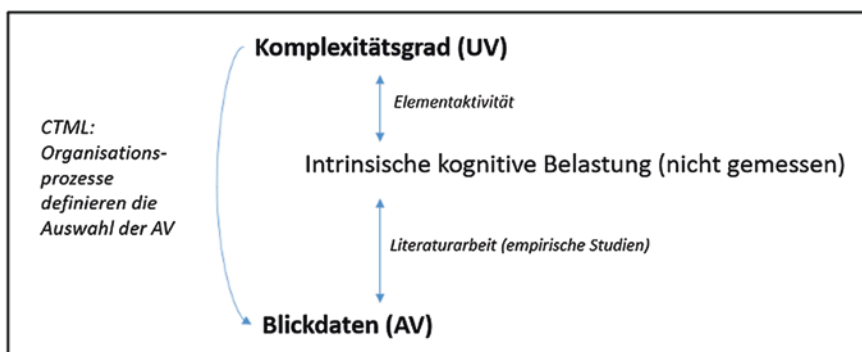


Abb. 12.5 Strukturübersicht der Variablen Komplexitätsgrad, Blickdaten und kognitive Belastung

keine Aussagen getroffen werden können und daher an dieser Stelle nicht weiter diskutiert werden sollen. Hier zeigt sich die Notwendigkeit, weitere Eye-Tracking-Metriken hinzuzuziehen. Zudem fällt auf, dass sich zwischen VF1 und VF3 für alle Maße trotz verschiedener Komplexitätsgrade keine signifikanten Unterschiede ergaben. Dies impliziert, dass der Unterschied, der sich hierbei in der Dimension der Vektorfelder ergibt, für eine vergleichsweise geringe Aufgabenschwierigkeit kaum Auswirkungen auf die kognitive Verarbeitung besitzt.

Für VF4 und VF8 (beide $K = 5$) ist auffällig, dass sie zwar für alle Metriken die höchsten Werte aufweisen, sich ihre Mittelwerte allerdings signifikant unterscheiden. Eine mögliche Erklärung für diese Differenz innerhalb des höchsten Komplexitätsgrades ist, dass VF4 zwei gemischte Abhängigkeiten besitzt, wohingegen VF8 nur eine gemischte Abhängigkeit aufweist. Diese Beobachtung deckt sich mit den Ergebnissen von Gire und Price (2012), die in gemischten Abhängigkeiten eine große Problemquelle lokalisierten. Um diesen Erklärungsansatz näher zu untersuchen, wäre eine feinere Klassifizierung der Komplexitätsgrade unter Einbeziehung gemischter Abhängigkeiten notwendig. Eine weitere Begründung für die Diskrepanz der Eye-Tracking-Maße der Vektorfelder mit $K = 5$ könnte außerdem in einem Lerneffekt liegen, da VF4 das zeitlich erste Vektorfeld mit gemischten Abhängigkeiten in zwei Komponenten darstellt. Auf VF4 folgen hingegen drei Vektorfelder mit variablen Komponenten und/oder gemischten Abhängigkeiten ($K \geq 3$). Daher könnte der Mittelwertunterschied zwischen VF4 und VF8 auch in einem höheren Wissensstand infolge der Bearbeitung vorangegangener Items begründet sein.

Diese Überlegungen sind im Fall weiterführender Untersuchungen dringend zu berücksichtigen, da sie zeigen, dass weit mehr Aspekte als nur die intrinsischen Eigenschaften der Vektorfelder Einfluss auf die kognitive Verarbeitung haben. An dieser Stelle wird außerdem die Notwendigkeit deutlich, die Distraktoren in die Eye-Tracking-Analyse einzubeziehen, um zu untersuchen, inwiefern ein Zusammenhang zwischen dem Blickverhalten auf den Distraktoren und der Komplexität der Vektorfelder besteht. Hierbei sollten insbesondere weitere, vor allem bewegungsbezogene Metriken hinzugezogen werden, um ein detailliertes Bild zu erhalten, das über positions- und zeitbezogene Maße hinausgeht. Dies erscheint insbesondere im Hinblick auf die Ergebnisse von Phillips und Edelman (2008) sinnvoll, die herausstellten, dass Leistungsunterschiede im Kontext von *visual scanning*-Aufgaben okulomotorische Variablen bedingen, die durch sakkadische Metriken besser abgebildet werden können als durch Positionsmaße. Nichtsdestotrotz liefert der Einsatz zeitbezogener Metriken in diesem Kapitel erste wichtige Erkenntnisse für ein tieferes Verständnis der kognitiven Prozesse beim Repräsentationswechsel im Kontext von Vektorfeldern, an die in Zukunft anzuknüpfen ist, um effektive Konsequenzen für die Lehre ziehen und umsetzen zu können. Aus unserer Analyse können wir schließen, dass Eye-Tracking als zusätzliche Methode der Itemanalyse dienen kann, die zur Erforschung von Itemeigenschaften wie Komplexität auf einer prozessbezogenen Ebene beiträgt.

Literatur

- Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 3(2), 131–152. [https://doi.org/10.1016/S0360-1315\(99\)00029-9](https://doi.org/10.1016/S0360-1315(99)00029-9).
- Ainsworth, S. (2006). DeFT: A conceptual framework for learning with multiple representations. *Learning and Instruction*, 16(3), 183–198. <https://doi.org/10.1016/j.learninstruc.2006.03.001>.
- Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In J. K. Gilbert, M. Reimer, & M. Nakhleh (Hrsg.), *Visualization: Theory and practice in science education. models and modeling in science education* (3. Aufl., S. 191–208). Springer.
- Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences*, 11(1) 25–61. https://doi.org/10.1207/S15327809JLS1101_2.
- Barniol, P., & Zavala, G. (2010). Students' understanding of the concepts of vector components and vector products. *AIP Conference Proceedings*, 1289(1), 341–344. <https://doi.org/10.1063/1.3515240>.
- Bollen, L., van Kampen, P., & De Cock, M. (2015). Students' difficulties with vector calculus in electrodynamics. *Physical Review Special Topics – Physics Education Research*, 11(2). <https://doi.org/10.1103/PhysRevSTPER.11.020129>.
- Bollen, L., van Kampen, P., Baily, C., Kelly, M., & De Cock, M. (2017). Student difficulties regarding symbolic and graphical representations of vector fields. *Physical Review Physics Education Research*, 13(2). <https://doi.org/10.1103/PhysRevPhysEducRes.13.020109>.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332. doi: https://doi.org/10.1207/s1532690xci0804_2.
- Chen, S., Epps, J., Ruiz, N., & Chen, F. (2011). Eye activity as a measure of human mental effort in HCI. In *IUI '11: Proceedings of the 16th international conference on intelligent user interfaces* (S. 315–318). Association for Computing Machinery.
- De Cock, M. (2012). Representation use and strategy choice in physics problem solving. *Physical Review Special Topics – Physics Education Research*, 8(2). <https://doi.org/10.1103/PhysRevSTPER.8.020117>.
- De Leone, C. J., & Gire, E. (2006). Is instructional emphasis on the use of non-mathematical representations worth the effort? *AIP Conference Proceedings*, 818(1), 45–48. <https://doi.org/10.1063/1.2177019>.
- Dufresne, R. J., Gerace, W. J., & Leonard, W. J. (1997). Solving physics problems with multiple representations. *The Physics Teacher*, 35(5), 270–275. <https://doi.org/10.1119/1.2344681>.
- Even, R. (1998). Factors involved in linking representations of functions. *Journal of Mathematical Behavior*, 17(1), 105–121. doi: [https://doi.org/10.1016/S0732-3123\(99\)80063-7](https://doi.org/10.1016/S0732-3123(99)80063-7).
- Gire, E., & Price, E. (2012). Graphical representations of vector functions in upper-division E&M. *AIP Conference Proceedings*, 1413(1), 27–30. <https://doi.org/10.1063/1.3679985>.
- Hill, M. J. (2015). *Scientific representational fluency: defining, diagnosing, and developing*. Doctoral dissertation, University of Sydney.
- Jaeger, D., Müller, R., & Franz, T. (2017). Aufgabenschwierigkeit und Cognitive Load. In C. Maurer (Hrsg.), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis* (Bd. 37, S. 91–94). GDGP - Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Zürich 2016. Abgerufen am 1. Februar 2021 von www.gdcp.de/images/tagungsbaende/GDGP_Band37.pdf.
- Just, M. A., & Carpenter, P. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>.
- Klein, P., Viiri, J., & Kuhn, J. (2019). Visual cues improve students' understanding of divergence and curl: Evidence from eye movements during reading and problem solving. *Physical Review Physics Education Research*, 15(1). <https://doi.org/10.1103/PhysRevPhysEducRes.15.010126>.

- Klein, P., Viiri, J., Mozaffari, S., Dengel, A., & Kuhn, J. (2018). Instruction-based clinical Eye Tracking study on the visual interpretation of divergence: How do students look at vector field plots? *Physical Review Physics Education Research*, 14(1). <https://doi.org/10.1103/PhysRevPhysEducRes.14.010116>.
- Knight, R. (1995). The vector knowledge of beginning physics students. *The Physics Teacher*, 33(2), 74–77. <https://doi.org/10.1119/1.2344143>.
- Königsberger, K. (2004). *Analysis 2* (4. Aufl.). Springer.
- Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45, 1058–1072. <https://doi.org/10.3758/s13428-013-0334-1>.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Hrsg.), *The cambridge handbook of multimedia learning* (S. 31–48). University Press.
- Nieminen, P., Savinainen, A., & Viiri, J. (2012). Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning. *Physical Review Special Topics – Physics Education Research*, 8(1). <https://doi.org/10.1103/PhysRevSTPER.8.010123>.
- Niedersächsisches Kultusministerium. (Hrsg.). (2015). *Kerncurriculum für das Gymnasium Schuljahrgänge 5–10. Naturwissenschaften*. Abgerufen am 2. Februar 2021 von https://cuvo.nibis.de/cuvo.php?key_lev0_0=Dokumentenart&svalue_lev0_0=Kerncurriculum&skey_lev0_1=Schulbereich&svalue_lev0_1=Sek+I&skey_lev0_2=Schulform&svalue_lev0_2=Gymnasium-Sek.I&skey_lev0_3=Fach&svalue_lev0_3=Naturwissenschaften&docid=1061&p=detail_view.
- Niedersächsisches Kultusministerium. (Hrsg.). (2018). *Kerncurriculum für das Gymnasium – gymnasiale Oberstufe, die Gesamtschule - gymnasiale Oberstufe, das Fachgymnasium, das Abendgymnasium, das Kolleg. Mathematik*. Abgerufen am 2. Februar 2021 von https://cuvo.nibis.de/cuvo.php?key_lev0_0=Dokumentenart&svalue_lev0_0=Kerncurriculum&skey_lev0_1=Schulbereich&svalue_lev0_1=Sek+II&skey_lev0_2=Schulform&svalue_lev0_2=Gymnasiale+Oberstufe&skey_lev0_3=Fach&svalue_lev0_3=Mathematik&.
- Nistal, A., Dooren, W., & Verschaffel, L. (2012). What counts as a flexible representational choice? An evaluation of students' representational choices to solve linear function problems. *Instructional Science*, 40(6), 999–1019. <https://doi.org/10.1007/s11251-011-9199-9>.
- Papula, L. (2008). *Mathematik für Ingenieure und Naturwissenschaftler Band 3. Vektoranalysis, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Fehler- und Ausgleichsrechnung* (5. Aufl.). Vieweg+Teubner.
- Phillips, M. H., & Edelman, J. A. (2008). The dependence of visual scanning performance on search direction and difficulty. *Vision Research*, 48(21), 2184–2192. <https://doi.org/10.1016/j.visres.2008.06.025>.
- Rau, M. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, 29(4), 717–761. <https://doi.org/10.1007/s10648-016-9365-3>.
- Rakoczi, G. (2012). Eye Tracking in Forschung und Lehre. Möglichkeiten und Grenzen einesvielfersprechenden Erkenntnismittels. In G. Csanyi, F. Reichl, & A. Steiner (Hrsg.), *Digitale Medien - Werkzeuge für exzellente Forschung und Lehre* (Bd. 61, S. 87–98). Waxmann.
- Renkl, A. (1999). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology of Education*, 14, 477–488. <https://doi.org/10.1007/BF03172974>.
- Rosengrant, D., Heuvelen, A. V., & Etkina, E. (2007). An overview of recent research on multiple representations. *AIP Conference Proceedings*, 883(1), 149–152. <https://doi.org/10.1063/1.2508714>.

- Rosengrant, D., Heuvelen, A. V., & Etkina, E. (2009). Do students use and understand free-body diagrams? *Physical Review Special Topics - Physics Education Research*, 5(1). <https://doi.org/10.1103/PhysRevSTPER.5.010108>.
- Scheiter, K., & Eitel, A. (2017). The use of eye tracking as a research and instructional tool in multimedia learning. In K. Scheiter, & A. Eitel (Hrsg.), *Eye-tracking technology applications in educational research* (S. 143–164). IGI Global.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8).
- Schnotz, W., Baadte, C., Müller, A., & Rasch, R. (2010). Creative thinking and problem solving with depictive and descriptive representations. In L. Verschaffel, E. D. Corte, J. Elen, & T. de Jong (Hrsg.), *Use of external representations in reasoning and problem solving* (S. 11–35). Routledge.
- Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, 13(2), 227–237. [https://doi.org/10.1016/S0959-4752\(02\)00022-1](https://doi.org/10.1016/S0959-4752(02)00022-1).
- Singh, C., & Maries, A. (2013). Core graduate courses: A missed learning opportunity? *AIP Conference Proceedings*, 1513(1), 382–385. <https://doi.org/10.1063/1.4789732>.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123–138. <https://doi.org/10.1007/s10648-010-9128-5>.
- Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An Eye Tracking analysis. *Computers & Education*, 58(1), 375–385. <https://doi.org/10.1016/j.compedu.2011.07.012>.
- Van Heuvelen, A. (1991). Learning to think like a physicist: A review of research-based instructional strategies. *American Journal of Physics*, 59(10), 891–897. <https://doi.org/10.1119/1.16667>.
- Van Heuvelen, A., & Zou, X. (2001). Multiple representations of work–energy processes. *American Journal of Physics*, 69(2), 184–194. <https://doi.org/10.1119/1.1286662>.
- Yerushalmy, M. (1991). Student perceptions of aspects of algebraic function using multiple representation software. *Journal of Computer Assisted Learning*, 7(1), 42–57. <https://doi.org/10.1111/j.1365-2729.1991.tb00223.x>.
- Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. In *BELIV '16: Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (S. 8–85). Association for Computing Machinery.

Eye-Tracking in der Mathematik- und Naturwissenschaftsdidaktik

13

Jennifer Dröse, Philipp Neugebauer, Renate Delucchi Danhier
und Barbara Mertins

Inhaltsverzeichnis

13.1	Einleitung und Forschungslage	209
13.2	Theoretischer Hintergrund	211
13.3	Forschungsfragen	215
13.4	Material und Methode	215
13.5	Ergebnisse	220
13.6	Diskussion	222
	Literatur	224

13.1 Einleitung und Forschungslage

Eye-Tracking ermöglicht vertiefte Untersuchungen kognitiver Prozesse durch das Erfassen des visuellen Verhaltens in verschiedenen Messwerten (Strohmaier et al., 2019). Der Lese- und Textverständnisprozess als kognitiver Prozess ist in der (Psycho-)Linguistik bereits systematisch mittels Eye-Trackings untersucht

J. Dröse (✉)

Fachgruppe Didaktik der Mathematik, Universität Paderborn, Paderborn, Deutschland

E-Mail: jennifer.droese@math.uni-paderborn.de

R. D. Danhier · B. Mertins

IdSuL, Technische Universität Dortmund, Dortmund, Deutschland

E-Mail: renate.delucchi@tu-dortmund.de

B. Mertins

E-Mail: barbara.mertins@tu-dortmund.de

P. Neugebauer

IEEM, Technische Universität Dortmund, Dortmund, Deutschland

E-Mail: philipp2.neugebauer@tu-dortmund.de

worden, mit einer in Teilen ausdifferenzierten Forschungslage (Connor et al., 2015; Holmqvist et al., 2011; Rayner & Well, 1996; van der Schoot et al., 2012). In der mathematikdidaktischen Forschung ist die Methode des Eye-Trackings zunehmend für verschiedene Themengebiete und Zielsetzungen genutzt worden, bspw. Lesen von mathematischen Beweisen oder mathematischen Textaufgaben sowie Nutzung von mathematischen Darstellungen (Überblick: Strohmaier et al., 2020). Das Themengebiet der Lese- und Bearbeitungsprozesse von Textaufgaben wurde bereits früh mittels Eye-Trackings untersucht (z. B. Hegarty et al., 1992; Verschaffel et al., 1992). Aufgrund des Erfassens auch unbewusster kognitiver Prozesse durch ihre Entsprechung in Blickbewegungen erwies sich die Eye-Tracking-Methode als geeignet für die Untersuchung des Lesens und Bearbeitens von Textaufgaben sowie des Decodierens und Analysierens von Texten (Holmqvist et al., 2011; Strohmaier et al., 2020).

Das vorliegende Kapitel fokussiert die bisher kaum erforschte Unterscheidung des Bemerkens und Interpretierens syntaktischer Strukturen (Vorarbeiten sind unter „Syntaktische Strukturen in Textaufgaben und ihre Verarbeitung durch Lernende“ in Abschn. 13.2 dargelegt). Diese Verarbeitungsprozesse sind beim Lesen von mathematischen Textaufgaben von Bedeutung. In der vorliegenden Studie bezeichnen „syntaktische Strukturen“ die sprachlichen Repräsentanten von in Textaufgaben enthaltenen operationstragenden Propositionen auf Satzebene (also Relationen zwischen Textelementen, aus denen sich mathematische Operationen ableiten lassen) (Dröse, 2019; Kintsch & van Dijk, 1975). Die Verarbeitung dieser syntaktischen Strukturen lässt sich charakterisieren durch

- die in der Sprachdidaktik fokussierten Prozesse des unbewussten Wahrnehmens oder Bemerkens (bewusstes Wahrnehmen) (Wildemann et al., 2016) und
- den Prozess des Interpretierens von Bedeutungen der syntaktischen Strukturen (Bedeutungsklä rung für die dargestellte Situation basierend auf explizit oder implizit verfügbarem Wissen) als Grundlage einer erfolgreichen Mathematisierung (Brimo et al., 2017).

Bisherige qualitative Analysen der Prozesse lassen Unterschiede beim Bemerkens und Interpretieren syntaktischer Strukturen in Abhängigkeit der Sprachprofile von Lernenden (Mehrsprachigkeit, Lesekompetenz) vermuten (Dröse, 2019). Für die Mathematikleistung und das Lesen von Textaufgaben haben verschiedene Autor*innen bereits einen Einfluss der Sprachprofile von Lernenden aufgezeigt, bspw. einen hohen Zusammenhang von Mathematiktestleistung und Sprachkompetenz/Lesekompetenz bzw. Migrationshintergrund/Zeit des Deutscherwerbs (Verschaffel et al., 2010; Prediger et al., 2015; Abschn. 13.2).

Die bisherigen qualitativen Analysen zum Bemerkens und Interpretieren ermöglichen allerdings nur begrenzte Rückschlüsse auf zugrunde liegende kognitive Prozesse und ihre Unterschiede zwischen Lernenden (Dröse, 2019). Deshalb wurde Eye-Tracking für eine vertiefte Untersuchung gewählt, kombiniert mit der Erfassung der Sprachprofile der Lernenden und einer diesbezüglichen explorativen Untersuchung von Unterschieden. Die Erforschung der kognitiven Prozesse

ermöglicht dann ein tiefergehendes Verständnis des Lesens von Textaufgaben und die Entwicklung von darauf abgestimmten Lehr-Lernarrangements (Dröse, 2019).

In Abschn. 13.2 wird der theoretische Hintergrund zum Bemerken und Interpretieren syntaktischer Strukturen im Prozess des Lesens von Textaufgaben dargestellt, und darauf basierend werden in Abschn. 13.3 Forschungsfragen formuliert. Anschließend werden Material und Methode in Abschn. 13.4 präsentiert. Abschn. 13.5 zeigt die Ergebnisse der explorativen Untersuchung von Unterschieden bei Lernenden mit verschiedenen Sprachprofilen.

13.2 Theoretischer Hintergrund

Syntaktische Strukturen in Textaufgaben und ihre Verarbeitung durch Lernende

Bedeutung syntaktischer Strukturen in Textaufgaben

„Textaufgaben“ bezeichnen im Folgenden „in (ggf. komplexe) Sachsituationen eingekleidete Mathematikaufgaben, die aufgrund einer Problemfrage eindeutig zu bearbeiten sind“ (Dröse, 2019). Zur Bearbeitung ist ein Verständnis der dargestellten Situation und damit verbunden die Konstruktion eines adäquaten Situationsmodells notwendig. Das Situationsmodell ist grundlegend für weitere Bearbeitungsschritte, bspw. die Mathematisierung (Leiss et al., 2019).

Studien haben sowohl allgemein einen Zusammenhang zwischen der Sprachkompetenz von Lernenden und ihrer Mathematikleistung (Prediger et al., 2015; Verschaffel et al., 2010) als auch spezifisch zwischen sprachlichen Textaufgabenmerkmalen und ihrem Schwierigkeitspotenzial aufgezeigt (Leiss et al., 2019). Die identifizierten potenziell schwierigkeitsgenerierenden sprachlichen Textaufgabenmerkmale können auf lexikalischer und syntaktischer Ebene systematisiert werden (z. B. Dyrvold et al., 2015):

1. Lexikalische Komplexität: Wortarten (z. B. Präpositionen, Pronomen), Wortformen (z. B. Komparative, Passiv) oder Wortbildungen (z. B. Komposita)
2. Syntaktische Komplexität: Phrasen (z. B. Nominalphrasen), Satzglieder (z. B. adverbiale Strukturen) oder Satzmuster (z. B. Relativ- und Konditionalsätze).

Während Auswirkungen der Sprachprofile von Lernenden im Mathematikunterricht allgemein und die sprachlichen Textaufgabenmerkmale im Speziellen bereits untersucht wurden, sind Verarbeitungsprozesse sprachlicher Merkmale durch Lernende bisher wenig fokussiert worden (Leiss et al., 2019).

Herausforderungen in der Verarbeitung syntaktischer Strukturen durch Lernende

Die Zweitspracherwerbsforschung hat die Verarbeitung sprachlicher Strukturen bereits vielfach untersucht. Dabei ist besonders die Verarbeitung von Kasusmarkierungen und syntaktischen Reihenfolgen beim Verständnis von Sätzen

Konstrukt „Sprachbewusstheit“ auf Sprachbewusstheit für syntaktische Strukturen eingeschränkt (Dröse, 2019). Syntaktische Sprachbewusstheit zeigt sich im Bemerken und Interpretieren syntaktischer Strukturen mit folgender Charakterisierung (adaptiert und synthetisiert aus Portmann-Tselikas, 2001; Fandrych 2005a, b; Brimo et al., 2017):

- Bemerken (bewusstes Wahrnehmen und Identifizieren): Aufmerksamkeit bewusst auf die Strukturen zu richten, deren Erschließung für die Konstruktion eines adäquaten Situationsmodells notwendig ist. Dieser Prozess kann allerdings (insbesondere bei L1-Sprechenden) zunehmend auch unbewusst erfolgen.
- Interpretieren: Verarbeitung durch Bedeutungsklä rung für die Konstruktion des Situationsmodells einschl. Rückbezug auf explizites oder implizites Wissen über diese Strukturen.

Erste Herausforderungen im Wahrnehmen, Bemerken oder Überlesen syntaktischer Strukturen sind bereits identifiziert worden (Dröse, 2019; Kaulvers et al., 2016). Herausforderungen im Interpretieren dagegen lassen sich zuweilen auf fehlendes explizites oder implizites syntaktisches Wissen, bspw. von Zweitsprachenlernenden, zurückführen (Beese & Gürsoy, 2012).

Bemerken/Wahrnehmen und Interpretieren sind somit theoretisch unterscheidbare Verarbeitungsprozesse, die gleichermaßen bedeutsam für die Verarbeitung sprachlicher Textaufgabenmerkmale sind. Ergänzend zu den Erkenntnissen der Zweitspracherwerbsforschung legen erste qualitative Videoanalysen Unterschiede in diesen Verarbeitungsprozessen für verschiedene Sprachprofile von Lernenden nahe: Lernende mit geringer Lesekompetenz könnten größere Herausforderungen im *Wahrnehmen und Bemerken syntaktischer Strukturen* aufweisen als solche mit hoher Lesekompetenz. Für die *Interpretation syntaktischer Strukturen* ist explizites und implizites sprachliches Wissen notwendig, über das einige multilinguale Lernende durch vom Deutschen abweichende Deklinationssysteme weniger verfügen könnten als monolinguale (Dröse, 2019). Diese ersten Hinweise werden im Rahmen der Eye-Tracking-Studie methodisch operationalisiert und daran anschließend qualitativ exploriert.

Eye-Tracking zur Untersuchung des Textaufgabenleseprozesses

Der Leseprozess von Textaufgaben ist das älteste Anwendungsgebiet der Eye-Tracking-Methode in der Mathematikdidaktik (Strohmaier et al., 2020). Eye-Tracking erfasst das visuelle Verhalten in verschiedenen Messwerten. Die Maße des visuellen Verhaltens werden kategorisch in Fixationen und Sakkaden unterteilt. Der Eye-Mind-Hypothese liegt die Annahme zugrunde, dass das Auge jeweils die Texteinheit fixiert, die zu dem Zeitpunkt vom Gehirn verarbeitet wird. Sie geht davon aus, dass die Fixation und das Verarbeiten der Information im

Gehirn eng miteinander verknüpft sind. Sind bspw. die Fixationszeiten länger oder die Anzahl der Revisits erhöht, weist dies auf verzögertes Lesen und Schwierigkeiten beim kognitiven Verarbeiten dieser Textstellen hin. Die Blickbewegungen können deshalb auch herangezogen werden, um zu untersuchen, wie syntaktische Strukturen verarbeitet werden. In der Linguistik wurden deshalb schon früh die Maße der Verweildauer (*dwelt time*) und Revisits zur Untersuchung der Verarbeitung semantischer und syntaktischer Informationen von deutschen L1- oder L2-Sprechenden beim Lesen genutzt (Bernhardt, 1984). Die Anzahl der Fixationen bzw. Revisits lässt dabei wie beschrieben Rückschlüsse auf die Bedeutsamkeit einer Area of Interest (AOI) für den Lesenden zu, da verstärkte Revisits mit höherer Bedeutungszuschreibung einhergehen (Holmqvist et al., 2011). In der Mathematikdidaktik wurden beide Maße zur Untersuchung des Einflusses sprachlicher Merkmale auf den Textaufgabenleseprozess genutzt (neuere Studien: Strohmaier et al., 2019; ältere Studien: Hegarty et al., 1992; Verschaffel et al., 1992; Rayner & Well, 1996).

Eye-Tracking besitzt somit das Potenzial, zum einen Auswirkungen sprachlicher Strukturen auf den Leseprozess und zum anderen Zusammenhänge zwischen Leseprozessen und erfolgreichen Bearbeitungen zu erfassen. Folgende Merkmale können dabei Einfluss auf den Leseprozess von Textaufgaben haben, die sich in verschiedenen Eye-Tracking-Maßen zeigen:

1. Konsistenz sprachlicher Strukturen: Bei lexikalischen Sprachmitteln, deren Bedeutung inkonsistent zu der intendierten mathematischen Operation ist, zeigt sich ein langsames Lesen mit mehr Sakkaden und Revisits (Hegarty et al., 1992; Verschaffel et al., 1992).
2. Erwartbarkeit sprachlicher Strukturen: Rayner et al. (1996) bspw. zeigten, dass der Leseprozess sprachlicher Mittel durch ihre Erwartbarkeit beeinflusst ist. Wörter, die vom Lesenden erwartet werden, werden oft gar nicht oder nur kurz fixiert.

Für diese Studie wird somit erwartet, dass Bemerken kombiniert mit der Erwartbarkeit syntaktischer Strukturen einen Einfluss auf die Blickbewegungen hat (Hegarty et al., 1992; Rayner & Well, 1996; Verschaffel et al., 1992).

3. Mathematisierungserfolg: Strohmaier et al. (2019) untersuchten die Bearbeitung komplexer Textaufgaben durch Erwachsene und zeigten, dass längere durchschnittliche Fixationen und häufigere Revisits insbesondere bei schwierigeren Textaufgaben auftraten. Dieser Effekt zeigte sich stärker bei nicht erfolgreichen Bearbeitungen als bei erfolgreichen.

Für diese Studie erwarten wir deshalb, dass das Bemerken/Wahrnehmen syntaktischer Strukturen als Voraussetzung des Interpretierens einen Einfluss auf die Blickbewegungen im Zusammenhang mit dem Bearbeitungserfolg hat, der erfolgreiches Interpretieren und Mathematisieren syntaktischer Strukturen voraussetzt.

Aus den zuvor dargestellten empirischen Befunden lassen sich folgende *potenzielle Einflussfaktoren* auf die Untersuchung der zuvor fokussierten Verarbeitungsprozesse mittels Eye-Tracings identifizieren (Strohmaier et al., 2019; Dröse, 2019; Beese & Gürsoy, 2012; Clahsen & Felser, 2006; Schlesewsky et al., 2000; Bader & Menge 1999; Rayner & Well, 1996; Hegarty et al., 1992; Verschaffel et al., 1992):

1. Sprachprofil der Lernenden (Mehrsprachigkeit, Lesekompetenz)
2. Erwartungskonformität syntaktischer Strukturen
3. Mathematisierungserfolg

13.3 Forschungsfragen

Basierend auf den ausgeführten Erkenntnissen wird in dieser Studie die Unterscheidung des unbewussten/bewussten Wahrnehmens (Bemerkens) und Interpretierens bei der Verarbeitung syntaktischer Strukturen untersucht.

Über diese empirische Unterscheidung hinaus, untersucht dieses Kapitel explorativ, inwiefern sich diese Verarbeitungsprozesse zwischen verschiedenen Sprachprofilen der Lernenden (Einflussfaktor 1: Mehrsprachigkeit, Lesekompetenz) unterscheiden. Dabei werden die Einflussfaktoren 2 (Erwartungskonformität) und 3 (Mathematisierungserfolg) als Unterkriterien fokussiert. Die Forschungsfrage wird in zwei Teilfragen untergliedert (jeweils unterschieden nach Erwartungskonformität/Mathematisierungserfolg):

1. Inwiefern unterscheiden sich die Verarbeitungsprozesse zwischen Lernenden mit hoher und geringer Lesekompetenz?
2. Inwiefern unterscheiden sich die Verarbeitungsprozesse zwischen mono- und multilingualen Lernenden?

13.4 Material und Methode

Die vorliegende Studie wurde mit Fünftklässler*innen einer Gesamtschule von einem interdisziplinären Forschungsteam aus Mathematikdidaktiker*innen und Psycholinguistiker*innen durchgeführt. Folgend werden die Zusammensetzung der Stichprobe sowie die technische Ausrüstung, die Gestaltung des Untersuchungsdesigns und der Items ebenso wie die Datenanalyse dargestellt.

Erhebung von Hintergrundmerkmalen

Bei den teilnehmenden Lernenden der Gesamtstichprobe ($N=126$) wurden als Hintergrundmerkmale Alter, Geschlecht sowie Sprachprofil (Mehrsprachigkeit, Lesekompetenz) erhoben:

1. Mehrsprachigkeit: Erhoben über Nennung aller zu Hause gesprochenen Sprachen im Selbstauskunftsfragebogen. Multilinguale Lernende sind jene, die eine andere/weitere Sprache als Deutsch angeben (angegebene Sprachen: Russisch, Polnisch, Bosnisch, Italienisch, Niederländisch, Türkisch).
2. Allgemeine Lesekompetenz: Erhoben über den ELFE 1-6-Test, einen Lesescreeningtest mit einer Retest-Reliabilität von $r=0.93$ und einer Gesamtpunktzahl von 120 Punkten (Lenhard et al., 2009). Die Lernenden werden per Median-Split (bei 60 Punkten) in Lernende mit hoher oder geringer Lesekompetenz eingeteilt.

Stichprobe

Die Gesamtstichprobe besteht aus den 126 Lernenden der sechs Klassen der Jahrgangsstufe 5 einer Gesamtschule. Für die Eye-Tracking-Studie wurden aus jeder Klasse zufällig fünf Lernende ausgewählt (insgesamt 30 Lernende), von denen 27 Lernende im Aufnahmezeitraum für Aufnahmen zur Verfügung standen. Von diesen mussten fünf Lernende aufgrund unzureichender Eye-Tracking-Aufnahmen aus der Stichprobe ausgeschlossen werden (s. unten). Die Studie umfasst somit 22 Lernende, deren Blickbewegungen sicher mittels Eye-Trackings erfasst werden konnten.

Tab. 13.1 zeigt die Stichprobenzusammensetzung bzgl. Alter, Lesekompetenz und Mehrsprachigkeit für die Eye-Tracking-Stichprobe und die Gesamtstichprobe. Die Eye-Tracking-Stichprobe scheint sich somit kaum bzgl. Lesekompetenz und Alter von der Gesamtstichprobe zu unterscheiden.

Eye-Tracking-Ausrüstung

Die Blickbewegungen der Lernenden wurden mit einem Remote Eye-Tracker von SensoMotoric Instruments (SMI RED250) mit einer Samplingrate von 250 Hz monokular (rechtes Auge) aufgenommen. Die Lernenden saßen durchschnittlich 65 cm vom Monitor entfernt (Monitor: LCD-Display, Seitenverhältnis: 16:9, Auflösung: 1920×1080 Pixel, Länge: 53,7 cm, Höhe: 29,7 cm, 24 Zoll). Die Aufnahmen wurden altersbedingt in der Schule mit einem mobilen Computer (für die Experimentleiterin) und einem weiteren Monitor (für die Proband*innen) durchgeführt. Die Blickbewegungen der Lernenden auf dem Monitor wurden ohne zusätzliche Ausrüstung (Kinnstütze etc.) aufgenommen. Die durchschnittliche Aufnahme-Accuracy lag bei 0.4° (SD : 0.15° , Min.: 0.18° , Max.: 0.64°). Der maximal mögliche Messfehler beträgt 0.57 cm auf dem Bildschirm. Hierfür wurden die Entfernung vom Bildschirm und die Bildschirmgröße einbezogen.

Fünf der 27 Lernenden wurden aus der Stichprobe ausgeschlossen: zwei Lernende mit einer Aufnahme-Accuracy jenseits von 1.5° , drei weitere Lernende, weil die Kalibrierung nicht gelang. Die Datenanalyse nutzte die automatische Ereigniserkennung der BeGaze-Software. Der Algorithmus von BeGaze definiert

Fixationen als Verweilen von mindestens 80 ms bei einer maximalen Dispersion von 100 Pixeln zwischen zwei Sakkaden (SensoMotoric Instruments, 2014).

Eye-Tracking-Items

Alle Lernenden lasen acht Textaufgabenitems auf dem Monitor. Die Lernenden bearbeiteten und lösten die Aufgaben in Einzelarbeit im Kopf und nannten anschließend die Antwort auf die gestellte Frage mündlich. Die Items wurden zuvor für eine Bearbeitung ohne schriftliche Notizen optimiert. Die Items wurden den Lernenden aus zwei Listen mit zufälliger Itemreihenfolge randomisiert zugewiesen. Zwischen jeweils zwei Textaufgaben wurden Wimmelbilder mit versteckten Tieren als Filler bzw. Distraktoren eingesetzt, um den Teilnehmenden Abwechslung zu bieten, die Motivation hochzuhalten und Lerneffekte zu vermeiden.

Da die deutsche Sprache eine Veränderung der Satzgliedstruktur ermöglicht, werden die Items diesbezüglich variiert. Diese Variation hat sich als geeignete Erhebungsmethode für unterschiedlichen Aufmerksamkeitsfokussierungen erwiesen (weitere Ausführungen zur Variationstheorie: Dröse, 2019): Variation 1 – Satzgliedstellung: erwartet (Subjekt – Objekt) und unerwartet (Objekt – Subjekt) (Rayner & Well, 1996) (Abb. 13.2, links).

Um Einflüsse weiterer Merkmale der Textaufgaben und Pronomen auszu-schließen, wurden diese basierend auf Erkenntnissen der Mathematikdidaktik und Linguistik darüber hinaus systematisch entsprechend folgender Merkmale variiert: Variation 2 – Genus der Pronomen: weiblich und männlich (Eilers et al., 2019), Variation 3 – Schrittigkeit: ein- und zweischrittig (Strohmaier et al., 2019). In der vorliegenden Studie wird im Rahmen der Forschungsfragen als Charakteristikum der Textaufgaben nur die Variation 1, d. h. die Satzgliedstellung, untersucht.

Design der Areas of Interest

Die Textaufgaben wurden aufgrund der Höhe der Messgenauigkeit in der Bildschirmitte mit großer Schrift präsentiert. Vor der Datenerhebung wurden Areas of Interest (AOIs) in den variierten Sätzen über das Subjekt (im Nominativ) und das indirekte Objekt (im Dativ) gelegt. Dabei wurden die AOIs nur für die Pronomen (Personal- oder Possessivpronomen) eng anliegend definiert (in Abb. 13.2

Tab. 13.1 Stichprobenzusammensetzung hinsichtlich der Hintergrundmerkmale

	Allg. Lesekompetenz (max. 120) <i>M (SD)</i>	Alter in Jahren <i>M (SD)</i>	Mehrsprachigkeit: mono-/multilingual
Gesamtstichprobe (<i>N</i> = 126)	60.03 (13.51)	10.90 (0.72)	53 %/47 %
Eye-Tracking- Stichprobe (<i>n</i> = 22)	59.45 (13.77)	10.95 (0.72)	64 %/36 %

durch Fettdruck markiert). Darüber hinaus waren die Pronomen beim Itemdesign von anderen für die Aufgabenlösung relevanten Informationen (z. B. numerischen Daten) im Text bspw. durch Adverbien getrennt worden. Obwohl die Pronomen verschiedene Längen haben, sind sie kurze Wörter mit einer geringen Buchstabenvariabilität (Variabilität: drei bis sechs Buchstaben), sodass leichte Abweichungen der AOI-Größe zu vernachlässigen sind. Die AOI-Größe variierte zwischen 6612 und 17.139 Pixel, somit 0.3 % (1.6 cm) bis 0.7 % (3.8 cm) des Bildschirms. Abb. 13.2 (rechts) zeigt ein Beispiel für die Platzierung der AOIs.

Nach der Datenerhebung wurden die Eye-Tracking-Daten qualitativ visuell dahingehend überprüft, dass die Blickbewegungen durch die AOIs erfasst wurden. Die Genauigkeit (*accuracy*) sowie die Kalibrierung wurden ebenfalls manuell überprüft, um ggf. eine Abdriftkorrektur (*drift correction*) durchzuführen. Eine Korrektur war jedoch für die Proband*innen nicht notwendig, da ihre Fixationen mit den Textlinien optimal aligniert waren und somit durch die AOIs erfasst wurden.

Methoden der Datenanalyse

Als methodische Herausforderung war zu klären, wie die beiden Verarbeitungsprozesse mittels Eye-Trackings erfasst werden können. Dazu wurde für das Bemerkens und Wahrnehmens die Fixation auf die AOI betrachtet. Diese wurde über die Verweildauer in Form der *net dwell time* (NDT) und die Revisits operationalisiert. Die NDT wurde der einfachen *dwell time* vorgezogen, um die Betrachtungszeit einer AOI über die gesamte Zeit des Trials zu erfassen, inklusive der für die Analysen relevanten Revisits. Die NDT und die Revisits gehören zu gängigen Maßen zur Untersuchung der Verarbeitung syntaktischer Strukturen in der Linguistik (Bernhardt, 1984).

Tab. 13.2 gibt einen Überblick über die durchschnittliche Revisit-Anzahl pro AOI pro Aufgabenbearbeitung für unterschiedliche Erwartbarkeiten der Strukturen und in Bezug auf den Mathematisierungserfolg.

Eine Revisit-Anzahl unter 1.00 zeigt, dass die AOI nach dem ersten Lesen nicht in allen Bearbeitungen erneut fokussiert wird. Eine Revisit-Anzahl über 1.00 zeigt eine erneute Fixierung. Das Bemerkens lässt sich somit wie folgt mittels

Einschrittige Textaufgabe	Zweistrittige Textaufgabe	
Emma und Jonas sind Freunde Emma hat 20 € und Jonas hat 15 €. (Hier Satz A, B, C oder D einfügen)	Emma und Jonas sind Freunde Emma hat 20 € und Jonas hat 15 €. Emma bekommt von ihren Eltern 10 €. (Hier Satz A, B, C oder D einfügen)	Emma und Jonas sind Freunde. Emma hat 20 € und Jonas hat 15 €, Seine Freundin Emma gibt ihm gerne 5 €. Wie viel Geld hat Jonas?
Wie viel Geld hat Jonas?	Wie viel Geld hat Jonas?	
Seine Freundin Emma gibt ihm gerne 5 €.	Ihr Freund Jonas gibt ihr gerne 5 €.	
Seiner Freundin Emma gibt er gerne 5 €.	Ihrem Freund Jonas gibt sie gerne 5 €.	

Abb. 13.2 Links: Stimuli mit hervorgehobenen AOIs (aus Dröse et al., 2021). Rechts: Platzierung der AOIs auf dem Bildschirm (AOIs wurden für Lernende nicht gekennzeichnet)

Tab. 13.2 Durchschnittliche Revisits pro AOI pro Aufgabebearbeitung

	Durchschnittliche Revisits pro AOI							
Items	erwartungskonform				nicht erwartungskonform			
Mathematisierung	erfolgreich		nicht erfolgreich		erfolgreich		nicht erfolgreich	
AOI	Subjekt	Objekt	Subjekt	Objekt	Subjekt	Objekt	Subjekt	Objekt
Lesekompetenz								
• Lesestärke (n = 11)	0.66	0.47	0.90	1.00	0.54	1.04	0.93	0.53
• Leseschwache (n = 11)	1.41	1.59	1.13	0.25	0.88	2.15	0.82	1.41
Mehrsprachigkeit								
• Monolinguale (n = 15)	0.85	0.83	0.81	0.66	0.68	1.53	0.81	0.95
• Multilinguale (n = 7)	1.30	1.27	1.40	0.13	0.79	2.00	0.81	1.10

Eye-Trackings identifizieren: Da Strukturen mehr Bedeutung beigemessen wird, je häufiger sie fixiert werden (Holmqvist et al., 2011), zeigt eine Revisit-Anzahl über 1.00 ein bewusstes erneutes Lesen nach dem ersten Leseprozess des gesamten Textes oder der syntaktischen Struktur alleine und deutet auf ein *erfolgreiches Bemerken* hin. Eine Revisit-Anzahl unter 1.00 zeigt, dass die AOI nach dem ersten Lesen nicht in allen Bearbeitungen erneut fokussiert wird. Dies kann entweder auf ein *nicht erfolgreiches Bemerken* oder ein *nicht bewusstes Bemerken* (im Fall einer direkten Interpretation) hindeuten. Das *erfolgreiche Bemerken* kann somit durch Revisit-Anzahlen über 1.00 operationalisiert werden. Für Revisit-Anzahlen unter 1.00 ist eine Betrachtung der NDT relevant, denn durch NDT kombiniert mit Revisit-Anzahl können sowohl die „frühere“ Auseinandersetzung mit dem Text als auch die „spätere“ Textbearbeitung untersucht werden.

Zur Unterscheidung von nicht erfolgreichem und unbewusstem Bemerken ist außerdem die Betrachtung der Interpretation notwendig. Dafür wurden die mündlichen Lernendenantworten hinsichtlich Mathematisierungserfolg des variierten Satzes ausgewertet, d. h., ob die Struktur des variierten Satzes korrekt in eine mathematische Operation überführt wurde. Andere Mathematisierungsfehler wurden nicht berücksichtigt. Eine *erfolgreiche Interpretation* zeigt sich in einer erfolgreichen Mathematisierung (Dröse, 2019). Eine nicht erfolgreiche Mathematisierung lässt sich auf eine *nicht erfolgreiche Interpretation* bei erfolgreichem Bemerken oder bedingt durch nicht erfolgreiches Bemerken zurückführen.

Die gleichzeitige Betrachtung von Eye-Tracking-Merkmalen und mündlichen Antworten der Lernenden ermöglicht eine gezielte Unterscheidung des Bemerkens und Interpretierens in der Operationalisierung und die Rekonstruktion der in Abb. 13.3 dargestellten *Schwerpunkte der Verarbeitungsprozesse* syntaktischer Strukturen.

Mathematisierung \ Revisits (NDT)	Anzahl = 0 (kurz)		Anzahl ≥ 1 (lang)	
	Anzahl = 0 (kurz)		Anzahl ≥ 1 (lang)	
Nicht erfolgreich	1) Nicht erfolgreiches Bemerken (keine Wahrnehmung)		2) Erfolgreiches Bemerken (bewusste Wahrnehmung) und nicht erfolgreiches Interpretieren	
	3) Erfolgreiches Bemerken (keine bewusste Wahrnehmung) und erfolgreiches Interpretieren		4) Erfolgreiches Bemerken (bewusste Wahrnehmung) und erfolgreiches Interpretieren	

Abb. 13.3 Rekonstruierte Schwerpunkte der Verarbeitungsprozesse

Methoden der statistischen Datenauswertung

Zur Überprüfung des Einflusses der Lesekompetenz der Lernenden auf die NDT wurden die Lernenden normativ am Median ihrer Testergebnisse getrennt. Die ungepaarte Stichprobe der lesestarken und -schwachen Lernenden liefert metrische Ergebnisse im Hinblick auf den Anteil der korrekt bearbeiteten Aufgaben. Die Teilstichproben der lesestarken und -schwachen Lernenden sind normalverteilt und weisen vergleichbare Varianzen auf. Dies ermöglichte, die Teilgruppen mittels eines t-Tests für unabhängige Stichproben zu vergleichen, da damit alle Voraussetzungen zur Durchführung von t-Tests erfüllt sind. Für die Mehrsprachigkeit (Unterscheidung in mono- und multilinguale Lernende) gelten die gleichen Überlegungen, weshalb auch hier die Bedingungen zur Beschreibung der Unterschiede mittels eines t-Tests für unabhängige Stichproben erfüllt sind.

Die statistische Auswertung erfolgte mit R durch das Package Stats. Da die beschriebenen Lernenden aufgrund ihrer gemeinsamen Klassenzugehörigkeit nicht als einfache Stichprobe angesehen werden können, ist jeder Vergleich nur deskriptiv zu sehen und liefert nicht die nötige Teststärke. Die t-Tests sollen daher nur als Indikatoren gesetzt sein, die Hinweise liefern, wo bei hinreichender Testung weitere Beobachtungen bestätigt werden könnten.

13.5 Ergebnisse

In Bezug auf die Forschungsfragen zeigen die statistischen Tests (Tab. 13.3) keine signifikanten Gruppenunterschiede. Deshalb werden im Folgenden für eine explorative Untersuchung die absolute NDT (in Millisekunden) und die relative NDT (als Anteil der Gesamtlesezeit) im Vergleich der Untergruppen betrachtet.

Dabei werden die Lesekompetenz und die Mehrsprachigkeit als relevante Merkmale fokussiert (Tab. 13.3, erste Spalte). Da zudem auch Erwartungskonformität und Mathematisierungserfolg als Einflussfaktoren identifiziert wurden, werden sie als Kriterien mit aufgeführt. Aufgrund der Kasusmarkierung in den Pronomen, die zur Verarbeitung erfasst werden müssen, werden diese als weiteres Unterscheidungskriterium aufgegriffen (Tab. 13.3, Zeile 3).

Hinsichtlich Forschungsfrage 1 zeigen sich folgende Tendenzen: Leseschwache Lernende weisen weitestgehend eine höhere absolute und relative NDT als lesestarke Lernende auf. Eine Ausnahme bilden Objektpronomen bei nicht erfolgreich mathematisierten erwartungskonformen Items. Im Verhältnis der Wahrnehmung von Subjekt und Objekt im Lesekompetenzvergleich gibt es kaum Hinweise auf Unterschiede in der NDT unabhängig von *Erwartungskonformität* und *Mathematisierungserfolg*. Unterschiede zeigen sich nur bei *nicht erfolgreich mathematisierten erwartungskonformen Items*. Leseschwache Lernende fokussieren dabei tendenziell geringer auf die Objekt-AOI (Lesestärke: 484.26 ms/1.6 %, Leseschwache: 263.22 ms/1.0 %). Die geringere NDT kombiniert mit der nicht erfolgreichen Mathematisierung könnte gemäß Operationalisierung auf ein nicht erfolgreiches Bemerken hindeuten (Abb. 13.3).

Hinsichtlich Forschungsfrage 2 zeigen sich im Vergleich der Mehrsprachigkeit bei den *erwartungskonformen Items* für die *absolute* NDT ähnliche Tendenzen wie für den Vergleich nach Lesekompetenz. Hingegen gibt es Hinweise, dass sich monolinguale und multilinguale Lernende in der *relativen* NDT bei erwartungskonformen Items nicht in gleichem Maße wie Lesestärke und Leseschwache unterscheiden (Erwartungskonforme Items: monolinguale: 1.5–1.7 %, Multilinguale: 1.5–1.9 %). Deutlichere Unterschiede zwischen mono- und multilingualen Lernenden zeigen sich jedoch bei *nicht erwartungskonformen Items*: Monolinguale zeigen unabhängig vom Mathematisierungserfolg eine relative NDT von 1.8–2.0 %. Multilinguale hingegen zeigen bei *erfolgreich mathematisierten nicht erwartungskonformen Items* eine höhere NDT auf der Objekt-AOI. Eine längere NDT kombiniert mit einer erfolgreichen Mathematisierung deutet nach Operationalisierung auf ein erfolgreiches Bemerken sowie Interpretieren hin (Abb. 13.3). Währenddessen zeigt sich bei Multilingualen bei *nicht erfolgreich mathematisierten nicht erwartungskonformen Items* auf beiden AOIs eine ähnliche (geringere als bei Monolingualen) relative NDT. Die geringere NDT kombiniert mit der nicht erfolgreichen Mathematisierung könnte auf ein nicht erfolgreiches Bemerken hindeuten (Abb. 13.3).

Da sich ansonsten kaum Tendenzen bzgl. Unterschieden in der NDT im Vergleich (*nicht*) *erfolgreich mathematisierter erwartungskonformer Items* sowohl bei monolingualen als auch bei lesestarken Lernenden zeigen, könnte dies ein Hinweis auf eine mangelnde bewusste Steuerung des Bemerkens sein (Abb. 13.3).

13.6 Diskussion

Für die Eye-Tracking-Methode ist das Lesen von Textaufgaben ein klassisches Themengebiet in der mathematikdidaktischen Forschung (Strohmaier et al., 2020). Dieses Kapitel erweitert die Untersuchungen über die Auswirkungen von Aufgabenmerkmalen (Hegarty et al., 1992; Strohmaier et al., 2019; Verschaffel et al., 1992) auf die *Verarbeitungsprozesse* zu diesen Aufgabenmerkmalen, hier syntaktischen Strukturen. Diese sind bereits vielfach als Herausforderungen

dargestellt, aber in ihren kognitiven Prozessen selten untersucht worden (Leiss et al., 2019).

Zusammenfassung der Ergebnisse

Durch Eye-Tracking-Maße (Revisits, NDT) kombiniert mit codierten mündlichen Antworten der Lernenden konnten Bemerken und Interpretieren unterschieden und vier verschiedene Schwerpunkte in der Verarbeitung syntaktischer Strukturen rekonstruiert werden:

1. Nicht erfolgreiches Bemerken (keine Wahrnehmung)
2. Erfolgreiches Bemerken (bewusste Wahrnehmung) und nicht erfolgreiches Interpretieren
3. Erfolgreiches Bemerken (keine bewusste Wahrnehmung) und erfolgreiches Interpretieren
4. Erfolgreiches Bemerken (bewusste Wahrnehmung) und erfolgreiches Interpretieren

Bei der anschließenden explorativen Untersuchung von Gruppenunterschieden zeigen sich interessante erste Tendenzen von Unterschieden:

Im Vergleich der Lesekompetenzen zeigen leseschwache Lernende bei nicht erfolgreich mathematisierten Items mit erwartungskonformen syntaktischen Strukturen eine geringere Verweildauer als lesestarke Lernende. Dies könnte auf ein nicht erfolgreiches Bemerken der syntaktischen Strukturen hindeuten. Im Vergleich der Mehrsprachigkeit werden Items mit erwartungswidrigen syntaktischen Strukturen von multilingualen Lernenden bei erfolgreicher Mathematisierung länger fokussiert, bei nicht erfolgreicher Mathematisierung jedoch kürzer als von den monolingualen Lernenden. Dies könnte im ersten Fall auf ein erfolgreiches Bemerken der syntaktischen Strukturen und im zweiten Fall auf ein nicht erfolgreiches Bemerken hindeuten.

Diese Tendenzen sollten in Folgeuntersuchungen durch größere Fallzahlen bestätigt und die Veränderung des Bemerkens und Interpretierens in Interventionsstudien untersucht werden (für eine weiterführende Diskussion s. Dröse et al., 2021).

Grenzen der Studie und Ausblick

Die geringe Anzahl an teilnehmenden Lernenden und bearbeiteten Items begrenzen die Studienergebnisse, sodass nur erste Tendenzen berichtet werden können. Zukünftige Studien sollten die Untersuchungen auf folgende Aspekte ausweiten:

1. Größere Lernendenanzahl für statistische Tests zur Untersuchung von Gruppenunterschieden

2. Hintergrundmerkmale der Lernenden (z. B. sozioökonomischer Status, Jahrgangsstufe) zur Erfassung ihrer Auswirkungen
3. Itemanzahl und betrachtete syntaktische Strukturen zur Untersuchung der Unterschiede und Gemeinsamkeiten in Verarbeitungsprozessen

Eine Ausweitung der Untersuchung sollte eine Diskussion über die praktischen Implikationen für den Unterricht sowie die Aus- und Fortbildung von Lehrkräften ermöglichen.

Danksagung Diese Studie wurde durchgeführt in Kooperation des Instituts für Entwicklung und Erforschung des Mathematikunterrichts gemeinsam mit Prof. Dr. Susanne Prediger und der AG Diversitätsstudien – Institut für deutsche Sprache und Literatur der TU Dortmund. Wir danken Prof. Dr. Susanne Prediger für die Zusammenarbeit und allen beteiligten Lehrkräften sowie den Schüler*innen und Schülern für ihre Unterstützung.

Literatur

- Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research*, 28(2), 121–143.
- Beese, M., & Gürsoy, E. (2012). Bezüge im Deutschen und im Türkischen herstellen. Sprachliche Stolpersteine beim Mathematiklernen für zweisprachige Lernende. *Praxis der Mathematik in der Schule*, 54(45), 34–37.
- Bernhardt, E. B. (1984). *Text processing strategies of native, non-native experienced, and non-native inexperienced readers of German: Findings and implications for the instruction of German as a foreign language*. University of Minnesota.
- Brimo, D., Apel, K., & Fountain, T. (2017). Examining the contributions of syntactic awareness and syntactic knowledge to reading comprehension. *Journal of Research in Reading*, 40(1), 57–74.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27(1), 3–42.
- Connor, C. M., Radach, R., Vorstius, C., Day, S. L., Mclean, L., Morrison, F. J., & McDonald, C. (2015). Individual differences in fifth graders' literacy and academic language predict comprehension monitoring development: An eye-movement study. *Scientific Studies of Reading*, 19(2), 114–134.
- Dröse, J. (2019). *Textaufgaben lesen und verstehen lernen. Entwicklungsforschungsstudie zur mathematikspezifischen Leseverständnisförderung*. Springer Spektrum.
- Dröse, J., Prediger, S., & Marcus, A. (2018). *Strategien zum Verstehen von Textaufgaben – Fach- und sprachintegriertes Unterrichtsmaterial in Basis- und Regelfassung*. Open educational resources on simadzl.de/um.
- Dröse, J., Prediger, S., Neugebauer, P., Delucchi Danhier, R. D., & Mertins, B. (2021). Investigating students' processes of noticing and interpreting syntactic language features in word problem solving through eye-tracking. *International Electronic Journal of Mathematics Education*, 16(1), em0625, 1–17.
- Dyrvold, A., Bergqvist, E., & Österholm, M. (2015). Uncommon vocabulary in mathematical tasks in relation to demand of reading ability and solution frequency. *Nordic Studies in Mathematics Education*, 20(1), 5–31.
- Eilers, S., Tiffin-Richards, P., & Schroeder, S. (2019). Gender Cue Effects in Children's Pronoun Processing: A Longitudinal Eye Tracking Study. *Scientific Studies of Reading*, 23(6), 509–522.
- Fandrych, C. (2005a). Ordnung und Variation in Satz und Text. Wortstellung entdecken, erkunden, erproben. *Fremdsprache Deutsch*, 32, 5–11.

- Fandrych, C. (2005b). Schauplatz: Wortstellung – Wichtige Regeln und Tendenzen. *Fremdsprache Deutsch*, 32, 12–18.
- García, O. (2017). Critical multilingual language awareness and teacher education. In J. Cenoz, S. May, & D. Gorter (Hrsg.), *Language Awareness and Multilingualism* (S. 263–280). Springer.
- Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of Arithmetic Word Problems: Evidence From Students' Eye Fixations. *Journal of Educational Psychology*, 84(1), 76–84.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Kaulvers, J., Schlager, S., Isselbacher-Giese, A., & Klein, M. (2016). Entwickeln, Beraten, Unterstützen – QUA-LIS: Sprachliche Hürden in Mathematikaufgaben. *Schule NRW*, 07(08), 16–19.
- Kintsch, W., & van Dijk, T. A. (1975). Towards a model of text comprehension and production. *Psychological Review*, 85(5), 363–394.
- Leiss, D., Plath, J., & Schwiipert, K. (2019). Language and Mathematics – Key Factors influencing the Comprehension Process in reality-based Tasks. *Mathematical Thinking and Learning*, 21(2), 131–153.
- Lenhard, W., Lenhard, A., & Schneider, W. (2009). Diagnose und Förderung von Leseverständnis mit ELFE 1–6 und ELFE-Training. In W. Lenhard & W. Schneider (Hrsg.), *Diagnose und Förderung des Leseverständnisses* (S. 79–112). Hogrefe.
- Nortvedt, G. A. (2011). Coping strategies applied to comprehend multistep arithmetic word problems by students with above-average numeracy skills and below-average reading skills. *The Journal of Mathematical Behavior*, 30(3), 255–269.
- Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E., & Benholz, C. (2015). Sprachkompetenz und Mathematikleistung – Empirische Untersuchung sprachlich bedingter Hürden in den Zentralen Prüfungen 10. *Journal für Mathematik-Didaktik*, 36(1), 77–104.
- Portmann-Tselikas, P. R. (2001). Sprachaufmerksamkeit und Grammatiklernen. In P. R. Portmann-Tselikas & S. Schmölzer-Eibinger (Hrsg.), *Grammatik und Sprachaufmerksamkeit* (S. 9–48). Studien Verlag.
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504–509.
- Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). Preferences for grammatical functions in the processing of locally ambiguous wh-questions in German. In B. Hemforth & L. Konieczny (Hrsg.), *German sentence processing* (S. 65–94). Kluwer Academic Publishers.
- SensoMotoric Instruments. (2014). BeGaze Manual. *Version*, 3, 4.
- Strohmaier, A. R., Lehner, M. C., Beitlich, J. T., & Reiss, K. M. (2019). Eye movements during mathematical word problem solving – global measures and individual differences. *Journal für Mathematik-Didaktik*, 40(2), 255–287.
- Strohmaier, A. R., Mac Kay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, 104, 147–200.
- Verschaffel, L., de Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's Consistency Hypothesis. *Journal of Educational Psychology*, 84(1), 85–94.
- Verschaffel, L., van Dooren, W., Greer, B., & Mukhopadhyay, S. (2010). Reconceptualising Word Problems as Exercises in Mathematical Modelling. *Journal für Mathematik-Didaktik*, 31(1), 9–29.
- van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, 25(7), 1665–1690.
- Wildemann, A., Akbulut, M., & Bien-Miller, L. (2016). Mehrsprachige Sprachbewusstheit zum Ende der Grundschulzeit. Vorstellung und Diskussion eines Elizitationsverfahrens. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 21(2), 42–56.

Blickbewegungen als Indikator für den Lernerfolg? – Nachvollziehen und Vergleichen beim Lösungsvergleich

14

Antje Boomgaarden, Timo Leuders und Katharina Loibl

Inhaltsverzeichnis

14.1	Einleitung	227
14.2	Theoretischer Hintergrund	229
14.3	Forschungsfragen	231
14.4	Material und Methode	231
14.5	Ergebnisse	234
14.6	Diskussion	241
	Literatur	242

14.1 Einleitung

Zu den zentralen kognitiven Prozessen beim Wissenserwerb und beim Problemlösen gehört das Vergleichen von Beispielen oder Fällen (Alfieri et al., 2013). Viele solcher Vergleichsprozesse wurden im Zusammenhang mit dem Erwerb mathematischen Wissens untersucht (Rittle-Johnson & Star, 2011), in der Regel durch experimentelle Designs, bei denen ein Vergleich von richtigen und falschen

A. Boomgaarden (✉) · T. Leuders · K. Loibl

Institut für Mathematische Bildung, Pädagogische Hochschule Freiburg, Freiburg,
Deutschland

E-Mail: antje.boomgaarden@ph-freiburg.de

T. Leuders

E-Mail: leuders@ph-freiburg.de

K. Loibl

E-Mail: katharina.loibl@ph-freiburg.de

Lösungen einen größeren Lernerfolg nach sich zog als die Bearbeitung nur eines Lösungstyps (z. B. Durkin & Rittle-Johnson, 2012; Siegler, 2002).

Dieses indirekte Vorgehen der Untersuchung von Vergleichsprozessen durch experimentelle Variation ist notwendig, da die Vergleichsprozesse meist schnell und wenig bewusst ablaufen. Hier stellt sich die Frage, inwieweit durch die Analyse von Blickbewegungen der Prozess des Vergleichens direkter empirisch abgebildet werden kann, um so unterschiedliche Formen des Vergleichens zu erkennen und diese mit dem Lernzuwachs in Beziehung zu setzen. Strohmaier und Kollegen (2020) beschrieben den Mehrwert des Eye-Tracking-Einsatzes u. a. dadurch, dass Prozesse in den Blick genommen werden können. In der hier vorgestellten Studie wird eine solche Analyse von Blickbewegungen bei Vergleichsprozessen durchgeführt, und damit werden die Optionen für die Weiterentwicklung der Lehr-Lernforschung zum Vergleichen ausgelotet.

Die hier untersuchten Vergleichsprozesse sind ein zentraler Lernschritt innerhalb eines umfassenderen Instruktionsdesigns mit einer anfänglichen Problemlösephase, gefolgt von einer Instruktionsphase (PS-I; Loibl & Leuders, 2019; Loibl & Rummel, 2014). Folgende Problemlöseaufgabe zum Nacherfinden relativer Häufigkeiten bildet das Kernstück unseres mehrphasigen Lernszenarios:

Eine Gruppe von fünf Mädchen tritt gegen eine zehnköpfige Jungengruppe im Papierkorbball an. Jedes Kind wirft einmal. Die Mädchen treffen dreimal, die Jungen fünfmal. Vergleiche fair: Welche Gruppe hat gewonnen? (Prediger, 2011; Prediger et al., 2013)

Die Lernenden werden aufgefordert, ihre Entscheidung mit Bruchstreifen zu begründen. Die individuellen Lösungen dieser Aufgabe, die Lernende in der Problemlösephase generieren, sind in der Regel unvollständig und fehlerhaft. Sie können aber als Grundlage für einen Vergleich mit der normativen (d. h. mathematisch korrekten) Lösung in der Instruktionsphase genutzt werden. Viele Lernende fokussieren bei der Bearbeitung dieser Problemlöseaufgabe bspw. ausschließlich die Anzahl der Treffer und lassen die Größe der Mannschaft außer Acht. Das führt zu der in Abb. 14.1 dargestellten Vergleichssituation der individuellen (fehlerhaften) Schülerlösung mit der normativen Lösung.

Doch wann führt ein solcher Vergleich zu einem Lernzuwachs? Durch eine Kombination aus Blickbewegungs- und Interviewdaten untersuchen wir, inwieweit sich ein erfolgreicher Fehlerverarbeitungsprozess von einem weniger erfolgreichen Fehlerverarbeitungsprozess in der Abfolge von Teilprozessen beim Vergleichen unterscheidet.

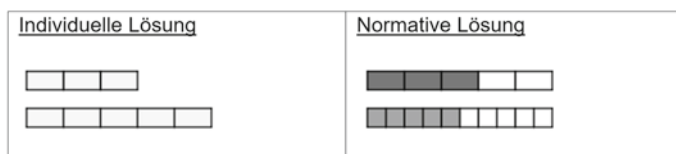


Abb. 14.1 Beispielhafte Gegenüberstellung einer individuellen Lösung und der normativen Lösung repräsentiert durch Bruchstreifen.

14.2 Theoretischer Hintergrund

Der Vergleich von Lösungen ist ein bedeutsamer Lernmechanismus, und sowohl Forscherinnen und Forscher aus den Kognitionswissenschaften als auch aus der Mathematikdidaktik betonen die Bedeutung dieses fundamentalen kognitiven Prozesses für das Lernen und den Transfer (z. B. Ball, 1993; Gick & Holyoak, 1983). Nach Gentner (2003, 2010) kann Vergleichen als Prozess aufgefasst werden, bei dem Beispiele oder Fälle in Beziehung gesetzt und Gemeinsamkeiten identifiziert werden. In PS-I-Lernumgebungen (z. B. Loibl & Leuders, 2019; Loibl & Rummel, 2014) stellen Vergleichsprozesse einen zentralen Lernschritt dar: Die Lösungen, die Lernende in der Problemlösephase generieren, werden in der anschließenden Instruktionsphase für einen Vergleich mit der normativen Lösung genutzt. Nach Rittle-Johnson und Star (2011) liegt der Fokus beim Vergleichen von normativen und fehlerhaften Lösungen auf der Frage, warum eine Lösung richtig ist und eine andere nicht. Dazu bedarf es nicht nur vergleichender Prozesse, sondern auch Prozesse des Nachvollziehens der normativen Lösung sowie des Nachvollziehens der fehlerhaften Lösung. Es ist anzunehmen, dass Letzteres Lernenden leichter fällt, wenn die dafür verwendete fehlerhafte Lösung der jeweils eigenen Lösung, z. B. aus einer vorangegangenen Problemlösephase, entspricht. Im Rahmen eines solchen Vergleichs der individuellen Lösung mit der normativen Lösung werden die Lernenden mit den Grenzen und Schwächen des eigenen mentalen Modells konfrontiert (Chi, 2000; Gadgil et al., 2012; Posner et al., 1982). Die Lernenden können Unterschiede zwischen den Lösungen erkennen, was eine Fehlerverarbeitung und eine Umstrukturierung des eigenen Wissens im Sinne eines Konzeptwechsels ermöglicht (Gadgil et al., 2012). Weiter können die Lernenden im Zuge solcher Vergleiche aber auch Übereinstimmungen der Lösungen erkennen, was den Aufbau abstrahierter Schemata begünstigt (Gentner et al., 2003).

Die Aufgabe aus dem Eingangsbeispiel wurde bereits von Loibl und Leuders (2018) mit Bezug auf die Bedeutung der Fehlerverarbeitung untersucht. Sie zeigten, dass die Entscheidung der Schüler*innen und Schüler, welche Mannschaft fairerweise gewonnen hat, Lösungen auf verschiedenen Stufen des konzeptuellen Verständnisses des Anteilskonzepts hervorbringt. Die Stufen bilden typische Präkonzepte im Sinne vorunterrichtlicher Erfahrungen und Vorstellungen ab. Einige Lernende (vgl. Eingangsbeispiel) interpretierten Bruchzahlen im Sinne natürlicher Zahlen, indem nur eine Komponente, nämlich der Zähler (Teil) fokussiert wird, statt das Verhältnis beider Komponenten zu berücksichtigen. Andere Lernende argumentieren ausschließlich über die Größe der Mannschaften und fokussieren damit auch nur eine Komponente: den Nenner (Ganzes). Dieses Phänomen ist bekannt unter dem Namen „natural number bias“ (Ni & Zhou, 2005). Die von Loibl und Leuders (2018) identifizierten Lösungstypen lassen sich basierend auf der SOLO-Taxonomie („Structure of Observed Learning Outcomes“: Biggs & Collis, 1982) als Stufen der konzeptuellen Integration und Abstraktion anordnen (Tab. 14.1).

Tab. 14.1 Stufen des konzeptuellen Verständnisses bei der Erarbeitung des Anteilskonzepts

Stufe	SOLO	Bruchkonzept bzw. Präkonzept	Kognitionen	Indikative Lösungen
1	Unistruktuell	Natürliche-Zahl-Konzept (<i>natural number bias</i>)	Fokus auf eine Größe: 1a) Teile <i>oder</i> 1b) Ganzes <i>oder</i> 1c) Rest	
2	Multistruktuell	Teil-Ganzes-Konzept	Erkennen der Abhängigkeit von zwei Größen: Teil <i>und</i> Ganzes	
3	Relational	Basales Anteils-konzept	3a) Ganzes als gleich ansehen (Verfeinern), 3b) Situation vergleichbar machen (Vergrößern)	
4	Erweitert abstrakt	Flexibles Anteils-konzept	Komplexere Schemata: gemeinsames Verfeinern und Vergrößern (nicht-symbolisches Erweitern)	

Rittle-Johnson und Star (2009) berichten, dass die allgemeine kognitionspsychologische Literatur nur bedingt Hinweise dafür liefert, in welchen Dimensionen sich zwei Beispiellösungen unterscheiden bzw. übereinstimmen sollen, um einen Lernzuwachs zu initiieren. Wir konzentrieren uns in fachdidaktischer Hinsicht auf die beschriebenen konzeptuellen Stufen und beleuchten die inhaltliche Ebene, indem die individuellen (fehlerhaften) Lösungen aus der vorangegangenen Problemlösephase zum Vergleich herangezogen werden.

Nicht jeder Vergleichsprozess führt automatisch zu einem Lernzuwachs. Einige experimentelle Studien konnten zwar grundlegend zeigen, dass es gewinnbringend ist, zwei Lösungen für einen Vergleich gleichzeitig zu präsentieren statt einzeln (Gentner, 2003; Oakes & Ribar, 2005), und dass sich instruktionale Unterstützungen wie Vergleichsaufforderungen förderlich auf den Lernzuwachs auswirken (Catrambone & Holyoak, 1989; Loibl & Leuders, 2018), dennoch profitieren einige Lernende mehr und andere weniger von einem Vergleich. Hier bedarf es einer Analyse der Prozessdaten in Form des Blickverhaltens der Lernenden, um zu untersuchen, inwieweit sich Muster und Strukturen im Vergleichsprozess erkennen und als Ursachen für einen unterschiedlichen Lernerfolg deuten lassen.

14.3 Forschungsfragen

Der Fehlerverarbeitungsprozess im Rahmen des Lösungsvergleichs einer normativen und einer individuellen Lösung bedarf nicht ausschließlich vergleichender Prozesse. Gleichmaßen notwendig sind Prozesse des Nachvollziehens der normativen und der individuellen Lösung. Dementsprechend unterteilen wir den Fehlerverarbeitungsprozess in die kognitiven Prozesse „Nachvollziehen“ und „Vergleichen“. Es ist naheliegend, dass sich die beiden Prozesse auch in den Blickbewegungen abzeichnen. Eine Betrachtung der Blickbewegungen beim Vergleich birgt also die Chance, unterschiedliche Prozesse abzubilden, was mit retrospektiven Methoden nicht möglich ist.

Vor diesem Hintergrund lässt sich das Forschungsinteresse durch folgende Fragestellungen präzisieren:

- Welche Abfolgen von kognitiven Prozessen bei der Fehlerverarbeitung lassen sich identifizieren durch die gemeinsame Interpretation von Blickbewegungs- und Interviewdaten?
- Wie unterscheiden sich Lernende mit hohen oder niedrigen Lernzuwächsen dabei?

14.4 Material und Methode

Vorgehen

An der vorliegenden Studie nahmen 63 Lernende der 5. Klasse teil, wobei die Eye-Tracking-Daten von fünf Lernenden aufgrund einer fehlerhaften Kalibrierung ausgeschlossen wurden. Die Lernenden wurden einzeln für eine Schulstunde in einen separaten Raum gebeten. Der Raum wurde so gewählt, dass möglichst optimale Bedingungen für eine Erhebung mit einem Eye-Tracker herrschten: Die Messung wurde in einem leicht abgedunkelten Raum unter stabilen Lichtverhältnissen durchgeführt. Der Stimulus wurde so positioniert, dass keine extremen Blickwinkel vonnöten waren. Weiter wurde bei den Lernenden darauf geachtet, dass keine starken Sehbeeinträchtigungen vorhanden waren, und sie wurden darum gebeten, ungeschminkt zu erscheinen (Holmqvist, 2011).

Die Lernenden lösten computerbasiert die eingangs vorgestellte Problemlöseaufgabe zur Erarbeitung der Bruchrechnung. Mithilfe des Kodierleitfadens von Loibl und Leuders (2018) zu den oben beschriebenen Kategorien wurde der individuelle Lösungstyp jedes Lernenden kodiert. Im Anschluss verglichen die Lernenden ihren individuellen Lösungstyp mit der normativen Lösung von Ole (Stufe 4; Abb. 14.1 rechts). Während des Vergleichsprozesses wurden die Blickbewegungen mithilfe eines monitorbasierten Eye-Trackers aufgezeichnet. Ein die Aufgabe überlagerndes Video der Blickbewegungen des jeweiligen Lernenden diente daraufhin als Stimulus für ein Stimulated Recall Interview. Durch das Video

Tab. 14.2 Einteilung von Schülergruppen im Zuge der Gegenüberstellung der erreichten Stufen in der Problemlösephase und im Posttest

		Stufe im Posttest				
		0	1	2	3	4
Stufe in Problemlösephase	0	○	○	~	+	+
	1	○	○	~	+	+
	2	○	○	○	+	+
	3	•	•	•	•	•
	4	•	•	•	•	•

1 = Natürliche-Zahl-Konzept, 2 = Teil-Ganzes-Konzept, 3 = basales Anteilskonzept, 4 = flexibles Anteilskonzept
+ = Lernende mit substanziellem Lernzuwachs, ~ = mit marginalem Lernzuwachs (nicht betrachtet), ○ = Lernende ohne Lernzuwachs, • = bereits in Problemlösephase leistungsstarke Lernende (nicht betrachtet)

werden die Lernenden dazu angeregt, sich an ihre Gedanken während dieses Ereignisses zu erinnern (Lyle, 2003). Abschließend bearbeiteten die Lernenden einen Posttest, damit Lernzuwächse diagnostizieren werden konnten. Der Posttest enthielt eine analoge Aufgabe zur Problemlöseaufgabe.

Zentrale Idee der Auswertung ist die Gegenüberstellung der Vergleichsprozesse von Lernenden mit substanziellem Lernzuwachs und von Lernenden ohne Lernzuwachs. Um diese Gruppen zu unterscheiden, wurden die in der Problemlösephase und im Posttest generierten Schülerlösungen den Lösungstypen 1 bis 4 aus Tab. 14.1 zugeordnet. Der Lernzuwachs wurde demnach über die erreichten Verständnisstufen in der Problemlösephase und im Posttest definiert (Tab. 14.2). Lernende, die schon in der Problemlösephase ein basales oder flexibles Anteilskonzept zeigten, wurden in dieser Studie nicht berücksichtigt, da sie durch den Vergleich keinen Lernfortschritt mehr erreichen konnten. Auch wurden die Lernenden mit marginalem Lernzuwachs (im Posttest noch kein basales Anteilskonzept) ausgeschlossen, um so Lernende aus den extremen Gruppen (ohne Lernzuwachs, mit substanziellem Lernzuwachs) gegenüberzustellen.

Die in den Stimulated Recall Interviews genannten Gedanken wurden induktiv kategorisiert und zu Phasen im Vergleichsprozess zusammengefasst.

Aufgabe und Lernumgebung

Die in der Problemlösephase zu bearbeitende Aufgabe (s. Eingangsbeispiel) stammt aus dem fachdidaktischen Forschungs- und Entwicklungsprojekt KOSIMA (Barzel et al., 2011). Loibl und Leuders (2018) untersuchten bereits im Paper-Pencil-Format die Lösungstypen, die von den Lernenden generiert werden. Weiter hat sich gezeigt, dass eine Transposition auf den Computer möglich ist, ohne die kognitiven Prozesse wesentlich zu verändern, da sowohl in der papierbasierten als auch in der computerbasierten Lernumgebung die gleichen

Lösungstypen auftreten (Boomgaarden et al., 2019). Im Rahmen der vorliegenden Studie bearbeiteten die Lernenden die Aufgabe in einer computerbasierten Lernumgebung. Die Bearbeitung der Aufgabe in der Lernumgebung sah eine grafische Lösung in Form von zwei Bruchstreifen sowie einen Antwortsatz vor. Der Lösungstyp der grafischen Lösung der Lernenden aus der Problemlösephase stellte die individuelle Lösung dar, welche in der Instruktionsphase mit der normativen Lösung von Ole (Tab. 14.1, Stufe 4) verglichen wurde. Um von dem Vergleich mit der normativen Lösung (Abb. 14.1) profitieren zu können, ist die Beachtung dreier relevanter Merkmale notwendig: die Anzahl, die Länge sowie die Farbe der Streifenabschnitte. So werden von beiden Gruppen (s. Eingangsbeispiel) sowohl die Anzahl der Kinder als auch die Anzahl an Treffern berücksichtigt.

Eye-Tracker

Der Eye-Tracker Tobii Pro X3-120 basiert auf der Auswertung der Position der Pupille und der Reflexion des ausgesendeten Infrarotlichtes an der Hornhaut (*pupil centre corneal reflection*). Für die Klassifizierung der Fixationen wird der geschwindigkeitsbasierte Algorithmus Velocity Threshold Identification (I-VT) angewendet. Unter idealen Bedingungen wird bei einer binokularen Messung eine Accuracy von 0.4° und eine Precision von 0.24° erwartet. Im Zuge einer 5-Punkte-Kalibrierung ergab sich bei uns eine mittlere Accuracy von 1.27° und eine mittlere Precision von 0.72° . Die Blickbewegungen wurden mit der Software Tobii Pro Lab mit 120 Hz aufgezeichnet. In unserem Fall blickten die Lernenden hierbei auf einen im Mittel 62 cm entfernten, 14 Zoll großen Bildschirm mit einer Auflösung von 1920×1080 Pixeln.

Methode

Auf der am Bildschirm dargestellten individuellen und der normativen Lösung wurden gleich große Areas of Interest (AOIs) definiert, sodass die Wechsel zwischen den AOIs in vertikale Transitionen des Nachvollziehens einer Lösung (individuell *oder* normativ) und in horizontale Transitionen des Vergleichens zwischen den Lösungen unterteilt werden konnten (Abb. 14.2).

Um die Vergleichsprozesse zu visualisieren, erfolgte die Auswertung in Cindy Script mithilfe eines selbst entwickelten Algorithmus, welcher die Blickdaten weiterverarbeitet:

1. Die Koordinaten der Blickpunkte pro Zeitstempel werden den AOIs 1 bis 6 zugeordnet.
2. Die Wechsel zwischen den AOIs werden als vertikale Transitionen (innerhalb der individuellen und innerhalb der normativen Lösung) und horizontale Transitionen klassifiziert.

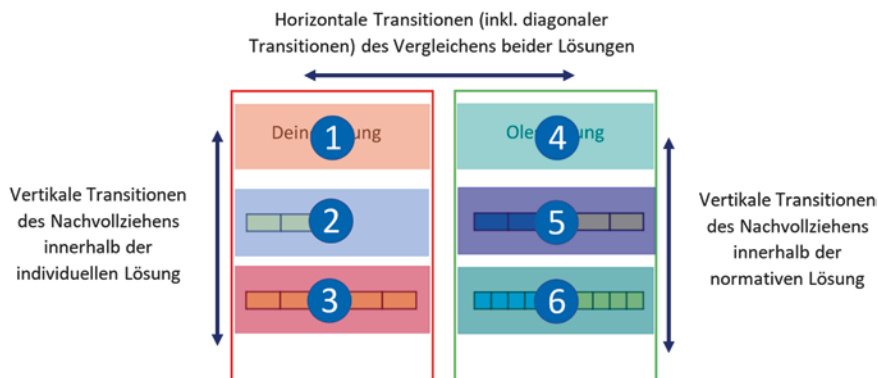


Abb. 14.2 Lokalisierung der AOIs und Definition der Transitionen

- Die zeitliche Abfolge dieser diskreten Transitionen wird zu einer Dichte geglättet und gegen die Zeit abgetragen, sodass das Diagramm sowohl isolierte Transitionen erkennen lässt als auch bei hoher Ballung von Transitionen die mittlere Anzahl der Transitionen pro Sekunde zeigt.

Das Diagramm der Transitionendichten jedes Lernenden wurde im Rahmen der Auswertung durch Beschreibungen aus den Interviews ergänzt.

14.5 Ergebnisse

Vergleichsprozesse können durch eine Reihe globaler Maße charakterisiert werden.

Tab. 14.3 zeigt die von uns analysierten Maße im Vergleich der Lernenden ohne und mit substanziellem Lernzuwachs. Nach unserer Einteilung hatten neun Lernende einen substanziellen Lernzuwachs, während 31 Lernende keinen Lernzuwachs hatten (Abschn. 14.2).

Die globalen Maße weisen keine Unterschiede auf, welche als Indikatoren für einen erfolgreichen Vergleichsprozess aufgefasst werden könnten. Bei der Betrachtung von Einzelfällen zeigt sich jedoch, dass Lernende, die ihre individuelle Lösung fast gänzlich außer Acht gelassen haben, keinen Lernzuwachs erreichen konnten (Abb. 14.3). In den Blickbewegungen ist das erkennbar durch die sehr geringe Anzahl vertikaler Transitionen innerhalb der individuellen Lösung.

Auch in Fällen mit gleichermaßen intensivem Nachvollziehen beider Lösungen (hohe Dichten der vertikalen Transitionen), kombiniert mit relativ wenigen vergleichenden Transitionen, fand kein Lernzuwachs statt (Abb. 14.4).

Trotz dieser markanten Fälle ist bei dem Großteil der Fälle kein Zusammenhang zwischen globalen Maßen der Blickbewegungsdaten und dem Lernzuwachs

Tab. 14.3 Ausgewählte globale Maße der Vergleichsprozesse

	Lernende mit substanziellem Lernzuwachs		Lernende ohne Lernzuwachs	
Dauer des Vergleichs- prozesses	$M = 38.1\text{ s}$	$SD = 21.35\text{ s}$	$M = 35.06\text{ s}$	$SD = 23.45\text{ s}$
Gesamtzahl an Transitionen	$M = 68.9$	$SD = 54.98$	$M = 83.27$	$SD = 89.80$
Relativer Anteil horizontaler (ver- gleichender) Transitionen	37.75 %		38.29 %	
Relativer Anteil vertikaler Transitionen innerhalb der individuellen Lösung	18.26 %		21.87 %	
Relativer Anteil vertikaler Transitionen innerhalb der normativen Lösung	44.00 %		39.85 %	

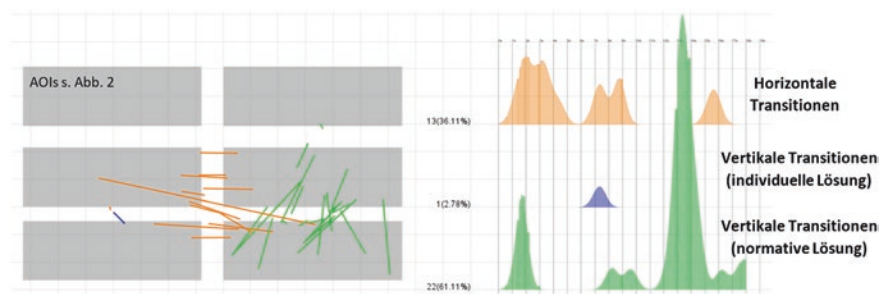


Abb. 14.3 Grafische Darstellung der Transitionen (links) und Transitionendichten (y-Achsen) im zeitlichen Verlauf in Sekunden (x-Achse) am Beispiel des Lernenden 38 ohne Lernzuwachs

festzustellen (Tab. 14.3). Dies lässt eine detailliertere Betrachtung der kognitiven Prozesse in ihrer zeitlichen Abfolge sinnvoll erscheinen. Hier zeigt sich, dass die Vergleichsprozesse von Lernenden mit substanziellem Lernzuwachs sich gegen Ende dadurch auszeichnen, dass die Lernenden zwischen Phasen der Analyse der eigenen Lösung und der Analyse der normativen Lösung wechseln. Allerdings ist dieses Muster auch in den Blickbewegungen einiger Lernender *ohne* Lernzuwachs zu finden. Dies macht eine gemeinsame Interpretation von Blickbewegungs- und Interviewdaten notwendig, bei der sich zeigt, inwiefern in den Blickbewegungen ersichtliche Phasen bei Lernenden unterschiedliche Funktionen hatten. Eine übergreifende Betrachtung der Interviewdaten ergab, dass der beschriebene Wechsel der Phasen bei Lernenden mit substanziellem Lernzuwachs Validierungszyklen darstellt, während er bei Lernenden ohne Lernzuwachs z. B. für wiederholte Zählvorgänge steht.

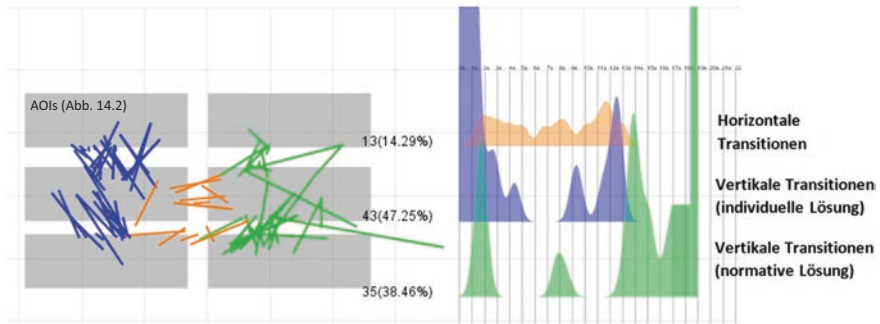


Abb. 14.4 Transitionen (links) und Transitionendichten (rechts) des Lernenden 1 ohne Lernzuwachs (z. T. oben abgeschnitten)

Im Folgenden werden die Vergleichsprozesse eines Lernenden mit substanziellem Lernzuwachs (Levi) und eines Lernenden ohne Lernzuwachs (Henri) exemplarisch gegenübergestellt, um den Unterschied im Detail deutlich zu machen. Bei der Auswahl der beiden Fallbeispiele wurden alle Fälle extrahiert, die sowohl gute Eye-Tracking-Daten als auch ausführliche Interviewdaten aufwiesen. Aus dieser Stichprobe wurden die beiden beschriebenen Fälle zufällig ausgewählt.

a) Lernender mit substanziellem Lernzuwachs (Levi)

Deskriptive Ergebnisse

Während Levi in der Erkundungsphase den Gewinner basierend auf den Fehlertreffern der Mannschaften ermittelte und demnach auf unstruktureller Ebene (Stufe 1) argumentierte, zeigte er im Posttest ein relationales Vorgehen. Er erkannte, dass bei den Jungen die Hälfte und bei den Mädchen mehr als die Hälfte getroffen hat. Diese Schlussfolgerung offenbart ein basales Anteilskonzept (Stufe 3).

Der Vergleichsprozess, welcher dieser Fehlerverarbeitung zugrunde liegt, dauerte 70 s und lässt sich in fünf Phasen untergliedern (Abb. 14.5).

Levi beginnt mit einer intensiven Vergleichsphase, in der er erkennt, welche Inhalte von den Bruchstreifen wiedergegeben werden:

„Erst mal habe ich gesehen, dass er alles mitgenommen hat, er hat auch Fehlertreffer, und, ja, die richtigen Treffer hat er mitgezählt, was ich nicht gemacht habe, und er hat auch eher die Treffer zählen lassen im Gegensatz zu mir.“ (AOI 2 & 5 (obere Streifen) und 3 & 6 (untere Streifen))

In der zweiten Hälfte der Vergleichsphase liegt der Fokus auf der inhaltlichen Interpretation der Farbgebung:

„Da habe ich erst mal verglichen, dass das Blau nicht die Fehlertreffer sind, sondern auch die Treffer, also ich hab da mit dem Gelben die Fehlertreffer gemeint und er mit dem Blauen die Treffer, und deswegen habe ich da erstmal so lange verglichen.“



Abb. 14.5 Phasen des Vergleichsprozesses von Levi

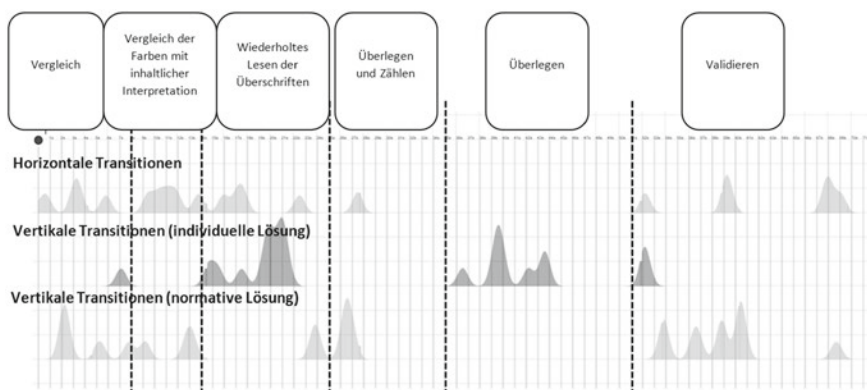


Abb. 14.6 Transitionendichten von Levi

In den Blickbewegungen ist diese erste Phase des Vergleichens durch ein kontinuierliches Auftreten horizontaler Transitionen gekennzeichnet (Abb. 14.6). Es folgt eine Phase der *Orientierung*, in der die Überschriften wiederholt gelesen und zugeordnet werden (lokales Maximum der Transitionen innerhalb der individuellen Lösung):

„Vielleicht kann es sein, dass ich irgendetwas mehrmals falsch lese, denn ich bin so einer, der muss manchmal mehrmals lesen.“

Auch dieser Wechsel zwischen den Überschriften zeigt sich in einem Überdauern horizontaler Transitionen.

Im Anschluss findet sich eine Phase des Nachvollziehens der Lösungen. Hier fällt das Augenmerk auf die Anzahl der Streifenabschnitte. Levi konnte durch *Zählen* der Abschnitte die individuelle Lösung als Teil der normativen Lösung wiederfinden:

„Ich habe da geguckt, dass ich auch alles richtig gemacht habe mit den Fehltreffern alles.“ (AOI 2 & 3 (individuelle Lösung) und 5 & 6 (normative Lösung))

Diese Phase ist in den Blickbewegungen insbesondere durch das längere Ausbleiben jeglicher Transitionen (Plateaus) gekennzeichnet.

Auch die anschließende Phase beinhaltet intensive *Überlegungen*:

„Da hab ich es mir die ganze Zeit gedacht, oh ne, da ist doch irgendetwas falsch.“

In den Blickbewegungen wird deutlich, dass der Fokus auf der individuellen Lösung liegt, da hier ein globales Maximum der vertikalen Transitionen innerhalb der individuellen Lösung zu erkennen ist. Weder horizontale noch vertikale Transitionen innerhalb der normativen Lösung treten auf, dafür jedoch ein Plateau von ca. 6 s.

Diese Überlegungen führten allerdings zu keiner weiteren Erkenntnis, sodass sich die abschließende Phase durch eine *Validierung* der bisherigen Erkenntnisse des Nachvollziehens beider Lösungen auszeichnet:

„Ich hab alles in meinem Kopf nochmal durchgespielt und dann hab ich geguckt, ja, okay, ja, Ole hat die Treffer und alles gezählt und ich die Fehltreffer, und der hat beides mitgenommen, also er hat alles aufgezeichnet und ich halt nicht [...].“

In Folge dessen zeigt sich ein erneuter Dichteanstieg aller hier betrachteten Typen von Transitionen. Durch ein überdauerndes Auftreten von Transitionen innerhalb der normativen Lösung mit lokalem Maximum wird der Fokus auf diese Lösung deutlich.

Interpretation

Auffallend ist, dass im Rahmen dieses Vergleichs alle relevanten Merkmale (Anzahl, Länge und Farbe der Streifenabschnitte) berücksichtigt und in einen inhaltlichen Zusammenhang mit den Variablen der Aufgabe gebracht werden. Die Kombination aus Blickbewegungs- und Verbaldaten zeigen ausführliche Phasen des Überlegens, verknüpft mit intensiven Analysen der individuellen Lösung. Besonders gewinnbringend erscheint in dieser Bearbeitung der Validierungsprozess. Die Teilprozesse „Zählen“ und „Validieren“ sind spezifische Prozesse des „Nachvollziehens“. Beim „Zählen“ sowie beim „Validieren“ werden nacheinander beide Lösungen betrachtet und nachvollzogen, ohne diese direkt in Beziehung zu setzen.

b) Lernender ohne Lernzuwachs (Henri)

Henri erarbeitete in der Problemlösephase eine Lösung auf unistruktueller Ebene. Beim Vergleich zweier Mannschaftsergebnisse im Papierkorbball bezog er sich bei der Wahl des Siegers auf die absolute Anzahl der Tore (Stufe 1). Im Posttest zeigte er weiterhin ein unistruktuelles Vorgehen. Zwar argumentierte Henri hier über die Anzahl der Kinder, doch fehlte die Erkenntnis, dass beide Komponenten, die Anzahl der Kinder und die Anzahl der Treffer, von Bedeutung sind. Der Vergleichsprozess erstreckte sich über 69 s und lässt sich in acht Phasen unterteilen (Abb. 14.7).

Henri startet mit einer Phase der *Orientierung*. Er liest die Überschriften und erkennt bereits, dass sich Anzahl und Länge der Streifenabschnitte unterscheiden:

„Da habe ich das gelesen.“(AOI 1 & 4 (Überschriften))



Abb. 14.7 Phasen des Vergleichsprozesses von Henri

„Hier habe ich geguckt, ähm gezählt, weil ob es die gleiche Breite wäre und die gleiche Größe.“ (AOI 2 & 5 (obere Streifen))

Es folgt eine Phase, in der die Streifenabschnitte aller Bruchstreifen *gezählt* werden:

„Hier habe ich dann die gezählt. Hier habe ich sie auch gezählt und hier auch, und da habe ich sie auch gezählt.“ (AOI 2 & 3 (individuelle Lösung) und 5 & 6 (normative Lösung))

Die Blickbewegungen zeichnen sich durch eine kontinuierliche Präsenz horizontaler Transitionen aus, während vertikale Transitionen vereinzelt auftreten (Abb. 14.8).

Es folgt ein erster *Vergleich*, in dessen Rahmen die beiden unteren Streifen in Bezug gesetzt werden. Das Augenmerk liegt auf der Anzahl und der Länge der Streifenabschnitte:

„Da habe ich es so, weil es hier breiter sind als die Kästchen, weil sie sind dünner als die. Sie sind ein bisschen breiter. Hier habe ich geguckt und, ähm, geschaut, als ob es die gleiche Menge wären.“ (AOI 3 & 6 (untere Streifen))

Diese Phase zeichnet sich ebenfalls durch eine Dominanz der horizontalen Transitionen aus. Zu Beginn der Phase wird ein intensiver Vergleich durch eine breite Dichtekurve visualisiert, welche durch die Überlagerung von Einzelereignissen entsteht.

Im Anschluss wiederholt sich die Phase der *Orientierung*, sodass noch einmal die Überschriften gelesen und mit den Lösungen in Bezug gesetzt werden, was sich

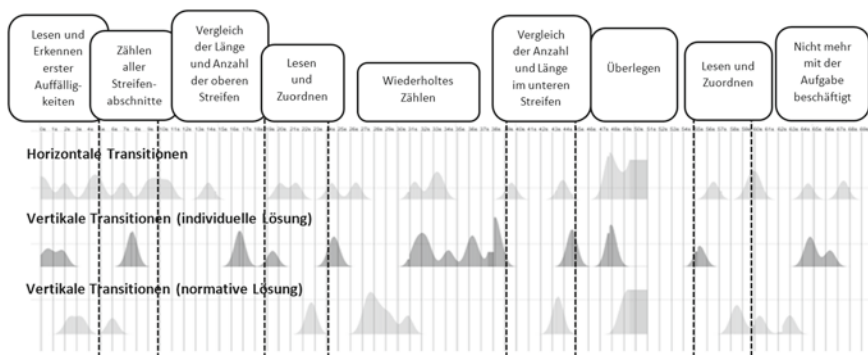


Abb. 14.8 Transitionendichten von Henri

in den Blickbewegungen durch einen Dichteanstieg der horizontalen Transitionen zeigt.

Es folgt eine Phase des wiederholten *Zählens* der Abschnitte der Bruchstreifen:

„Also hier habe ich geschaut, ob das die gleiche Größe wären, dann waren die und dann habe ich bemerkt, dass da noch zwei fehlen.“

In den Blickbewegungen bedeutet Zählen ein Verweilen in einer AOI, sodass in dieser Zeit auch keine Transitionen auftreten. Somit finden in dieser Phase Zählvorgänge zwischen zwei lokalen Maxima der Transitionendichten statt.

Im Anschluss werden diesmal die beiden unteren Streifen hinsichtlich der Anzahl und Länge der Streifenabschnitte *verglichen*:

„Beim Ersten habe ich es schnell gezählt, beim Zweiten langsam und beim Dritten habe ich so mit meinen Fingern so gemacht, ob sie gleich groß wären von der Menge also mit der Breite.“

Im Diagramm der Transitionendichten ist zu erkennen, dass sich Henri zunächst auf einen Streifen der individuellen Lösung fokussiert, da die vorherige Phase mit Transitionen innerhalb der individuellen Lösung endet. Es folgt der Wechsel zu einem Streifen der normativen Lösung, was durch den minimalen Dichteanstieg der horizontalen Transitionen zu erkennen ist. Es schließt sich ein Wechsel zurück zum Streifen der individuellen Lösung an, wobei den vertikalen Transitionen keine Bedeutung beigemessen wird.

Es folgt eine Phase des *Überlegens*:

„Hier habe ich wieder überlegt.“

In dieser Zeit wanderten die Augen vermehrt aus dem Bereich des Stimulus heraus, sodass der Dichteanstieg der vertikalen Transitionen innerhalb der individuellen Lösung zu vernachlässigen ist.

Der Vergleichsprozess endet mit einer erneuten *Orientierungsphase*, in der Überschriften gelesen und den Lösungen zugeordnet werden.

„Immer wenn ich etwas mache, muss ich es immer ganz oft lesen.“

Für die darauffolgenden Blickbewegungen konnte keine Erklärung gegeben werden.

Interpretation

Auffallend an diesem Vergleichsprozess ist, dass alle relevanten Merkmale mit Ausnahme der Farben bewusst betrachtet wurden. Allerdings wurden diese Merkmale in keinen inhaltlichen Zusammenhang gebracht, sodass der Prozess aus einem reinen Vergleich der Oberflächenmerkmale bestand. Der abschließende Wechsel einer hohen Transitionendichte innerhalb der individuellen Lösung und einer hohen Transitionendichte innerhalb der normativen Lösung stellte im Fall

des Lernenden mit substanziellem Lernzuwachs einen Validierungszyklus dar. In diesem Beispiel hingegen stellt sich der Wechsel allein als Kontrolle des richtigen Erkennens der Oberflächenmerkmale (Zählen der Streifenabschnitte) heraus. Weiter legt die Kombination aus Blickbewegungen und Verbaldaten die Vermutung nahe, dass es dem Lernenden schwerfiel, die Aufmerksamkeit auf dem Stimulus ruhen zu lassen.

14.6 Diskussion

Die Befunde dieser Studie geben Aufschluss darüber, wie Fehlerverarbeitungsprozesse beim Lösungsvergleich als Kombination von Prozessen des Nachvollziehens und des Vergleichens vonstattengehen können. Durch die Betrachtung von Einzelfällen ist eine vollständige Erfassung aller Phänomene keinesfalls gewährleistet. Dennoch stützt die Tatsache, dass ein Außerachtlassen der individuellen Lösung keinen Lernzuwachs nach sich zieht, die Vermutung von Loibl und Leuders (2018), dass das Nichtbeachten der individuellen Lösung die Vorteile eines Vergleichsprozesses unterbindet. Weiter wird dadurch die Bedeutung von Aufforderungen zum Vergleich unterstrichen (z. B. Berthold et al., 2007; Schworm & Renkl, 2007).

Erst durch die Kombination von Transitionendichten und Interviewdaten ergab sich ein umfassendes Bild der individuellen Vergleichsprozesse. Die Exploration der Interviewdaten ergab, dass der Lernzuwachs mit der Ebene zusammenhängt, auf welcher der Vergleich stattfand. Lernende ohne Lernzuwachs erkennen zwar die relevanten Merkmale, doch für einen substanziellen Lernzuwachs ist zusätzlich das inhaltliche In-Beziehung-Setzen notwendig. Für diese Studie wurde die Oberfläche hinsichtlich der Auswertung der Blickbewegungen möglichst einfach gehalten. Es ist zu vermuten, dass sich die Möglichkeit eines erneuten Lesens der Aufgabenstellung während des Vergleichs positiv auf den inhaltlichen Bezug ausgewirkt hätte.

Die fehlenden Informationen zur Aufgabenstellung während des Vergleichs und auch die generell kurzen Vergleichsdauern können möglicherweise erklären, warum der Großteil der Lernenden keinen Lernzuwachs erreichte.

Weiterhin zeigte sich beim Lernenden mit substanziellem Lernzuwachs insbesondere der abschließende Validierungszyklus als gewinnbringend, was an Ergebnisse aus der Expertiseforschung anknüpft (z. B. Streit et al., 2019).

Zusammenfassend bestätigt die vorliegende Studie, dass es durch die Analyse von Blickbewegung möglich ist, Prozesse des Nachvollziehens und des Vergleichens zu trennen. Die grafische Darstellung der Transitionendichten ermöglicht einen ersten Eindruck von den Fehlerverarbeitungsprozessen beim Vergleich von Lösungen und erste Vermutungen über den Lernerfolg. Jedoch generiert erst die gemeinsame Interpretation von Blickbewegungs- und Interviewdaten ein umfassendes Verständnis des Fehlerverarbeitungsprozesses im Zusammenhang mit dem Lernzuwachs. Die so gewonnenen Erkenntnisse können

für die Optimierung künftiger Vergleichssituationen im Rahmen von Fehlerverarbeitungsprozessen genutzt werden. An diese Studie anknüpfend, bedarf es nun der Auswertung aller 63 Fälle, um so weitere Implikationen für die Forschung und für die Unterrichtspraxis ableiten zu können.

Literatur

- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, 48(2), 87–113.
- Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal*, 93, 373–397.
- Barzel, B., Prediger, S., Leuders, T., & Hußmann, S. (2011). *Kontexte und Kernprozesse – Aspekte eines theoriegeleiteten und praxiserprobten Schulbuchkonzepts*. WTM.
- Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, 17(5), 564–577.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluation the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. Academic Press.
- Boomgaarden, A., Loibl, K., & Leuders, T. (2019). Computer-based problem solving to prepare for adaptive consolidation. *Proceedings of the 13th international conference on computer-supported collaborative learning (CSCL 2019)*.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problemsolving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1147–1156.
- Chi, M. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in Instructional Psychology*, 5, 161–238.
- Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, 22(3), 206–214.
- Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, 22(1), 47–61.
- Gentner, D. (2003). Why we're so smart. In D. Gentner und S. Goldin-Meadow (Hrsg.) *Language in mind: Advances in the study of language and thought*, 195–235.
- Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, 34(5), 752–775.
- Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, 95(2), 393.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15(1), 1–38.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Loibl, K., & Leuders, T. (2018). Errors during exploration and consolidation – The effectiveness of productive failure as sequentially guided discovery learning. *Journal für Mathematik-Didaktik*, 39(1), 69–96.
- Loibl, K., & Leuders, T. (2019). How to make failure productive: Fostering learning from errors through elaboration prompts. *Learning and Instruction*, 62(1), 1–10.
- Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction*, 34, 74–85.
- Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal*, 29(6), 861–878.
- Ni, Y., & Zhou, Y. D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, 40(1), 27–52.

- Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy*, 7(1), 85–98.
- Prediger, S. (2011). Vorstellungsentwicklungsprozesse initiieren und untersuchen. *Der Mathematikunterricht*, 57(3), 5–14.
- Prediger, S., Barzel, B., Hußmann, S., & Leuders, T. (2013). *Mathewerkstatt 6*. Cornelsen.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
- Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, 101(3), 529.
- Rittle-Johnson, B., & Star, J. R. (2011). The power of comparison in learning and instruction: Learning outcomes supported by different types of comparisons. *Psychology of learning and motivation*, 55, 199–225.
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, 101(4), 836.
- Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for selfexplaining examples. *Journal of Educational Psychology*, 99(2), 285–296.
- Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Garnott & J. Parziale (Hrsg.), *Microdevelopment: Transition processes in development and learning* (S. 31–58). Cambridge University Press.
- Streit, C., Rüede, C., Weber, C., & Graf, B. (2019). Zur Verknüpfung von Lernstandeinschätzung und Weiterarbeit im Arithmetikunterricht: Ein kontrastiver Vergleich zur Charakterisierung diagnostischer Expertise. *Journal für Mathematik-Didaktik*, 40(1), 37–62.
- Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, 104, 147–200.



Unterschiede im Leseverhalten von Algorithmen und Programmen

15

Philipp Kather und Jan Vahrenhold

Inhaltsverzeichnis

15.1	Einleitung und Forschungslage	245
15.2	Theoretischer Hintergrund	246
15.3	Forschungsfragen und Hypothesen	247
15.4	Material und Methodik	250
15.5	Ergebnisse	255
15.6	Diskussion	258
	Literatur	259

15.1 Einleitung und Forschungslage

Der Begriff „Algorithmus“ ist vor allem in der frühen Literatur oft synonym mit dem Begriff „Programm“ verwendet worden. Das Verständnis von Programmen, d. h. kleinen, in sich abgeschlossenen Einheiten von Programmtext, ist vielfach untersucht worden. Eine frühe Übersicht hierzu entwickelter Theorien und Modelle findet sich bei von Mayrhauser und Vans (1995), der aktuelle Forschungsstand wird von Duran (2020) zusammengefasst. Forschungsarbeiten im Bereich des Programmverständnisses greifen in Teilen auf Eye-Tracking als Forschungsmethodik zurück (Obaidellah et al., 2018).

P. Kather (✉) · J. Vahrenhold

Institut für Informatik, Westfälische Wilhelms-Universität Münster, Münster, Deutschland

E-Mail: philipp.kather@uni-muenster.de

J. Vahrenhold

E-Mail: jan.vahrenhold@uni-muenster.de

Durch Haberman et al. (2005) wurde der Begriff „Algorithmus“ von Programmen dadurch abgegrenzt, dass zum einen ein Verfahren (z. B. als Programm) beschrieben und zum anderen eine schriftliche Analyse des Problems und der Eigenschaften des Verfahrens, wie z. B. dessen Korrektheit oder Effizienz, vorhanden ist (Haberman et al., 2005, S. 74). Um das Anspruchsniveau mit dem einer formal-mathematisch orientierten Algorithmmikvorlesung abzugleichen, fordern wir zudem, dass dieser Nachweis aus Sicht der Lesenden nicht trivial ist und dass es sich um einen formalen Beweis und nicht um eine Plausibilitätsbetrachtung handelt (Kather & Vahrenhold, 2021).

Mit dem Verständnis von Algorithmen in diesem Sinne haben sich bislang nur wenige Arbeiten beschäftigt. Einige Studien beschrieben, inwieweit allgemeine Verständnis- und Lerntheorien sich auf Algorithmen übertragen lassen (da Rosa, 2010, 2015; Narayanan & Hergarty, 2002). Andere gehen näher auf Eigenheiten von Algorithmen ein, wie das Verständnis von Korrektheit von Algorithmen oder das Verständnis spezieller Arten von Algorithmen (Ben-David Kolikant, 2005; Velázquez-Iturbide, 2013). Durch den in diesem Kapitel beschriebenen Einsatz von Eye-Tracking erhoffen wir uns weitere Erkenntnisse über den Unterschied zwischen Algorithmen- und Programmverständnis. Konkret wird der Fragestellung nachgegangen, ob Erkenntnisse aus dem bereits gut untersuchten Feld des Programmverständnisses auf Algorithmen übertragbar sind. Erkenntnisse könnten langfristig dazu beitragen, dass Algorithmen effektiver didaktisch aufbereitet werden können, indem z. B. algorithmenspezifische Lesestrategien berücksichtigt werden.

15.2 Theoretischer Hintergrund

Wir legen den Ansatz von Soloway et al. (1983) zugrunde, der auf einem Transfer der Schematheorie (Rumelhart, 1980) in den Bereich des Programmverständnisses beruht. In diesem Ansatz werden als Entsprechung zu Rumelharts Schemata sog. Pläne definiert, die stereotypisierte Lösungen eines (Teil-)Problems darstellen und die dann zur Lösung eines größeren Problems kombiniert werden können. Soloway und Ehrlich (1984) haben festgestellt, dass fortgeschrittene Studierende oftmals mit solchen Plänen so gut vertraut sind, dass sie Programme, die diese ausnutzen, besser verstehen können als Programme, die keine solchen Pläne verwenden. Abb. 15.1 zeigt zwei beispielhafte Programmausschnitte.

Eye-Tracking ist bereits als Methodik zur Untersuchung des Programmverständnisses eingesetzt worden: Obaidellah et al. (2018) haben über 60 solcher Eye-Tracking-Studien untersucht. Von diesen beschäftigten sich 26 Arbeiten mit dem Verstehen von Programmtext und 19 Arbeiten mit dem Finden bewusst eingebauter Fehler; die restlichen Arbeiten fokussierten auf dem Verständnis von Diagrammen, kollaborativem Arbeiten oder dem Abgleich von Anforderungen und

```
List<Integer> list = new ArrayList<>();

for (int i = 0; i < numbers.length && numbers[i] != 999; i++){

    list.add(numbers[i]);

}

int r = 0;

for (int i = 0; i < list2.size(); i++){

    r += list2.get(i);

}
```

Abb. 15.1 Der obere Programmausschnitt ist aufgrund der Abbruchbedingung ungewöhnlich und setzt keinen Plan um. Der zweite Programmausschnitt setzt einen *running total variable*-Plan (Soloway & Ehrlich, 1984) um

Umsetzungen in der professionellen Softwareentwicklung. Weder in dieser Arbeit¹ noch in einer aktuellen Synopse von Arbeiten aus der Mathematikdidaktik (Strohmaier et al., 2020) werden jedoch Studien erwähnt, die Eye-Tracking zur Untersuchung von Algorithmenverständnis nach unserer Definition einsetzen.

Für unsere Studie ist vor allem die Arbeit von Crosby und Stelovsky (1990) von Interesse, die ergab, dass Experten die meiste Zeit des Programmlesens auf komplexen Programmabschnitten verbringen. Die Untersuchung der im nächsten Abschnitt formulierten Hypothese 1 (a) dient der Überprüfung dieser Ergebnisse für unsere Studiengruppe mit dem Ziel, eine gemeinsame Basislinie zu finden. Die Untersuchung der ebenfalls im nächsten Abschnitt formulierten Hypothese 1 (b) dient der Beantwortung der Frage, ob diese Ergebnisse auf den Bereich des Algorithmenverständnisses übertragbar sind. Bei diesen Untersuchungen nutzen wir die Umsetzung bzw. das Fehlen von Plänen, deren Kenntnis und Beherrschung wir bei fortgeschrittenen Studierenden als vorhanden annehmen können, stellvertretend für die Komplexität von Programmabschnitten.

15.3 Forschungsfragen und Hypothesen

Wir betrachten das Leseverhalten von fortgeschrittenen Studierenden im Kontext des Algorithmenverständnisses. Basierend auf den im vorherigen Abschnitt beschriebenen Planschema weisen die von uns verwendeten Programme und Algorithmen sowohl Abschnitte auf, die Planschemata umsetzen und somit als „einfach“ klassifiziert werden können, als auch „komplexe“ Programmabschnitte, die zu keinem Planschema korrespondieren.

¹In dieser Übersicht werden zwar einige Arbeiten erwähnt, die „Algorithmen“ untersuchen, in allen Fällen liegt jedoch die eingangs erwähnte synonyme Verwendung der Begriffe „Programm“ und „Algorithmus“ vor. Insbesondere fehlt den jeweiligen Materialien die gemäß unserer Definition notwendige Komplexität bzw. formalisierte Betrachtung von Programmeigenschaften.

Basierend auf den Erkenntnissen von Crosby und Stelovsky (1990) erwarten wir, dass Programmabschnitte, die Planschemata umsetzen, schneller verstanden werden als Programmabschnitte, die keinem Planschema entsprechen. Einen Zusammenhang zwischen Kognition und visueller Aufmerksamkeit wurde erstmals in der Eye-Mind-Hypothese (Just & Carpenter, 1980) formuliert. Chen et al. (2011) und die dort zitierten Arbeiten bringen ebenfalls die Verweildauer auf Abschnitten mit deren Komplexität und Schwierigkeiten beim Verständnis in Verbindung (s. auch weitere in den Synopsen zum Programmverständnis [Obaidellah et al., 2018] und in der Mathematik [Strohmaier et al., 2020] genannte Arbeiten). Dementsprechend erwarten wir, dass weniger Zeit mit dem Verständnis von planbasierten „einfachen“ Programmabschnitten verbracht wird.

Bei Algorithmen liegt neben dem Programmtext noch eine Beweisskizze über eine Eigenschaft des Programms vor. Dies ist eine weitere Informationsquelle über das Programm. In einer qualitativen Studie konnten wir beobachten, dass zusätzliche Beweisskizzen das Leseverhalten von Programmtext maßgeblich beeinflussen können (Kather & Vahrenhold, 2021). Wir untersuchen daher, ob die zusätzliche Präsentation dieser Informationsquelle das Verhalten fortgeschrittener Studierender beim Lesen von Programmtext, genauer das Verhältnis von Verweildauer auf den unterschiedlichen Abschnitten, quantitativ beeinflussen kann.

Konkret betrachten wir zunächst die folgenden Hypothesen:

Hypothese 1 (a): Bei Programmen fokussieren sich fortgeschrittene Studierende überwiegend auf komplexen Programmtext.

Wir betrachten die Situation, in der ein Programmtext gelesen und das Programm mental für eine gegebene Eingabe ausgewertet werden muss. Wir erwarten, dass bei Programmen die normalisierte Verweildauer² auf komplexem Programmtext größer ist als die normalisierte Verweildauer auf einfachem Programmtext.

Ein grundlegender Unterschied zwischen Programmen und Algorithmen ist, dass der Algorithmus zusätzlich zum Programmtext, welcher beiden gemein ist, zusätzlich einen Beweis oder eine Beweisskizze aufweist. Im Rahmen unserer Studie wurden die Teilnehmer und Teilnehmerinnen aufgefordert, beim ersten Lesen des Algorithmus zunächst ausschließlich den Programmtext zu lesen. Zur Untersuchung der folgenden Hypothese werden ausschließlich die in dieser Phase erfassten Verweildauern untersucht:

Hypothese 1 (b): Bei Algorithmen fokussieren sich fortgeschrittene Studierende während des Programmcodelesens überwiegend auf komplexen Programmtext.

Wir betrachten die Situation, in der ein Programmtext in dem Wissen gelesen werden soll, dass im Nachgang ein Beweis zu einer Eigenschaft des Programms

²Bei dem Vergleich wurde berücksichtigt, dass komplexer Programmtext oft länger als einfacher Programmtext ist (Abschn. 15.4). Aus diesem Grund wurde für jeden Programmabschnitt betrachtet, aus wie vielen Zeilen sinnvollen Programmtext er besteht. Das heißt Zeilen, die z. B. keinen Programmtext oder nur eine schließende Klammer enthalten, wurden nicht gezählt. Durch diese Anzahl an Zeilen wurde die Fixationsdauer anschließend geteilt.

untersucht werden soll. Da sich der Leseauftrag jedoch ausschließlich auf den Programmtext bezieht, erwarten wir wie bei Hypothese 1 (a), dass beim Programmlesen von Algorithmen die normalisierte Verweildauer auf komplexem Programmtext größer ist als die normalisierte Verweildauer auf einfachem Programmtext.

In der Kombination eines 2×2 -Designs dienen diese Hypothesen somit der Untersuchung der Frage, ob sich das Leseverhalten beim Lesen des Programmtexts von Algorithmen von dem beim Lesen von Programmen unterscheidet. Hypothese 1 (a) ist angelehnt an die Erkenntnisse von Crosby und Stelovsky (1990); ihre Untersuchung dient somit auch dem Abgleich mit den aus der Literatur bekannten Erkenntnissen und soll zeigen, dass die Bedingungen dieser Studie überhaupt geeignet sind, solche Effekte festzustellen.

Sollte ein Interaktionseffekt, aber kein durch Hypothese 1 (b) nachweisbarer Haupteffekt erkannt werden, würde dies bedeuten, dass, wenn ein Unterschied vorliegt, dieser bei Algorithmen nicht so stark ausgeprägt ist wie bei Programmen; dies würde insbesondere einen Unterschied zwischen dem Lesen von Programmen und dem Lesen des Programmtextes von Algorithmen nahelegen. Die im Folgenden formulierte Hypothese 2, die in einem solchen Fall als Fortsetzungsfragestellung untersucht wird, ist durch unsere qualitativen Ergebnisse (Kather & Vahrenhold, 2021) motiviert, die suggerieren, dass das Verständnis komplexer Programmabschnitte durch den Beweis unterstützt wird:

Hypothese 2: Beim Lesen von Algorithmen fokussieren sich fortgeschrittene Studierende während des Programmlesens weniger stark auf komplexen Programmtext als beim Lesen von Programmen.

Wir vergleichen das Verhältnis der jeweils auf einfachem und komplexen Programmtext in den entsprechenden Zeiträumen verbrachten Zeit. Ein Bestätigen dieser Hypothese würde zur der in zukünftigen Arbeiten qualitativ zu untersuchenden Folgehypothese führen, dass fortgeschrittene Studierende beim Lesen des Programmtextes von Algorithmen das Verständnis komplexerer Programmtexte in der Erwartung hintanstellen, sich diese nach dem Lesen des Beweises einfacher erschließen zu können. Im Kontext unserer Eye-Tracking-Studie betrachten wir vorbereitend hierzu rein quantitative Aspekte dieser Folgehypothese:

Hypothese 3: Fortgeschrittene Studierende nutzen die zu einem Algorithmus gegebene Beweisskizze, um sich komplexen Programmtext zu erschließen.

Hierzu vergleichen wir abermals das Verhältnis auf beiden Programnteilen, diesmal allerdings zwischen dem Programmlesen und dem Beweislesen. Wir erwarten, dass das Verhältnis zwischen der auf einfachen und komplexen Programmabschnitten verbrachten Zeit beim Beweislesen geringer ist als beim Programmlesen, was bedeuten würde, dass während des Beweislesens der komplexe Programmtext verhältnismäßig mehr Aufmerksamkeit erhält als beim Programmlesen.

15.4 Material und Methodik

Stichprobe

Die Probandinnen und Probanden für diese Studie wurden als Zielstichprobe ausgewählt, indem fortgeschrittene Informatikstudierende persönlich angesprochen wurden. Als Auswahlkriterien galten Erfahrungen in (diskreter) Mathematik und Erfahrungen mit dem Lesen und Schreiben von Java-Programmen. Da die Materialien vorab in keiner Lehrveranstaltung genutzt wurden, waren keine Ausschlusskriterien notwendig. Insgesamt wurden elf Probandinnen und Probanden rekrutiert und entsprechend den Institutsrichtlinien kompensiert. Die Zusammensetzung (acht männliche, drei weibliche Probanden) ist repräsentativ für die Demografie der zugrunde liegenden Population. Da keine institutionellen Vorgaben für die Beteiligung einer Ethikkommission bestanden, wurden die Probandinnen und Probanden gemäß der universitätsweiten Datenschutzrichtlinien informiert und gaben ihre schriftliche Einwilligung zur Teilnahme.

Technische Rahmendaten

Die Studie wurde in einem dedizierten Eye-Tracking-Labor durchgeführt. Zu jedem Zeitpunkt war neben dem Versuchsleiter maximal eine weitere Person anwesend, sodass sich keine Probandinnen bzw. Probanden begegneten. Die Laboreinrichtung entsprach den Empfehlungen von Holmqvist et al., (2011, S. 17–21) und nutzte einen Tobii-Pro-Spectrum-300-Hz-Eye-Tracker mit Tobii Pro Lab (Tobii AB, 2020) als Aufnahmesoftware. Dieses System ermöglicht binokulares Verfolgen der Augen durch *pulsed infrared* und befand sich unter einem 23.8-Zoll-Bildschirm mit einer Auflösung von 1600×1200 Pixeln. Die Ungenauigkeit lag unterhalb von zehn Pixeln bzw. 3.3 mm; dies wurde bei der Konstruktion der Areas of Interest (AOIs) berücksichtigt. Zur Auswertung wurden zunächst Daten durch Tobii Pro Lab und dem IV-T-Filter extrahiert und anschließend durch eigene Programmskripte analysiert.

Ablauf der Experimente

Während der Experimente wurden etablierte Richtlinien eingehalten, um eine möglichst gute Datenerhebung zu ermöglichen (Holmqvist et al., 2011, S. 116–128). So wurde z. B. der Raum regelmäßig gelüftet, und es wurden Abschminktücher bereitgestellt, um die optische Datenerfassung zu verbessern. Während der Experimente betrug der Bildschirmabstand 65 ± 10 cm.

Zu Beginn wurden die Teilnehmenden über den Ablauf des Experiments aufgeklärt. Als Motivation wurde genannt, das Leseverhalten von Programmtext genauer zu untersuchen. Es wurde angesprochen, dass ein gewisser Abstand vom

Bildschirm eingehalten werden müsse, zudem wurde erläutert, welche Daten durch den Eye-Tracker erhoben werden. Erst dann wurde die Einwilligung zur Erhebung und Verarbeitung der Daten eingeholt.

Teil I: Programmverständnis Im ersten Teil der Sitzung (ca. 40 min) wurden den Probandinnen und Probanden nacheinander sechs verschiedene Programme mit mehreren Eingaben pro Programm präsentiert. Die Probandinnen und Probanden wurden darüber informiert, dass jedes Programm nur für wenige Minuten präsentiert würde und dass sie innerhalb dieses Zeitraums die Ausgabe des Programms für möglichst viele der gegebenen Eingaben bestimmen und eingeben sollten. Zwischen den einzelnen Programmen hatten die Probandinnen und Probanden die Möglichkeit, sich nach eigenem Ermessen zu erholen. Diese Erholungsphase betrug nicht mehr als 5 min.

Hierzu wurde der Bildschirm in drei Bereiche unterteilt. Im Zentrum wurde ein in der Programmiersprache Java geschriebenes Programm angezeigt. Am oberen linken Bildschirmrand wurde eine Eingabe präsentiert, am oberen rechten Bildschirmrand befand sich eine Eingabemaske, in der die Ausgabe eingetragen werden konnte. Nach jedem Eintragen einer Ausgabe wurde eine neue Eingabe präsentiert.

Nach der Kalibrierung des Eye-Trackers folgte ein Beispielprogramm, sodass sich die Probandinnen und Probanden mit der Anwendung vertraut machen und Fragen stellen konnten. Während des Experiments waren Fragen nicht erwünscht; hierüber wurde im Vorfeld informiert.

Teil II: Algorithmenverständnis Nach einer kurzen Pause folgte der zweite Hauptteil der Sitzung (ca. 15 min). Hier wurden den Probandinnen und Probanden nacheinander zwei Algorithmen präsentiert. Im Vorfeld wurde erklärt, dass ohne ein Zeitlimit die Problemstellung, danach der Programmtext und anschließend die Beweisskizze gelesen und dann ein Fehler gefunden und korrigiert werden sollte.

Hierzu wurde der Bildschirm in drei Bereiche unterteilt (Abb. 15.2). Im linken Bereich befand sich ein Java-Programm. Im rechten oberen Bereich befand sich die Problembeschreibung, im rechten unteren Bereich die Skizze eines Korrektheitsbeweises.

Sobald die Probandinnen und Probanden das Gefühl hatten, verstanden zu haben, warum das Programm das Problem korrekt löst, sollten sie eine Schaltfläche anklicken. Hierdurch wurde eine Aussage im Beweis hervorgehoben, die über eine nicht näher spezifizierte Zeile im Programmtext argumentierte. In dieser Zeile befand sich jedoch ein auf den ersten Blick unauffälliger Fehler, beispielsweise ein verschobener Index. Die Probandinnen und Probanden sollten nun zügig die zur hervorgehobenen Zeile im Beweis korrespondierende Stelle im Programmtext markieren und mündlich einen Vorschlag zur Fehlerkorrektur machen. Die Phase nach dem Anklicken der Schaltfläche ist jedoch für diesen Artikel nicht relevant, da die Probandinnen und Probanden instruiert wurden, in der Phase vor dem Anklicken der Schaltfläche den Algorithmus zu verstehen, wohingegen sie danach eine Zuordnung durchführen sollten.

Betrachte eine Reihe von N aufeinander folgenden Säulen. Jede Säule besteht aus mehreren Platten. Insgesamt gibt es $N \cdot H$ viele Platten. Dies wird modelliert als Folge $S = s_1, s_2, \dots, s_n$, wobei jedes Folgenglied eine natürliche Zahl ist. Ein Beispiel wäre die Folge 5, 6, 7, 10, mit $H = 7$. Das Ziel ist es, die kleinste Anzahl an Zügen zu bestimmen, sodass alle Säulen die Höhe H haben. In jedem gültigen Zug können dabei beliebig viele Platten von einer Säule auf eine benachbarte umverteilt werden. Im Beispiel wäre ein gültiger Zug, wenn man 3 Platten von der letzten Säule auf die vorletzte bewegt. Dies führt zu der Konfiguration 5, 6, 10, 7.

Beweisskizze

Sei k der erste Eintrag, für den gilt, dass $\sum_{i=1}^k s_i = k \cdot H$. Dieser Bereich kann unabhängig von allen nachfolgenden Säulen in $k - 1$ Zügen in einen Zustand überführt werden, sodass die ersten k Säulen die Höhe H haben. Der Beweis dieser Aussage entfällt, kann aber als valide angenommen werden.

Der Algorithmus bestimmt zunächst für jede Säule, wie viele Platten links von ihr liegen. Dazu berechnet er die Summe aller Elemente in der Folge bis zum jeweiligen Index.

Durch diese Summe findet er die erste Säule k_0 , sodass von der ersten Säule bis vor die Säule k_0 genauso viele Platten liegen, dass durch Umverteilen, wie oben beschrieben, alle Säulen vom Anfang bis k_0 auf die Höhe H gebracht werden können.

Es wird vermehrt, dass eine solche Teilfolge von Säulen gefunden wurde.

Von da an betrachtet der Algorithmus nur noch die verbleibende Teilfolge ab k_0 .

Dazu wird der Index k_0 , an dem die neu zu betrachtende Teilfolge beginnt, aktualisiert.

Dies ist ein korrektes Vorgehen, da von der ursprünglichen Summe die Anzahl an Platten links von k_0 abgezogen werden kann. Es gilt nämlich

$$\sum_{i=1}^t s_i - s_1 = \sum_{i=1}^{t-1} s_i.$$

Anschließend wird von k_0 an analog ein Index k_1 gesucht, sodass sich links von k_1 bis einschließlich k_0 wieder so viele Platten befinden, dass sich alle Säulen von k_0 bis vor die Säule k_1 auf die Höhe H bringen lassen.

Auf diese Weise wird iterativ vorgegangen, bis das Ende der Folge erreicht ist. Jedes Mal, wenn eine weitere solche Teilfolge gefunden wird, wird die Anzahl der gefundenen Teilfolgen weiter erhöht. Aufgrund der Voraussetzung, dass es $N \cdot H$ viele Platten gibt, lässt sich jede Teilfolge auf diese Weise weiter aufteilen.

Durch Weiterführen dieses Vorgehens werden insgesamt t Teilfolgen gefunden. Da $\sum_{i=1}^t S_{k_i} = N \cdot H$ gilt, folgt, dass die Säulen in $N - t$ Zügen auf die Höhe

H umformbar sind.

Get hier!

```
private int compute(int[] numbers, int b) {
    Integer[] array = new Integer[numbers.length + 1];
    array[0] = 0;
    for (int i = 1; i < array.length; i++) {
        array[i] = numbers[i - 1] + array[i - 1];
    }

    int s = 0;
    int c = 0;
    for (int i = 1; i < array.length; i++) {
        if (array[i] - array[s] == (i - s) * b) {
            c++;
            s = i + 1;
        }
    }
    return numbers.length - c;
}
```

Abb. 15.2 Präsentation des *fence leveling*-Algorithmus. Zwei AOIs sind blau hervorgehoben. Die obere umfasst einen Programmabschnitt, der einem Planschema entspricht, wohingegen der untere keinem Planschema entspricht

Auch hier wurde zunächst ein Beispiel präsentiert, um sich mit der Anwendung vertraut zu machen und ggf. Fragen stellen zu können.

Teil III: Abschluss Zum Abschluss wurde mit den Probandinnen und Probanden jede Aufgabe der Sitzung in einem halbstrukturierten Interview besprochen. Die Aufnahme des Eye-Trackers wurde dabei zur Unterstützung verwendet. Da diese Interviews den Rahmen dieses Kapitels sprengen würden und für die untersuchten Hypothesen nicht relevant sind, verzichten wir an dieser Stelle auf Details.

Materialien

Zur Untersuchung unserer Hypothesen wurden zwei Programme und zwei Algorithmen ausgewählt, die für die Versuchsdurchführung zentrale Eigenschaften aufwiesen:

1. Das zugehörige Programm enthält mindestens einen Programmabschnitt, der ein Planschema umsetzt und daher als „einfach“ klassifiziert werden kann.
2. Das zugehörige Programm enthält mindestens einen nichttrivialen Programmabschnitt, der kein Planschema umsetzt und daher als „komplex“ klassifiziert werden kann.
3. Diese Programmabschnitte sind getrennt voneinander dargestellt.

Problem 1: Rainfall Das zugehörige Programm löst eine Vereinfachung des *rainfall problem*, das von Soloway (1986) in einer Studie zum Programmverständnis genutzt wurde; dieses Programm wird zur Überprüfung von Hypothese 1 genutzt. Die hier verwendeten AOIs finden sich beispielhaft in Abb. 15.1.

Problem 2: Balanced Subset Das zweite Programm stammt – wie auch die verwendeten Algorithmen – aus einer Serie vergleichbarer Arbeiten (Ginat & Blau, 2017). Das diese Aufgabe lösende Programm wurde zur Überprüfung von Hypothese 1 genutzt.

Die beiden nächsten Algorithmen wurden genutzt, um Hypothese 1 (b), 2 und 3 zu untersuchen. Sie stammen aus derselben Serie von Arbeiten wie Balanced Subset. Zusätzlich ist jedoch noch eine Beweisskizze der Korrektheit bereitgestellt.

Problem 3: Fence Leveling Der erste betrachtete Algorithmus adressiert die Lösung des *fence-leveling problem* (Ginat, 2019a). Zusätzlich zum Programmcode wird eine Beweisskizze (Ginat, 2019b) bereitgestellt. Ein Screenshot des gesamten Materials findet sich in Abb. 15.2.

Problem 4: Longest Plateau Der zweite Algorithmus löst das *longest-plateau problem* (Ginat & Blau, 2017). Hier wurde ebenfalls neben dem Programmcode eine Beweisskizze bereitgestellt.

Operationalisierung der Überprüfung der Hypothesen

Um komplexe und einfache Programmabschnitte zu vergleichen, haben wir die oben diskutierten Programmabschnitte als AOIs aufgefasst und in gewissen Zeitfenstern die gesamte Fixationsdauer auf ihnen bestimmt. Beispiele für die AOIs finden sich in Abb. 15.1 und 15.2. Die Fixationsdauer wurde bereits von Crosby und Stelovsky (1990) als Maß für Aufmerksamkeit verwendet; zusätzlich haben Chen et al. (2011) gezeigt, dass sie als Maß für die wahrgenommene Komplexität geeignet ist. Eine höhere Komplexität bedeutet in diesem Zusammenhang, dass mehr Zeit zum Verständnis benötigt wird, da der gelesene Inhalt weniger naheliegend ist (Abschn. 15.3). Da die Abschnitte unterschiedlich viele Zeilen Programmtext umfassen, haben wir die Fixationsdauer normalisiert; hierfür wurde die Fixationsdauer auf der jeweiligen AOI durch die Anzahl der Zeilen an Programmtext dieser AOI (Klammern nicht mit eingerechnet) geteilt.

Wurden innerhalb eines der betrachteten Zeitfenster weniger als 85 % der Augenbewegungen erfasst, wurde dieser Datensatz nicht in die Analyse einbezogen. Dieser Grenzwert stellte sich als klarste Lücke heraus, bei der die Qualität der Daten noch hinreichend gut ist und gleichzeitig die geringste Anzahl an Daten ausgeschlossen werden muss. Für die Analyse der Experimente standen somit neun Datensätze zur Verfügung.

Da die Probandenzahl für eine quantitative Studie verhältnismäßig gering war, haben wir versucht, durch ein *within subjects*-Design die Qualität der Analysen zu erhöhen. Den Empfehlungen von Holmqvist et al. (2011, S. 71–79) und dem Vorgehen von Chitalkina et al. (2020) folgend betrachten wir Phasen mit hoher Konzentration. Hierzu analysieren wir jeweils einen kurzen Ausschnitt der Bearbeitungszeit der Programme und Algorithmen und vergleichen die normalisierte Verweildauer auf den einzelnen Abschnitten. Bei den Programmen umfasste dieser Zeitbereich den Zeitraum, bei dem die erste Eingabe für das Programm in eine Ausgabe umgewandelt wurde. Zu Beginn dieses Zeitraums ist das Programm noch unbekannt, und am Ende muss es hinreichend gut verstanden worden sein, damit eine Ausgabe angegeben werden konnte. Bei den Algorithmen umfasste dieser Bereich zum einen den „Programmlesen“-Zeitraum vom ersten Blick auf den Programmtext bis hin zum ersten Blick auf den Beweis und zum anderen den „Beweislesen“-Zeitraum vom ersten Blick auf den Beweis bis zum Anklicken der *Get Hint*-Schaltfläche. Da die Probandinnen und Probanden in allen Fällen nach Anklicken der Schaltfläche sinnvolle Zuordnungen zwischen Beweis und Programmtext vornahmen bzw. sinnvolle Ausgaben für die Programme angegeben haben und diese in nachgelagerten Interviews begründen konnten, gehen wir davon aus, dass zum Ende des zweiten Zeitraums ein hinreichendes Verständnis vorlag.

15.5 Ergebnisse

Zur Untersuchung von Hypothese 1 wurde eine *two-way repeated measures ANOVA* verwendet, für Hypothese 2 und 3 wurde zunächst die Normalverteilungsannahmen mithilfe eines Shapiro-Wilk-Tests überprüft, bevor dann ein *paired-samples t-test* durchgeführt wurde.

Hypothese 1

Zur Untersuchung von Hypothese 1 wurden die normalisierten Fixationsdauern auf den jeweils als „einfach“ bzw. „komplex“ klassifizierten Abschnitten verglichen (s. Abb. 15.3 für eine Gegenüberstellung aller Mittelwerte).

Zur Prüfung von Hypothese 1 (a) („Bei Programmen fokussieren sich fortgeschrittene Studierende überwiegend auf komplexen Programmtext“) und Hypothese 1 (b) („Bei Algorithmen fokussieren sich fortgeschrittene Studierende während des Programmcodelesens überwiegend auf komplexen Programmtext“) wurde ein 2×2 -Design verwendet. Hierbei wurden für die durchgeführte *two-way repeated measures ANOVA* als abhängige Variable die Verweildauer und als Faktoren die Art des gelesenen Programmabschnitts (einfach bzw. komplex) sowie die Art des Materials (Programm bzw. Algorithmus) modelliert.

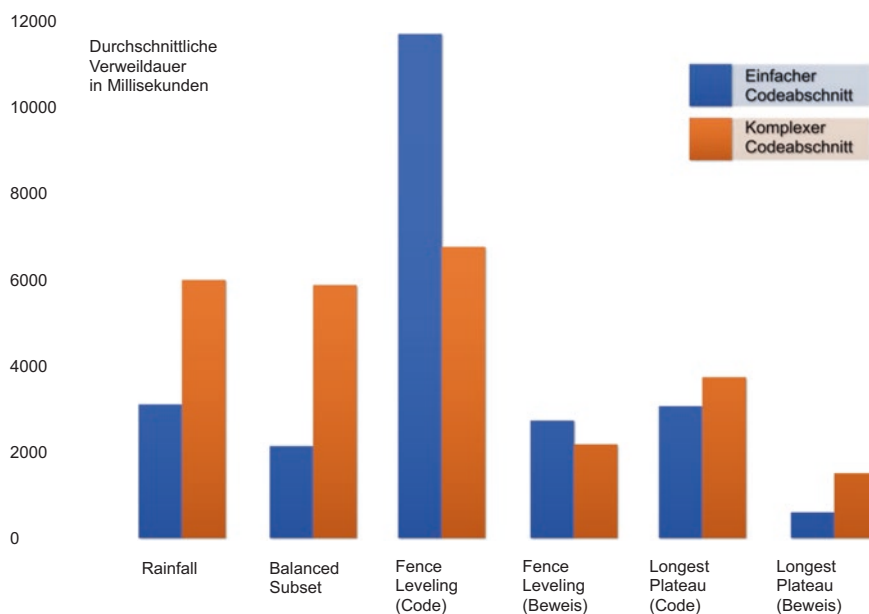


Abb. 15.3 Visualisierung der durchschnittlichen Verweildauer auf einfachen (blau) und komplexen (orange) Abschnitten des Programmcodes

Die nach Student angepassten Residuen für die Verweildauern auf einfachem und auf komplexem Programmtext bei Algorithmen waren mit einer schwach linksseitigen Ausrichtung nicht normalverteilt (Shapiro-Wilk-Test, $p = .019$ bzw. $p = .009$). Da sich die Hypothesen nur auf die relativen Verweildauern beziehen, wurde eine monotone Datentransformation für alle Daten mithilfe der Wurzelfunktion durchgeführt (Fernandez, 1992). Die Analyse der nach Student angepassten Residuen für diese transformierten Daten zeigte, dass die Normalverteilungsannahme in allen Fällen nun nicht mehr verletzt wurde (Shapiro-Wilk-Test, $p > .05$). Ebenso zeigte die Betrachtung von Residuen außerhalb von ± 3 Standardabweichungen, dass keine Ausreißer vorlagen. Für den Interaktionsterm wurde Sphärizität angezeigt (Mauchly-Test, $p > .05$).

Insgesamt zeigte sich eine statistisch signifikante Interaktion zwischen Art des Programmabschnitts (Komplexität) und Art des Materials (Algorithmus/Programm), $F(1, 17) = 26.27, p < .001, \eta^2 = .607$.

Nachfolgend wurde eine Analyse der einfachen Haupteffekte durchgeführt; wie in Abb. 15.3 werden alle Messwerte in nichttransformierter Form angegeben.

Beim Lesen von Programmen unterschied sich die auf einfachen Programmabschnitten ($M = 2651ms, SD = 901.49$) verbrachte Zeit signifikant von der auf komplexen Programmabschnitten ($M = 6003ms, SD = 2877.875$), $F(1, 17) = 55.32, p < 0.001$. Daher kann Hypothese 1 (a) nicht verworfen werden.

Beim Lesen des Programmcodes von Algorithmen dagegen unterschied sich die auf einfachen Programmabschnitten ($M = 7470ms, SD = 5658.770$) verbrachte Zeit nicht signifikant von der auf komplexen Programmabschnitten ($M = 5316ms, SD = 3665.267$) verbrachten Zeit, $F(1, 17) = 3.06, p = .099$. Daher muss Hypothese 1 (b) verworfen werden.

Zusammenfassung Während auf der Basis der von uns erfassten Daten Hypothese 1 (a) („Bei Programmen fokussieren sich fortgeschrittene Studierende überwiegend auf komplexen Programm“) nicht abgelehnt werden muss, sind die zu Hypothese 1 (b) („Bei Algorithmen fokussieren sich fortgeschrittene Studierende während des Programmcodelesens überwiegend auf komplexen Programmtext“) ausgewerteten Daten weniger klar und scheinen darauf hinzudeuten, dass gewisse Unterschiede im Leseverhalten von Programmen und Algorithmen bestehen.

Hypothese 2

Zur Untersuchung von Hypothese 2 („Bei Algorithmen fokussieren sich fortgeschrittene Studierende während des Programmlesens weniger stark auf komplexen Programmtext als beim Lesen von Programmen“) wurden die Verhältnisse der normalisierten Fixationsdauern während des Programmlesens betrachtet (ein Proband hatte extrem geringe Werte beim *fence leveling*-Problem, sodass dieser als Ausreißer dort nicht betrachtet wird). Ein höherer Wert bedeutet, dass der komplexe Programmtext relativ gesehen weniger Aufmerksamkeit erhalten hat. Tab. 15.1 zeigt, dass dieses Verhältnis für Programme geringer als für

Tab. 15.1 Vergleich der Verhältnisse der normalisierten Fixationsdauern (Mittelwerte mit Standardabweichungen in Klammern) auf einfachem bzw. komplexen Programmtext während des Programmlesens. Die Werte wurden für jedes Vergleichspaar separat normalisiert und nach Überprüfung der Normalverteilungsannahmen mittels eines t-Tests verglichen.

Programm	Algorithmus	t	p	d
Balanced Subset .407 (.198)	Longest Plateau .917 (.289)	t(8) = 4.786	.001	1.595
Rainfall .529 (.398)	Longest Plateau .917 (.289)	t(8) = 2.102	.069	.701
Balanced Subset .391 (.205)	Fence Leveling 1.762 (.970)	t(7) = 3.651	.008	1.291
Rainfall .550 (.420)	Fence Leveling 1.762 (.970)	t(7) = 4.153	.004	1.468

Algorithmen ist. Bis auf einen Fall sind diese Unterschiede immer signifikant; die Hypothese kann also auf Basis unserer Daten nicht abgelehnt werden.

Hypothese 3

Zur Untersuchung von Hypothese 3 („Fortgeschrittene Studierende nutzen die zu einem Algorithmus gegebene Beweisskizze, um sich komplexen Programmtext zu erschließen“) wurden innerhalb der betrachteten Algorithmen die Quotienten der Fixationsdauern auf einfachem bzw. komplexem Programmtext beim Programmlesen und beim Beweislesen verglichen.

Hypothese 3: Fence Leveling Nach Entfernung eines Ausreißers verletzten die Daten ($n = 8$) die Normalverteilungsannahme nicht ($p = .287$). Der Quotient zeigt, dass der Fokus während des Beweislesens mehr auf dem komplexen Programmtext liegt ($M = .723, SD = .695$) als beim Programmlesen ($M = 1.762, SD = .970$); dies ist jedoch aufgrund der hohen Varianzen nicht signifikant, $t(7) = 2.307, p = .054, d = .816$.

Hypothese 3: Longest Plateau Wie oben erwähnt, wurden die Daten dreier Probanden, die den Programmtext während des Beweislesens gar nicht beachtet haben, bei der Analyse ausgeschlossen. Die Differenzen der verbleibenden Quotienten ($n = 6$) verletzten die Normalverteilungsannahme nicht ($p = .993$). Hier zeigte sich der schon beim *fence leveling*-Algorithmus beobachtete, aber ebenfalls nicht signifikante Effekt: Der Quotient während des Beweislesens ($M = .152, SD = .890$) war geringer als beim Programmlesen ($M = .944, SD = .304, t(5) = 1.978, p = .105, d = .808$).

Zusammenfassung Auf der Basis der von uns analysierten Daten kann Hypothese 3 zwar in der Tendenz, nicht aber statistisch bestätigt werden.

15.6 Diskussion

Gegenstand dieser Studie war die Untersuchung, wie unterschiedlich komplexe Programmabschnitte in Programmen und Algorithmen von fortgeschrittenen Studierenden fokussiert werden. Die Ergebnisse zeigen, dass die Erkenntnisse von Crosby und Stelovsky (1990) auch in unserem Studiendesign und für unsere Studiengruppe zutreffen: Bei Programmen haben sich unsere Probandinnen und Probanden eher auf komplexen als auf einfachen Programmtext konzentriert.

Bei Algorithmen sind diese Unterschiede weniger deutlich. Aufgrund der geringen Stichprobengröße sind diese Ergebnisse zwar mit Vorsicht zu interpretieren, aber sie erlauben es, die Gleichheit von Algorithmenlesen und Programmlesen per se zu hinterfragen: Tatsächlich hat sich gezeigt, dass sich unsere Probandinnen und Probanden beim Lesen von Programmtext bei Algorithmen weniger stark auf komplexen Programmtext fokussieren als bei Programmen.

In einer qualitativen Studie haben wir beobachten können, dass Probandinnen und Probanden den Beweis oder auch einen beschreibenden Text nutzen, um Hypothesen über den Programmtext zu generieren, die anschließend nur noch verifiziert werden müssen (Kather & Vahrenhold, 2021).

Insbesondere werden solche zusätzlichen Materialien genutzt, um Schwierigkeiten beim Verständnis komplexer Programmabschnitte zu kompensieren. Eine quantitative Bestätigung konnte in der hier betrachteten Studien zwar in der Tendenz, nicht aber signifikant erzielt werden.

Mögliche Kritikpunkte an der Validität der Ergebnisse können trotz der Verwendung einer Zielstichprobe an der geringen Zahl an Probandinnen und Probanden festgemacht werden. Zudem ist festzuhalten, dass die Probandinnen und Probanden nur von einer einzigen Institution stammten und somit eine mögliche Verallgemeinerung der Erkenntnisse mit äußerster Vorsicht betrachtet werden muss.

Zusammenfassend lässt sich jedoch festhalten, dass die in dieser Studie erhobenen quantitativen Daten andeuten, dass ein Leseverhalten, wie es aus dem reinen Programmverständnis bekannt ist, nicht zwingend beim Lesen von Algorithmen zu finden ist. In Kombination mit den qualitativ erzielten Erkenntnissen (Kather & Vahrenhold, 2021) scheint es somit nicht ausreichend zu sein, das Verständnis von Algorithmen auf Basis disjunkter Erkenntnisse aus dem Programm- und Beweisverständnis erklären zu wollen. Es ergibt sich daher die Notwendigkeit weiterer Studien sowie des Einsatzes qualitativer Methoden, die die Nutzung von Eye-Tracking-Technologien flankieren. Ein möglicher Anknüpfungspunkt unserer Arbeit für die interventionsbezogene Forschung stellt die Untersuchung sog. *subgoal labels* (Catrambone, 1998) dar, die sich als für das Programmverständnis als förderlich erwiesen haben (Morrison et al., 2020) und deren Einsatz auch in den die Programme flankierenden Beweisen eines Algorithmus untersucht werden sollte.

Danksagung Wir danken R. Duran und M. Puurtinen für ihre Beratung sowie S. Schukajlow-Wasjutinski für die Erlaubnis zur Nutzung seines Eye-Tracking-Labors. Während des Begutachtungsprozesses sind viele hilfreiche Kommentare in die Überarbeitung dieses Kapitels eingeflossen, für die wir ebenfalls dankbar sind.

Literatur

- Ben-David Kolikant, Y. (2005). Students' alternative standards for correctness. *International Computing Education Research Workshop*, 2005, 37–43. <https://doi.org/10.1145/1089786.1089790>
- Catrambone, R. (1998). The subgoal labeling model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, 127(4), 355–376.
- Chen, S., Epps, J., Ruiz, N., & Chen, F. (2011). Eye activity as a measure of human mental efforts in HCI. *Proceedings of the 16th International Conference on Intelligent User Interfaces*, 315–318. <https://doi.org/10.1145/1943403.1943454>.
- Chitalkina, N., Bednarik, R., Puurtinen, M., & Gruber, H. (2020). When you ignore what you see: How to study proof-readers' error in pseudocode reading. *ACM Symposium on Eye Tracking Research and Applications, Short Papers*, 53(51–53), 55. <https://doi.org/10.1145/3379156.3391979>
- Crosby, M. E., & Stelovsky, J. (1990, January). How do we read algorithms? A Case Study. *IEEE Computer*, 23(1), 24–35. <https://doi.org/10.1109/2.48797>.
- da Rosa, S. (2010). The Construction of the concept of binary search algorithm. *Proceedings of the 22th Annual Workshop of the Psychology of Programming Interest Group*, 100–111.
- da Rosa, S. (2015). The construction of knowledge of basic algorithms and data structures by novice learners. *Proceedings of the 26th Annual Workshop of the Psychology of Programming Interest Group*, 37–48.
- Duran, R. (2020). *Cognitive complexity of comprehending computer programs* [PhD Thesis, School of Science, Aalto University]. <http://urn.fi/URN:ISBN:978-952-60-3944-2>.
- Fernandez, G. C. J. (1992, April). Residual analysis and data transformations: Important tools in statistical analysis. *HortScience*, 27(4), 297–300. <https://doi.org/10.21273/HORTSCI.27.4.297>.
- Ginat, D. (2019a, March). COLORFUL CHALLENGES: Fence Leveling. *ACM Inroads*, 10(1), 28–29. <https://doi.org/10.1145/3306135>.
- Ginat, D. (2019b, September). COLORFUL CHALLENGES: Letter Game. *ACM Inroads*, 10(3), 19–20. <https://doi.org/10.1145/3341096>.
- Ginat, D., & Blau, Y. (2017). Multiple levels of abstraction in algorithmic problem solving. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, 237–242. <https://doi.org/10.1145/3017680.3017801>.
- Haberman, B., Averbuch, H., & Ginat, D. (2005). Is it really an algorithm: The need for explicit discourse. *Proceedings of the 10th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education*, 74–78. <https://doi.org/10.1145/1067445.1067469>.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Just, M. A., & Carpenter, P. A. (1980, July). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>.
- Kather, P., & Vahrenhold, J. (2021). Is algorithm comprehension different from program comprehension? *Proceedings of the 29th IEEE/ACM International Conference on Program Comprehension*, 455–466. <https://doi.org/10.1109/ICPC52881.2021.00053>.
- Morrison, B. B., Margulieux, L. E., & Decker, A. (2020). The curious case of loops. *Computer Science Education*, 1–28. <https://doi.org/10.1080/08993408.2019.1707544>.

- Narayanan, N. H., & Hergarty, M. (2002, October). Multimedia design for communication of dynamic information. *International Journal of Human-Computer Studies*, 57(4), 279–315. <https://doi.org/10.1006/ijhc.2002.1019>
- Obaidallah, U., Al Haek, M., & Cheng, P. C.-H. (2018, January). A survey on the usage of eye-tracking in computer programming. *ACM Computing Surveys*, 51(1), 5:1–5:58. <https://doi.org/10.1145/3145904>.
- Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Hrsg.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence and education* (S. 33–58). L. Erlbaum Associates.
- Soloway, E. (1986, September). Learning to program = Learning to construct mechanisms and explanations. *Communications of the ACM*, 29(9), 860–858. <https://doi.org/10.1145/6592.6594>.
- Soloway, E., & Ehrlich, K. (1984, September). Empirical studies of programming knowledge. *IEEE Transactions on Software Engineering*, SE-10(5), 595–609. <https://doi.org/10.1109/TSE.1984.5010283>.
- Soloway, E., Bonar, J., & Ehrlich, K. (1983, November). Cognitive structures and looping constructs: An empirical study. *Communications of the ACM*, 26(11), 853–860. <https://doi.org/10.1145/182.358436>.
- Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020, June). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, 104(2), 147–200. <https://doi.org/10.1007/s10649-020-09948-1>.
- Tobii AB. (2020). *What's new in Tobii Pro Lab? | Learn More*. <https://www.tobii.com/learn-and-support/learn/tobii-pro-lab/>.
- Velázquez-Iturbide, J. Á. (2013, November). An experimental method for the active learning of greedy algorithms. *ACM Transactions on Computing Education*, 13(4), 18:11–18:23. <https://doi.org/10.1145/2534972>.
- von Mayrhauser, A., & Vans, A. M. (1995, August). Program comprehension during software maintenance and evolution. *IEEE Computer*, 28(8), 44–55. <https://doi.org/10.1109/2.402076>.
- Yang, K.-L., & Lin, F.-L. (2008, January). A Model Of Reading Comprehension of Geometry Proof. *Educational Studies in Mathematics*, 67(1), 59–76. <https://doi.org/10.1007/s10649-007-9080-6>.



Willkommen zu den Springer Alerts

Unser Neuerscheinungs-Service für Sie:
aktuell | kostenlos | passgenau | flexibel

Mit dem Springer Alert-Service informieren wir Sie individuell und kostenlos über aktuelle Entwicklungen in Ihren Fachgebieten.

Abonnieren Sie unseren Service und erhalten Sie per E-Mail frühzeitig Meldungen zu neuen Zeitschrifteninhalten, bevorstehenden Buchveröffentlichungen und speziellen Angeboten.

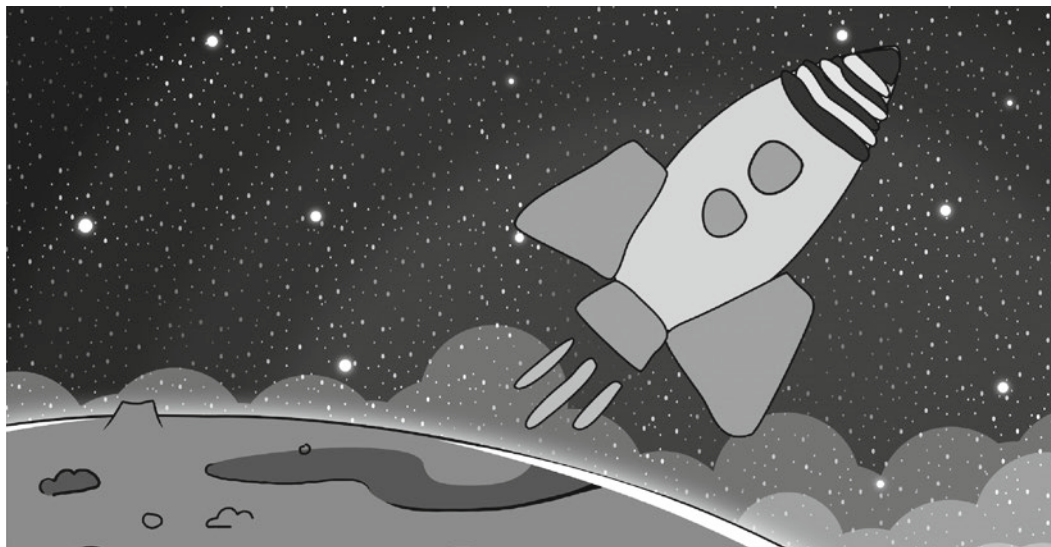
Sie können Ihr Springer Alerts-Profil individuell an Ihre Bedürfnisse anpassen. Wählen Sie aus über 500 Fachgebieten Ihre Interessensgebiete aus.

Bleiben Sie informiert mit den Springer Alerts.

Jetzt
anmelden!

Mehr Infos unter: springer.com/alert

Part of **SPRINGER NATURE**



Auf zu neuen Lernwelten - mit Springer Spektrum!

Die neue Springer Lehrbuchplattform für dein Physikstudium

Mit Lexikon, Prüfungsfragen und Videos für das Bachelor- und Masterstudium

- Du willst dich online auf deine Prüfungen vorbereiten?
- Du hast das passende Lehrbuch noch nicht gefunden?
- Du stehst kurz vor einer Prüfung, hast aber noch keinen Überblick?

Kein Problem: Auf lehrbuch-physik.springer.com haben wir wichtige Springer Spektrum-Lehrbücher zu allen Themen der Physik zusammengestellt! Dazu gibt es umfangreiche Materialien und Lernangebote: Lernkarten zu den einzelnen Lehrbüchern, ein umfangreiches Lexikon, interessante, thematisch passende Videos, unsere Facebook News und vieles mehr! Für Dozentinnen und Dozenten gibt es exklusiv vorbereitete Präsentationen mit Abbildungen passend zur Vorlesung.



lehrbuch-physik.springer.com

Part of **SPRINGER NATURE**