Der Einfluss von Stress auf Prozesse beim diagnostischen Urteilen – eine Eye Tracking-Studie mit mathematischen Textaufgaben

Sara Becker · Birgit Spinath · Beate Ditzen · Tobias Dörfler

Online publiziert: 6. Juli 2020
© Der/die Autor(en) 2020

Zusammenfassung In realen Unterrichtssituationen diagnostizieren Lehrkräfte aufgrund enormer Interaktionsdichten sowie hoch komplexer Unterrichtssituationen unter psychischer Anspannung wie Stress. Durch Stress können kognitive Kapazitäten gebunden werden, was Prozesse beim Diagnostizieren, wie die Wahrnehmung und die Interpretation von Aufgabenschwierigkeiten, beeinträchtigen kann.

Die vorliegende Studie untersuchte in einem experimentellen Design den Einfluss von Stress auf Wahrnehmungs- und Interpretationsprozesse beim diagnostischen Urteilen. Hierzu schätzten $N=64$ Mathematiklehramtsstudierende die Schwierigkeit von acht mathematischen Textaufgaben im Bereich von Brüchen ein. Die Wahrnehmungsprozesse wurden mithilfe von Blickbewegungen durch die Methode des Eye Trackings erhoben und anhand ausgewählter globaler Maße (Dauer des Blickpfades, Fixationsdauer, Anzahl der Fixationen ab 250 ms und Sakkadenamplitude) analysiert. In Kombination mit Verbalprotokollen wurden die Interpretationsprozesse ausgewertet. Anhand der Analyse von Cortisolwerten aus Speichelproben während des Diagnostizierens wurde das Stressniveau der Teilnehmenden objektiv erfasst. Aufgrund der Bindung kognitiver Kapazitäten durch Stress wurden Veränderungen sowohl in den Wahrnehmungs- als auch in den Interpretationsprozessen erwartet. Die Ergebnisse der Blickbewegungsanalysen zeigten eine kürzere Blickpfaddauer, eine kürzere Fixationsdauer sowie eine geringere Anzahl an Fixationen mit einer Dauer ab 250 ms bei Teilnehmenden unter Stresserleben. Die Verbalprotokolle un-

S. Becker (✉) · T. Dörfler
Institut für Psychologie, Pädagogische Hochschule Heidelberg, Keplerstraße 87, 69120 Heidelberg, Deutschland
E-Mail: becker@ph-heidelberg.de

B. Spinath
Psychologisches Institut, Universität Heidelberg, Heidelberg, Deutschland

B. Ditzen
Institut für Medizinische Psychologie, Universitätsklinikum Heidelberg, Heidelberg, Deutschland

terschieden sich qualitativ zwischen den Gruppen. Die veränderten Maße können als Indikatoren der veränderten Wahrnehmungs- und Interpretationsprozesse der Experimentalgruppe interpretiert werden.

Die Studie gibt erste Hinweise auf den Einfluss von Stress auf diagnostische Prozesse im anwendungsbezogenen Schulkontext.

Schlüsselwörter Diagnostische Prozesse · Stress · Eye Tracking · Textaufgaben

The influence of stress on processes of diagnostic judgement—an eye tracking study based on mathematical word problems

Abstract In real-life teaching situations, teachers make diagnoses under psychological tension like stress, based on high interaction densities as well as on highly complex teaching situations. Stress can bind cognitive capacities which could impair diagnostic processes, such as the perception and the interpretation of task difficulties.

For this purpose, $N=64$ student teachers of mathematics estimated the difficulty of eight mathematical word problems in the area of fractions. The perceptual processes were documented with the help of eye tracking and analyzed using global measures (duration of gaze path, fixation duration, number of fixations with a duration of 250 ms and saccade amplitudes). The interpretation processes were evaluated in combination with verbal protocols. Based on the analysis of cortisol values from saliva samples during diagnosis, the stress level of the participants was objectively assessed. Due to the binding of cognitive capacities by stress experience, changes in perception, which can be shown by global eye movement measures, and interpretation processes, which can be shown by verbal protocols, were expected.

Results of eye movement analyses showed a shorter scanpath duration, a shorter fixation duration and a lower number of fixations with a duration of 250 ms or more for participants under stress experience. The verbal protocols differed in quality between groups. The altered measures can be interpreted as indicators of altered perception and interpretation processes of the experimental group.

The study provides initial indications of the influence of stress on diagnostic processes in an application-related school context.

Keywords Diagnostic processes · Stress · Eyetracking · Word problems

1 Einleitung

Die kognitiven Anforderungen verschiedener Lernaufgaben für Schülerinnen und Schüler fortlaufend zutreffend beurteilen zu können, erfordert diagnostisch kompetente Lehrkräfte (Schrader und Helmke 1987; Weinert 2000). Insbesondere in Unterrichtsfächern wie der Mathematik, in denen Aufgaben den Unterricht dominieren und deren Bearbeitung durch die Schülerinnen und Schüler einen Großteil der Unterrichtszeit einnehmen (Hiebert et al. 2003), bildet eine akkurate Schwierigkeitseinschätzung von Aufgaben durch die Lehrkraft die Grundlage, ein optimales Anforderungsniveau des Unterrichts für die Lerngruppe zu erreichen (Leuders et al.

2018). Die Schwierigkeitseinschätzung von Aufgaben wird meist der adaptiven Planungskompetenz einer Lehrkraft zugeordnet (Brühwiler 2014). Lehrkräfte müssen jedoch auch während des Unterrichtsverlaufs Aufgaben zutreffend interpretieren und im Hinblick auf den Unterrichtsprozess sowie die jeweilige Anforderungssituation einschätzen (Bromme und Haag 2008). In realen Unterrichtssituationen diagnostizieren Lehrkräfte oftmals unter Zeitdruck oder aufgrund hoch komplexer Unterrichtssituationen sowie enormer Interaktionsdichten unter psychischer Anspannung. Es ist bereits gut dokumentiert, dass Lehrkräfte im Unterrichtsgeschehen einem erheblichen Stresspotential ausgesetzt sind (u. a. Rothland und Terhart 2007; Scheuch und Knothe 1997). Ergebnisse aus der Belastungs- und Stressforschung belegen, dass physische und insbesondere psychische Zustände, wie Stress, kognitive Kapazitäten binden (Chen und Chaiken 1999). Diese Kapazitäten stehen – so die Annahme – für diagnostische Prozesse, wie die Wahrnehmung und die Interpretation von Aufgabenschwierigkeiten, nicht mehr zur Verfügung (Het et al. 2005). Dennoch besteht wenig empirische Evidenz zum möglicherweise verzerrenden Effekt von Stress auf Diagnoseprozesse im anwendungsbezogenen Schulkontext.

Die vorliegende Studie untersucht in einem experimentellen Design den Einfluss von Stress auf Prozesse beim Diagnostizieren. Mithilfe der Methode des Eye Trackings in Kombination mit Verbalprotokollen wird der Einfluss von Stress auf Wahrnehmungs- und Interpretationsprozesse beim diagnostischen Urteilen fokussiert. Es wird angenommen, dass sich die Bindung kognitiver Kapazitäten durch Stresserleben in der Veränderung ausgewählter globaler Blickbewegungsmaße zeigen wird, die insbesondere Aussagen über die Wahrnehmung und die Verarbeitung von Aufgaben als Ganzes widerspiegeln. Die veränderte Interpretation der wahrgenommenen Aufgabenschwierigkeiten unter Stress wird sich – so die Annahme der vorliegenden Studie – in einer geringeren Qualität der Verbalprotokolle zeigen.

2 Theoretischer Hintergrund diagnostischer Urteile im Mathematikunterricht

In einem Großteil der bisherigen empirischen Studien zur diagnostischen Kompetenz von Lehrkräften wird die Qualität diagnostischer Urteile als Urteilsakkuratheit operationalisiert. Die Urteilsakkuratheit beschreibt das Maß der Übereinstimmung von Lehrkräfteurteilen mit tatsächlichen Merkmalsausprägungen, wie zum Beispiel empirisch erhobenen Aufgabenschwierigkeiten (Hoge und Coladarei 1989; Südkamp et al. 2012). Metaanalysen belegen jedoch eine erhebliche Streuung der interindividuellen Lehrkräfteurteile (Südkamp et al. 2012). Südkamp et al. (2012) berichten in ihrer Untersuchung, dass sich zwei Drittel der Varianz der Urteile hinsichtlich der Schülerergebnisse nicht alleine durch die Schülerleistungen selbst erklären lassen, sondern auf personale Bedingungen der Lehrkräfte während des Diagnostizierens zurückzuführen sind.

Studien untersuchten bereits die Auswirkungen von verschiedenen Einflussfaktoren, insbesondere von stabilen Personencharakteristika, auf die Akkuratheit diagnostischer Urteile. Befunde im Bereich der Mathematikdidaktik konnten den positiven Einfluss von zeitlich stabilen Personencharakteristika einer Lehrkraft, wie

das Fachwissen (=content knowledge, kurz CK) und das fachdidaktische Wissen (=pedagogical content knowledge, kurz PCK), in Anlehnung an Shulman (1986, 1987), auf die Akkuratheit diagnostischer Urteile aufzeigen (Brunner et al. 2011; Karing 2009; Kunter et al. 2011). So konnte die Interventionsstudie von Ostermann et al. (2017) eine gesteigerte Akkuratheit der Lösungsrate zukünftiger Lehrkräfte durch die Vermittlung von PCK nachweisen. In anderen Fachdidaktiken, wie zum Beispiel Deutsch, konnte jedoch kein positiver Einfluss von verschiedenen Wissenskomponenten der Lehrkraft auf die Urteilsakkuratheit gefunden werden (Rausch et al. 2015). Die in Metaanalysen berichtete erhebliche interindividuelle Varianz der Urteilsakkuratheit zwischen den Lehrkräften konnte durch zeitlich stabile Personencharakteristika, wie PCK, demnach nicht hinreichend über alle Domänen hinweg erklärt werden (Helmke et al. 2004; Hoge und Coladarcì 1989; Spinath 2005; Schrader 2013). Situativ aktivierte Personencharakteristika einer Lehrkraft, wie Stress oder Anspannung, hervorgerufen beispielsweise durch das Unterrichtsgeschehen selbst, wurden in Studien zur Diagnosekompetenz bisher weitgehend außer Acht gelassen.

Erklärungsansätze für die heterogene Befundlage sowie die berichtete interindividuelle Varianz bisheriger Forschungsergebnisse könnten neben den Einflussfaktoren in der Genese diagnostischer Urteile sowie deren Einflussfaktoren auf den Diagnoseprozess begründet liegen. Wie diagnostische Urteile entstehen und welche Faktoren die Genese bedingen, ist jedoch weitgehend ungeklärt. Aufgrund dessen entwickelte sich neben dem produktorientierten Ansatz der Urteilsakkuratheit ein neues Forschungsgebiet zur Untersuchung der kognitiven Prozesse während des Diagnostizierens. Obwohl erste Modelle zur Konzeptualisierung vorliegen (vgl. Herppich et al. 2018) sowie empirische Studien zur Untersuchung der kognitiven Prozesse beim Diagnostizieren durchgeführt wurden (vgl. Ostermann et al. 2017), kann der Forschungsstand über die zugrundeliegenden kognitiven Prozesse beim diagnostischen Urteilen immer noch als unbefriedigend angesehen werden (Leuders et al. 2018; Loibl et al. 2020). Es fehlte bislang ein studienübergreifender Rahmen, der zum einen den Fokus auf die kognitiven Prozesse beim Diagnostizieren richtet, sowie zum anderen die möglichen Einflussfaktoren der diagnostizierenden Lehrkraft sowie der Diagnosesituation berücksichtigt.

Das Modell DiaCoM (*Explaining Teachers' Diagnostic Judgements by Cognitive Modeling*) (Loibl et al. 2020) bietet hier eine Rahmung für die theoretische Konzeptualisierung und die experimentelle Untersuchung kognitiver Prozesse beim diagnostischen Urteilen von Lehrkräften. Es wird auch als theoretischer Rahmen für die vorliegende Studie herangezogen. Das Rahmenmodell DiaCoM umfasst vier Komponenten: a) die *Situationscharakteristika* einer diagnostischen Situation, die den inhaltlichen Rahmen der Situation sowie Hinweisreize, zum Beispiel von Aufgaben umfassen, b) die *Personencharakteristika* einer diagnostizierenden Person, die in zeitlich stabile Dispositionen einer Person (*traits*), wie zum Beispiel PCK und CK, sowie in situativ aktivierte Zustände einer Person (*states*), wie zum Beispiel Stresserleben, unterschieden werden können, c) das *Diagnostische Denken*, das den Informationsverarbeitungsprozess einer Diagnose beschreibt, sowie d) das *Diagnostische Verhalten*, das sich aus den Indikatoren für den zugrundeliegenden Prozess sowie den Ergebnisindikatoren zusammensetzt. Durch die experimentelle Variation der *Situationscharakteristika* und der *Personencharakteristika* sowie aufgrund von

Abb. 1 Für die vorliegende Studie spezifiziertes DiaCoM-Rahmenmodell. (vgl. Loibl et al. 2020)

weiterführenden theoretischen Annahmen können die Auswirkungen dieser zwei Komponenten auf das *Diagnostische Denken* (sowie das *Diagnostische Verhalten*) unter Nutzung des Rahmenmodells in empirischen Studien überprüft werden.

Das Rahmenmodell DiaCoM wird für die empirische Untersuchung der vorliegenden Studie spezifiziert (siehe Abb. 1): Die Diagnosesituation der vorliegenden Studie bilden mathematische Textaufgaben im Bereich der Bruchrechnung. Die Schwierigkeiten der Textaufgaben werden über schwierigkeitsgenerierende Aufgabenmerkmale (Hinweisreize) operationalisiert. Fokussiert wird in der vorliegenden Studie der Einfluss des *situativ aktivierten Personencharakteristikums* Stress auf das *Diagnostische Denken*, bestehend aus der Wahrnehmung sowie der Interpretation von Aufgabenschwierigkeiten für Schülerinnen und Schüler. Blickbewegungen, aufgezeichnet mithilfe der Methode des Eye Trackings, stellen Indikatoren für die zugrundeliegenden (latenten) Wahrnehmungsprozesse des diagnostischen Denkens in der vorliegenden Studie dar. Verbalprotokolle ermöglichen Rückschlüsse über die (latenten) Interpretationsprozesse wahrgenommener Hinweisreize von Aufgabenschwierigkeiten durch die diagnostizierende Person. Die einzelnen Komponenten sowie die Operationalisierungen der vorliegenden Studie werden im Folgenden differenziert dargestellt.

2.1 Stress als *Personencharakteristikum* im Urteilsprozess

Der Fokus der vorliegenden Studie liegt auf dem Einfluss des situativ aktivierten Personencharakteristikums Stress auf die Informationsverarbeitungsprozesse, bestehend aus den Wahrnehmungs- sowie den Interpretationsprozessen, während einer Diagnoseaufgabe. Stress entsteht nach Steckler et al. (2005), wenn ein Stressor in Form eines externen oder internen Reizes vorliegt, welcher vom Individuum als aversiv erlebt wird und eine adäquate Reaktion zur Regulation des Individuums erfordert. Die Stressreaktion äußert sich dabei nicht nur behavioral und emotional, sondern auch kognitiv und physiologisch (Steckler et al. 2005). Die physiologische Stress-

reaktion zeigt sich unter anderem in einer schnellen Aktivierung des sympathischen Nervensystems (De Kloet et al. 2005), in der Erhöhung der Herzrate, des Blutdrucks sowie der Muskelaktivität (Birbaumer und Schmidt 2010). Um dem Körper die für die Stressreaktion benötigte Energie zu liefern wird zusätzlich die Aktivität der Hypothalamus-Hypophysen-Nebennierenrinden-Achse (HHNA) durch das Gehirn angeregt und damit unter anderem eine gesteigerte Ausschüttung des Stresshormons Cortisol aktiviert (Dickerson und Kemeny 2004). Da Stress eine latente Variable darstellt, die nicht direkt von außen beobachtbar ist, werden die in Untersuchungen erfassten Anstiege des Cortisolspiegels als Indikator für eine zugrundeliegende Stressreaktion des Individuums angesehen (Kirschbaum und Hellhammer 1999).

Studien, in denen Auswirkungen von Stress auf kognitive Funktionen untersucht wurden, belegen, dass physische und insbesondere psychische Anspannungen kognitive Kapazitäten binden, die zur Erfüllung einer Aufgabe nicht oder nur bedingt genutzt werden können (Het et al. 2005; Schoofs et al. 2009). So konnte in einer Studie von Schoofs et al. (2009) der Einfluss eines schmerzbezogenen Stressfaktors auf die kognitiven Kapazitäten aufgezeigt werden. Der Stressfaktor aktivierte eine verstärkte Ausschüttung von Stresshormonen, die die verfügbaren kognitiven Kapazitäten verringerten.

Zur Stressinduktion im Labor-Setting arbeiten Studien zur Auswirkung von Stress auf kognitive Funktionen häufig mit dem „Sozialevaluativen Kaltwasserstresstests“ (engl. „*social evaluated cold-pressor test*“, SECPT). Der SECPT gilt als zuverlässige Methode zur endogenen Stressinduktion, indem das sympathische Nervensystem sowie die Hypothalamus-Hypophysen-Nebennierenrinden-Achse (HHNA) stimuliert werden (Schwabe et al. 2008). Er beruht auf dem weit verbreiteten und standardisierten „Kaltwasserstresstest“ (engl. „*cold-pressor test*“, CPT) und wurde durch sozialevaluative Komponenten, wie die Beobachtung durch eine weibliche Versuchsleiterin und die Aufzeichnung durch eine Videokamera, erweitert (al’Absi et al. 2002). Die Ergebnisse der Metaanalyse von Dickerson und Kemeny 2004 belegen die Schlüsselfunktion von sozialevaluativer Bewertung einerseits und dem Gefühl der Unkontrollierbarkeit der Stresssituation andererseits für die Aktivierung der HHNA und einer dadurch signifikant gesteigerten Ausschüttung des Stresshormons Cortisol.

Entsprechende physiologische Reaktionen sind in Unterrichtssituationen, in denen eine Lehrkraft unter Anspannung und Stress eine adäquate Einschätzung treffen muss, plausibel.

2.2 Informationsverarbeitungsprozesse als *Diagnostisches Denken*

Das *Diagnostische Denken* beschreibt latente, von außen nicht zugängliche Wahrnehmungs- und Interpretationsprozesse von Informationen beim diagnostischen Urteilen. Der Fokus auf die kognitiven Prozesse im Sinne eines Informationsverarbeitungsparadigmas ist im Rahmen der Bildungsforschung aus Konzepten wie *professional vision* (Seidel et al. 2010) sowie *noticing and knowledge-based reasoning* bekannt. Unter *noticing* wird hierbei die selektive Wahrnehmung von lernrelevanten Ereignissen und Merkmalen in professionellen Situationen, wie dem Unterrichten, verstanden. Die wahrgenommenen Merkmale müssen im Anschluss von der

Lehrperson vor dem Hintergrund ihres Wissens und ihrer kognitiven Ressourcen als relevant eingeschätzt sowie wissensgestützt interpretiert und verarbeitet werden (*knowledge-based reasoning*). Daran anlehnend beschreibt das *Diagnostische Denken* im zugrundeliegenden DiaCoM-Modell den Informationsverarbeitungsprozess als Zusammenspiel der Komponenten *Wahrnehmen* sowie *Interpretieren* und ergänzt diese durch den Bereich des *Entscheidens*, welcher in der vorliegenden Studie nicht näher betrachtet wird. Das Wahrnehmen und das Interpretieren von relevanten Aufgabenmerkmalen stellen Informationsverarbeitungsprozesse dar, die die Grundlage bilden, eine Aufgabenschwierigkeit adäquat vorherzusagen.

2.3 Blickbewegungen sowie Verbalprotokolle als Prozessindikatoren des *Diagnostischen Verhaltens*

Das Aufzeichnen von Blickbewegungen (Eye Tracking) stellt in der Leseforschung bereits seit vielen Jahren eine weit verbreitete objektive Messmethode dar, um Wahrnehmungsprozesse abzubilden (Engbert und Kliegl 2001; Rayner et al. 2012; Reichle et al. 2003). Lesen beschreibt die Fähigkeit, schriftliche Informationen visuell wahrzunehmen sowie hinsichtlich der Aufgabenstellung zu interpretieren (Rayner et al. 2012). Blickbewegungen stellen somit Schlüsselkomponenten der Informationswahrnehmung und der Informationsverarbeitung dar. In den vergangenen 40 Jahren wurde die Methode des Eye Trackings mit zunehmender Intensität auch für die Mathematikdidaktik interessant. Verschiedene Studien untersuchten bereits die Abhängigkeit der Blickbewegungen von situativen Aufgabenmerkmalen, wie sprachlichen oder mathematischen Merkmalen (van der Schoot et al. 2009; Verschaffel et al. 1992). Darüber hinaus konnte in einer Studie von Inglis und Alcock (2012) der Einfluss von stabilen Personencharakteristika in Form von Expertise beim Beurteilen von Beweisen aufgezeigt werden. Wenig empirische Evidenz ist über den Einfluss von situativ aktivierten Personencharakteristika, wie Stress, auf Blickbewegungen beim Einschätzen von Aufgaben verfügbar.

Die Analyse von Eye Tracking-Daten basiert zum einen auf der *eye-mind* Hypothese, die postuliert, dass Wörter so lange mit dem Auge fixiert werden, wie das Wort kognitiv verarbeitet wird, sowie zum anderen auf der *immediacy* Hypothese, die besagt, dass visuell wahrgenommene Reize in der Regel unmittelbar kognitiv verarbeitet werden (Just und Carpenter 1980). Die Blickbewegungen sind gekennzeichnet durch spezifische Blickdauern auf definierte Blickorte (Fixationen) sowie durch eine spezifische Anzahl und Abfolge von Blicksprüngen zwischen den Blickorten (Sakkaden), die als Blickpfad zusammengefasst werden.

Während den Fixationen findet der Prozess der visuellen Wahrnehmung und der Informationsaufnahme statt. Die Anzahl sowie die Dauer der Fixationen können Auskunft darüber geben, welche Inhalte mit welcher Intensität fixiert werden. Die einzelnen Fixationen sind durch Blicksprünge (Sakkaden), die meist zwischen 30 bis 100 Millisekunden dauern, verbunden. Während der Sakkaden bewegt sich das Auge so schnell, dass das visuelle Wahrnehmungsvermögen stark eingeschränkt ist und kaum eine Informationsaufnahme oder Informationsverarbeitung möglich ist (Rötting 1999).

Anhand globaler Blickbewegungsmaße können Lese- und Wahrnehmungsprozesse verglichen werden. Daher bietet sich die Aufzeichnung und Analyse von Blickbewegungen insbesondere im Bereich von Textaufgaben an. Globale Maße ermöglichen Aussagen über die Wahrnehmung und die Verarbeitung der Aufgabe als Ganzes und ermöglichen den Vergleich der Maße über verschiedene Aufgaben hinweg. So stellt die Dauer des Blickpfades die Bearbeitungszeit der Teilnehmenden dar und kann als Hinweis auf zugrundeliegende kognitive Verarbeitungsprozesse angesehen werden (Joos et al. 2003; Rötting 2001). Die Länge der Blicksprünge wird als Sakkadenamplitude bezeichnet. Eine kleinere Sakkadenamplitude kann als Indikator für ein detailliertes und genaueres Lesen angesehen werden (Holmqvist et al. 2011). Die Fixationsdauer, als Zeitspanne zwischen zwei aufeinanderfolgenden Sakkaden, ist abhängig von der Aufgaben- beziehungsweise Textschwierigkeit. Auf Grundlage der *eye-mind* Hypothese wird sie auch als Maß für die kognitive Verarbeitungstiefe interpretiert (Holmqvist et al. 2011; Radach et al. 2012). So konnte Deubel (2008) zeigen, dass Fixationen ab einer Länge von 250 ms einen tiefergehenden und analytischeren kognitiven Prozess widerspiegeln. Schindler und Lilienthal (2019) konnten in ihrer Studie erste Indizien feststellen, die darauf hinweisen, dass insbesondere in emotionalen Erregungszuständen, wie Stress oder Panik, Fixationsdauern Veränderungen aufzeigen. Die Abhängigkeit der Fixationsdauer von der kognitiven Verarbeitung sowie Belastung lassen begründete Rückschlüsse auf die meist unbewussten und hochautomatischen Informationsverarbeitungsprozesse zu. Eye Tracking-Daten sind vielseitig, reich an Informationen und müssen hinsichtlich des betrachteten Gegenstandes interpretiert werden (Holmqvist et al. 2011; Triesch et al. 2003). Die Interpretationsprozesse wahrgenommener Aufgabenschwierigkeiten können mithilfe von Verbalprotokollen abgebildet werden (vgl. Philipp 2018). Verbalprotokolle stellen audiographierte und transkribierte Daten dar, die Äußerungen von Probanden erfassen, um Denkvorgänge auf der Suche nach Zusammenhängen und Erklärungen reflektieren zu können (Heine 2005). Durch die Verbalisierung von Gedanken können Einblicke in die Verarbeitungs- sowie Interpretationsprozesse von Informationen gewährt werden (Heine 2014). Verbalprotokolle in Zusammenhang mit Eye Tracking-Daten ermöglichen Aussagen darüber, ob Aufgabenmerkmale nicht nur wahrgenommen, sondern auch als relevant für die Aufgabenschwierigkeit von der diagnostizierenden Person interpretiert werden.

3 Forschungsfragen und Hypothesen

Aufgrund der dargestellten Forschungslage ist das Hauptziel der vorliegenden experimentellen Eye Tracking-Studie die Generierung empirischer Evidenz über den Einfluss von Stress auf Prozesse beim diagnostischen Urteilen, bestehend aus Wahrnehmungs- sowie kognitiven Interpretationsprozessen.

Die vorliegende Studie geht insbesondere der Frage nach, ob sich Unterschiede in den Wahrnehmungsprozessen von Aufgaben als Ganzes, gemessen anhand von globalen Blickbewegungen, zwischen gestressten Personen (Experimentalgruppe) und nicht gestressten Personen (Kontrollgruppe) zeigen werden. Auf Grundlage einer physiologisch begrenzten Arbeitsgedächtniskapazität sowie der zusätz-

lich gebundenen kognitiven Kapazität durch Stresserleben werden Unterschiede in den ausgewählten globalen Blickbewegungsmaßen zwischen den Gruppen erwartet (Hypothese 1). Es wird präziser eine kürzere Blickpfaddauer sowie eine kürzere Fixationsdauer für die Experimentalgruppe angenommen. Auf Grundlage der *eye-mind* Hypothese können die Blickpfaddauer sowie die Fixationsdauer als Maße für die kognitive Verarbeitungstiefe angesehen werden. Darüber hinaus wird von einer geringeren Anzahl von Fixationen ab einer Länge von 250 ms, die einen tiefergehenden und analytischeren Prozess widerspiegeln können (Deubel 2008), in der Experimentalgruppe ausgegangen. Eine größere Sakkadenamplitude als Hinweis auf ein Überfliegen des Textes sowie ein ungenaueres Lesen des Textes wird in der Experimentalgruppe erwartet (Radach et al. 2012).

Die vorliegende Studie geht darüber hinaus der Frage des Einflusses von Stress auf Interpretationsprozesse wahrgenommener Aufgabenschwierigkeiten nach, die sich anhand von Verbalprotokollen aufzeigen lassen. Es wird angenommen, dass sich die Effekte von Stress anhand einer eingeschränkten Interpretation der schwierigkeitsgenerierenden Aufgabenmerkmale in der Experimentalgruppe aufzeigen lassen. Die Erwartung ist, dass die Experimentalgruppe pro Aufgabe weniger relevante schwierigkeitsgenerierende Aufgabenmerkmale nennen wird (Hypothese 2a) sowie insgesamt im Mittel weniger Aufgabenmerkmale pro Aufgabe (Hypothese 2b) als relevant interpretieren wird.

4 Methode

4.1 Stichprobe

Die Basis für die Untersuchung bildete eine Stichprobe von $N=64$ Lehramtsstudierenden mit dem Fach Mathematik der Pädagogischen Hochschule Heidelberg (Alter: $M=22,72$, $SD=4,16$, 73 % Frauen). Die Teilnehmenden wurden randomisiert einer Experimentalgruppe (Stressbedingung) oder einer Kontrollgruppe zugeordnet. Bei der Experimentalgruppe ($n=33$) wurde im Gegensatz zu der Kontrollgruppe zusätzlich vor der Diagnoseaufgabe mithilfe des SECPT eine physiologische Stressreaktion erzeugt.

4.2 Studiendesign und Erhebungsinstrumente

Jeder Teilnehmende wurde in einem Einzelsetting in einem ruhigen Raum der Universität getestet. Die 45-minütige Untersuchung erfolgte zu einem mit den Teilnehmenden vereinbarten Einzeltermin zwischen 14 und 17 Uhr, da der Cortisolspiegel der Teilnehmenden in dieser Zeit am Nachmittag als stabil erachtet werden kann (vgl. Het et al. 2005).

4.2.1 Erfassung soziodemographischer Daten anhand eines Fragebogens

Mit Hilfe eines Fragebogens wurden zunächst soziodemographische Angaben der Teilnehmenden erfasst. Daneben wurden vorausgegangene körperliche Aktivitäten

der Teilnehmenden am Tag der Untersuchung erhoben, um mögliche Einflussfaktoren auf den Cortisolspiegel zu kontrollieren.

4.2.2 Erfassung des objektiven Stresszustandes durch die Messung von Cortisolwerten

Die Experimentalgruppe wurde vor der Diagnoseaufgabe mithilfe des SECPT künstlich unter Stress gesetzt. Hierzu legte die Versuchsperson ihre Hand für wenige Minuten in ein Wasserbad, welches in der Experimentalbedingung eine Temperatur von 0–4 °C aufwies (Schwabe et al. 2008; Velasco et al. 1997). In der Kontrollbedingung betrug die Wassertemperatur 35–37 °C (Schwabe et al. 2008). Die Teilnehmenden wurden angewiesen, ihre Hand so lange wie möglich im Wasserbad zu halten. Um Schäden am Gewebe der Hand zu verhindern, wurde der Test spätestens nach drei Minuten von der Versuchsleitung beendet (vgl. Schwabe et al. 2008), sofern die Versuchsperson ihre Hand nicht schon vorher aus dem Wasserbad gezogen hatte. Die sozialevauiativen Komponenten der Stressinduktion sind auf die Beobachtung durch eine weibliche Versuchsleiterin sowie eine Videoaufnahme während der Stressinduktion zurückzuführen.

Um den objektiven Stresszustand der Teilnehmenden während des Experiments zu erheben, gaben die Teilnehmenden vor, während und nach der Diagnoseaufgabe insgesamt vier Speichelproben ab (zum Zeitpunkt 0 als Ausgangswert, 5 min nach der Stressinduktion (SI), 15 min und 25 min nach der SI) (vgl. Schwabe et al. 2008). Zwischen 15 und 25 min nach der Stressinduktion erreicht der Cortisolwert in der Regel seinen Maximalwert (vgl. Schwabe et al. 2008). Die Analyse des Speichels im Hinblick auf Cortisol ließ eine Dokumentation des objektiven Stressniveaus der Teilnehmenden zu. Die Objektivität war durch die standardisierte Durchführung sowie durch die Auswertung der Proben hinsichtlich des Cortisolwertes gegeben. Die Validität der Stressinduktion konnte bei der Entwicklung des SECPT durch Schwabe und Kollegen gezeigt werden (vgl. Schwabe et al. 2008).

4.2.3 Diagnoseaufgabe am Eye Tracker

Die Diagnoseaufgabe der Teilnehmenden bestand in der Einschätzung von acht mathematischen Textaufgaben: Diese sollten hinsichtlich der Schwierigkeit für Real-schülerinnen und -schüler am Ende der Klassenstufe 6 beziehungsweise am Anfang der Klassenstufe 7 auf einer zehnstufigen Skala eingeschätzt werden. Es handelte sich um Textaufgaben der Mathematik aus dem Bereich der Addition und der Subtraktion von Brüchen. Die Aufgaben variierten in ihrer Schwierigkeit in Anlehnung an das Klassifikationsschema von Jordan et al. (2006) sowie weiterer Befunde zu schwierigkeitsgenerierenden Aufgabenmerkmalen aus dem Bereich der Textaufgaben (Gürsoy et al. 2013; Maier und Schweiger 1999) sowie der Bruchrechnung (Padberg und Wartha 2017). Die Schwierigkeit variierte insbesondere in den mathematischen Bereichen *Bruchschwierigkeit* und *Anzahl der Rechenschritte* sowie in der sprachlogischen Komplexität hinsichtlich der Bereiche *Lexikologie* und *Syntax* (siehe Beispielitems Abb. 2 und 3).

Aufgabe 1:

Tina geht einkaufen und trägt $\frac{1}{5}$ kg Butter und $\frac{2}{5}$ kg Mehl nach Hause. Wie schwer ist Tinas Einkauf?

Abb. 2 Leichtes Beispielitem: Die Bruchschwierigkeit ist niedrig, da echte, gleichnamige Brüche mit kleinem Zähler und Nenner zu verarbeiten sind. Die Lösung der Aufgabe besteht aus einem Rechenschritt, der eine Addition des ersten und des zweiten Bruchs umfasst. Die sprachlogische Komplexität des Beispielitems ist im Bereich der Lexikologie sowie im Bereich der Syntax niedrig

Aufgabe 8:

Im Chemieunterricht werden $\frac{2}{5}$ Liter Wasserstoff mit $\frac{2}{6}$ Liter Sauerstoff von der Lehrperson zusammengemischt, sodass Wasserstoffoxide entstehen. Welche Menge erhält die Lehrkraft, wenn bereits $\frac{3}{7}$ Liter Wasserstoffoxide im Reagenzglas enthalten waren?

Abb. 3 Mittelschweres Beispielitem: Die Bruchschwierigkeit ist im mittleren Bereich, da echte, jedoch ungleichnamige Brüche mit kleinem Zähler und Nenner zu verarbeiten sind. Zur Lösung der Aufgabe müssen zwei Rechenschritte, zwei hintereinander ausgeführte Additionen, berechnet werden. Die sprachlogische Komplexität ist sowohl im Bereich der Lexikologie aufgrund der chemischen Fachbegriffe als auch im Bereich der Syntax aufgrund der Passivkonstruktion hoch

Die Diagnoseaufgabe wurde für die Nutzung am Eye Tracker konzipiert und programmiert. Es wurde mit dem EyeLink 1000 Plus gearbeitet. Der Bildschirm wurde 60 cm vom Auge des Teilnehmenden entfernt aufgebaut. Die Teilnehmenden beider Gruppen betrachteten den Bildschirm beidäugig, aber nur die Augenbewegungen des rechten Auges wurden mit einer Messrate von 500 Hertz aufgezeichnet. Eine Kopfstütze wurde eingesetzt, um die Kopfbewegungen zu minimieren und dadurch die Genauigkeit der Aufzeichnung der Blickbewegungen zu erhöhen. Zu Beginn der Diagnoseaufgabe am Eye Tracker wurde bei jeder Versuchsperson eine Kalibrierung mit einem 9-Punkte Bildschirm durchgeführt, um optimale Eye Tracking-Ergebnisse zu erzielen (Jepma und Nieuwenhuis 2011; Merkley und Ansari 2010). Für eine optimale Durchführungsobjektivität, wurde den Teilnehmenden die Vorgehensweise der Studie sowie die Aufgabenstellung durch eine Einführung auf dem Bildschirm erklärt.

Die Teilnehmenden wurden gebeten, acht Textaufgaben hinsichtlich der Schwierigkeit auf einer zehnstufigen Skala einzuschätzen sowie ihre Einschätzung nach jeder Aufgabe in einem Verbalprotokoll zu begründen. Neben der Adressatengruppe, Realschülerinnen und -schüler am Ende der Klassenstufe 6 beziehungsweise am Anfang der Klassenstufe 7, wurden den Teilnehmenden die Aufgabenstellung sowie die schwierigkeitsgenerierenden Merkmale der Textaufgaben *Bruchschwierigkeit*, *An-*

zahl der Rechenschritte, *Lexikologie* und *Syntax* als Orientierungsrahmen genannt. Jede Aufgabe wurde einzeln auf dem Bildschirm präsentiert. Durch Drücken der Enter-Taste gelangten die Teilnehmenden selbstständig zur jeweils nächsten Bildschirmansicht. Für die Teilnehmenden bestand bei der Bearbeitung der Aufgaben kein Zeitlimit. Sie wurden gebeten, die Aufgaben so schnell und so akkurat wie möglich einzuschätzen. Nachdem die Teilnehmenden eine Entscheidung getroffen hatten, betätigten sie die Enter-Taste und interpretierten die wahrgenommene Aufgabe hinsichtlich ihrer Schwierigkeit. Die Interpretation der Teilnehmenden wurde mithilfe eines externen Diktiergerätes durch Verbalprotokolle erfasst. Die jeweilige Aufgabe wurde den Teilnehmenden auch während der Aufzeichnung der Verbalprotokolle präsentiert. Der Eye Tracker wurde vor jedem Verbalprotokoll, durch das Drücken der Enter-Taste, angehalten. Die Blickbewegungen der Teilnehmenden wurden ausschließlich während den Wahrnehmungsprozessen aufgezeichnet, um Überlagerungen der Blickbewegungen während den Interpretationsprozessen zu vermeiden. Die Aufnahme der Verbalprotokolle erfolgte nach jeder Aufgabe. Die Teilnehmenden konnten während der Dauer des Verbalprotokolls ihr Urteil korrigieren, sie erhielten kein Feedback bezüglich der Genauigkeit ihres Urteils. Die Verbalprotokolle der Teilnehmenden wurden anschließend hinsichtlich der Aufzählung der vier relevanten schwierigkeitsgenerierenden Aufgabenmerkmale (*Bruchschwierigkeit*, *Anzahl der Rechenschritte*, *Lexikologie* und *Syntax*) ausgewertet und binär kodiert. Aufgrund der zeitlichen Anlage der Studie bearbeiteten die Teilnehmenden die Diagnoseaufgabe im Zeitfenster des maximalen Cortisolwertes, 15 bis 25 min nach der Stressinduktion. Es wurde daher erwartet, dass sich die Stressinduktion der Experimentalgruppe sowohl in den Wahrnehmungsprozessen anhand globaler Eye Tracking-Maße als auch in den Interpretationsprozessen anhand der Qualität der Verbalprotokolle zeigen wird.

5 Ergebnisse

Die Unterschiede der Blickbewegungsdaten zwischen der Experimental- und der Kontrollgruppe wurden durch eine multivariate Varianzanalyse (MANOVA) berechnet. Die globalen Maße der Blickbewegungen sowie die Aufzählung der schwierigkeitsgenerierenden Aufgabenmerkmale *Bruchschwierigkeit*, *Anzahl der Rechenschritte*, *Lexikologie* und *Syntax* wurden als abhängige Variablen operationalisiert. Die Stressinduktion stellte die unabhängige Variable dar. Das Signifikanzniveau betrug fünf Prozent, die gerichteten Hypothesen wurden einseitig getestet.

5.1 Auswertung der Cortisolreaktionen

Um die Entwicklung der Cortisolwerte zu bestimmen, wurde der sogenannte *Slope*, welcher den Anstieg der Regressionsgeraden durch die verschiedenen Messzeitpunkte des Cortisols während des Experiments darstellt, berechnet. Hierbei wurde die Differenz der mittleren Cortisolkonzentration der Baseline und der mittleren Cortisolkonzentration des Messzeitpunktes 4 berechnet. Die Baseline bildete den individuellen Ausgangswert der Cortisolwerte der Teilnehmenden zum Zeitpunkt 0

Tab. 1 Übersicht der globalen Eye Tracking-Maße

	Experimentalgruppe <i>n</i> = 33 <i>M</i> (<i>SD</i>)	Kontrollgruppe <i>n</i> = 31 <i>M</i> (<i>SD</i>)
Blickpfaddauer (<i>ms</i>)	23.499,75 (7526,73)	29.006,05 (9408,98)
Fixationsdauer (<i>ms</i>)	20.106,29 (6927,77)	23.245,87 (6980,65)
Fixationen ab 250ms	27,23 (10,88)	32,39 (11,21)
Sakkadenamplitude (<i>Grad</i>)	2,17 (0,37)	2,23 (0,53)

des Experiments. Es konnte eine signifikant gesteigerte Cortisolreaktion bei der Experimentalgruppe ($M = 1,20$, $SD = 2,29$) im Vergleich zur Kontrollgruppe ($M = -0,20$, $SD = 1,78$) gezeigt werden ($t(62) = -2,73$, $p = 0,004$).

5.2 Auswertung der Eye Tracking-Daten hinsichtlich globaler Maße

Die Eye Tracking-Daten wurden mit der Software E-Prime ausgewertet. Die Analyse der Eye Tracking-Daten wurde auf ausgewählte globale Maße begrenzt: die *Blickpfaddauer* (Angaben in ms) als Bearbeitungszeit der Aufgabe sowie als Hinweis für die zugrundeliegenden kognitiven Verarbeitungsprozesse, die *Fixationsdauer* (Angaben in ms) als Indikator für die kognitive Verarbeitungstiefe, die *Anzahl der Fixationen ab einer Dauer von 250ms*, ab der eine metakognitive Verarbeitung möglich ist sowie die *Sakkadenamplitude* (Angaben in Grad) als Indikator für die Genauigkeit beim Lesen (Tab. 1).

Die Berechnung der MANOVA zeigte erwartungskonforme (vgl. Hypothese 1), signifikante Haupteffekte von Stresserleben auf die *Blickpfaddauer* (ms) der Experimentalgruppe im Vergleich zur Kontrollgruppe ($F(1,62) = 5,50$, $p = 0,006$). Die Kontrollgruppe bearbeitete jede Aufgabe im Durchschnitt 5,5s länger als die Experimentalgruppe. Es konnte ein signifikanter Haupteffekt von Stresserleben auf die *Fixationsdauer* pro Aufgabe ($F(1,62) = 3,97$, $p = 0,038$) sowie ein signifikanter Haupteffekt der Anzahl an *Fixationen ab einer Dauer von 250ms* im Vergleich zur Kontrollgruppe errechnet werden ($F(1,62) = 4,70$, $p = 0,033$). Hinsichtlich der *Sakkadenamplitude* konnte kein signifikanter Haupteffekt gezeigt werden ($F(1,62) = 0,16$, $p = 0,295$).

5.3 Auswertung der Verbalprotokolle

Die Verbalprotokolle dienten in der vorliegenden Studie der stringenteren Auswertung der Eye Tracking-Daten, indem sie die Interpretationsprozesse der wahrgenommenen Aufgabemerkmale der Teilnehmenden abbildeten. Die Verbalprotokolle der Teilnehmenden wurden hinsichtlich der Aufzählung der relevanten schwierigkeitsgenerierenden Aufgabenmerkmale ausgewertet und binär kodiert: „(richtig) genannt“ oder „nicht genannt“. Es kam in der vorliegenden Stichprobe nicht vor, dass ein Merkmal genannt, aber falsch interpretiert wurde. Um die Werte über die Aufgaben hinweg zwischen den Gruppen vergleichen zu können, wurden Prozentwerte gebildet, indem die Summe der binär kodierten Werte durch die Gesamtzahl der Aufgaben

Tab. 2 Übersicht über die mittleren Nennungshäufigkeiten der variierten Aufgabenmerkmale in den Verbalprotokollen

	Experimentalgruppe <i>n</i> = 33	Kontrollgruppe <i>n</i> = 31
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Bruchschwierigkeit	63,58 (19,28)	81,05 (17,03)
Anzahl Rechenschritte	46,92 (27,87)	76,21 (31,51)
Lexikologie	16,16 (22,23)	34,41 (30,40)
Syntax	8,08 (16,73)	21,51 (22,02)

($N=8$) geteilt und dann mit 100 multipliziert wurde. Die Verbalprotokolle wurden vom Erstautor und einer studentischen Mitarbeiterin mit hoher Interrater-Reliabilität doppelt kodiert. Der Cohens-Kappa-Koeffizient (vgl. Cohen 1988) lag bei den relevanten Aufgabenmerkmalen zwischen 0,86 und 0,95. Tab. 2 zeigt eine Übersicht über die Häufigkeiten der Aufzählungen der vier variierten Aufgabenmerkmale.

Bei allen Aufgaben wurden die variierten mathematischen Aufgabenmerkmale *Bruchschwierigkeit* ($t(62)=3,82$, $p<0,001$) und *Anzahl der Rechenschritte* ($t(62)=3,94$, $p<0,001$) sowie die sprachlichen Aufgabenmerkmale im Bereich der *Lexikologie* ($t(62)=2,75$, $p=0,004$) und im Bereich der *Syntax* ($t(61)=2,75$, $p=0,004$) erwartungskonform signifikant häufiger von der Kontrollgruppe als von der Experimentalgruppe genannt (vgl. Hypothese 2a). Die Kontrollgruppe ($M=53,29$, $SD=12,40$) berücksichtigte darüber hinaus im Mittel mehr relevante schwierigkeitsgenerierende Aufgabenmerkmale pro Aufgabe im Vergleich zur Experimentalgruppe ($M=33,69$, $SD=09,18$) ($t(62)=7,21$, $p<0,001$) (vgl. Hypothese 2b).

6 Diskussion

Die Befunde der vorliegenden Studie weisen zusammenfassend darauf hin, dass sowohl Wahrnehmungs- als auch Interpretationsprozesse beim diagnostischen Urteilen durch Stresserleben negativ beeinflusst werden. Die Annahme, dass durch die Stressinduktion kognitive Kapazitäten gebunden werden, die nicht mehr für die kognitiven Prozesse beim diagnostischen Urteilen zur Verfügung stehen, kann durch die dargelegten Ergebnisse bekräftigt werden.

Anhand der Analyse der Cortisolwerte kann davon ausgegangen werden, dass die Versuchspersonen der Experimentalgruppe in der vorliegenden Studie durch den SECPT hinreichend gestresst waren. Die Wahrnehmungsprozesse, die in der vorliegenden Studie anhand globaler Blickbewegungsmaße analysiert wurden, zeigten in drei der vier untersuchten Maße erwartungskonforme signifikante Unterschiede zwischen der Kontroll- und der Experimentalgruppe (vgl. Hypothese 1). Der geringe Unterschied der Sakkadenamplitude zwischen der Experimental- und der Kontrollgruppe lässt sich auf die unterschiedliche Sichtweise auf die Bedeutung der Sakkadenamplitude im Wahrnehmungsprozess zurückführen. Die Größe der Sakkadenamplitude der Kontrollgruppe ohne Stressinduktion lässt sich durch ein

detailliertes Lesen der Aufgaben erklären (Holmqvist et al. 2011). Diese Annahme wird durch die höhere Blickpfaddauer der Kontrollgruppe zusätzlich bekräftigt (Joos et al. 2003; Rötting 2001). Die Größe der Sakkadenamplitude der Experimentalgruppe lässt sich durch einen höheren kognitiven Aufwand beziehungsweise durch weniger zur Verfügung stehende kognitive Ressourcen begründen, die sich an anderen Blickbewegungsmaßen signifikant nachweisen ließen.

Die Annahme der geringer zur Verfügung stehenden kognitiven Ressourcen der Experimentalgruppe mit Stressinduktion wird durch die signifikant kürzere Fixationsdauer, die auf Grundlage der *eye-mind* Hypothese als Maß für die kognitive Verarbeitungstiefe interpretiert werden kann (Holmqvist et al. 2011; Radach et al. 2012), sowie die geringere Anzahl an Fixationen ab einer Dauer von 250 ms, die einen tiefergehenden und analytischeren kognitiven Prozess widerspiegeln (Deubel 2008), unterstützt.

Die Interpretation der wahrgenommenen Aufgabenmerkmale wurde in der vorliegenden Studie als Aufzählung der schwierigkeitsgenerierenden Aufgabenmerkmale operationalisiert. Entsprechend unserer Hypothese 2a nannte die Kontrollgruppe ohne Stressinduktion (zum Teil hoch signifikant), in allen vier untersuchten Merkmalsbereichen, mehr schwierigkeitsgenerierende Merkmale pro Aufgabe im Vergleich zur Experimentalgruppe mit Stressinduktion. Dies weist daraufhin, dass die Aufgabenmerkmale von den Teilnehmenden der Kontrollgruppe signifikant häufiger für das diagnostische Urteil als relevant eingestuft und berücksichtigt wurden. Darüber hinaus nannte die Kontrollgruppe im Mittel insgesamt mehr schwierigkeitsgenerierende Aufgabenmerkmale pro Aufgabe als die Experimentalgruppe (Hypothese 2b).

Die vorliegende Studie konnte zeigen, dass sowohl Wahrnehmungsprozesse von Aufgaben als Ganzes als auch Interpretationsprozesse beim diagnostischen Urteilen aufgrund der Bindung kognitiver Kapazitäten durch Stresserleben negativ beeinflusst werden. Wahrnehmungen sowie deren Interpretationsprozesse bilden die Grundlage für adäquate diagnostische Entscheidungen während des Unterrichtsgeschehens. Sie sind ausschlaggebende Voraussetzung für adaptiven Unterricht und erfolgreiches Lernen (Praetorius und Südkamp 2017).

6.1 Limitationen der vorliegenden Studie

Einschränkend bezüglich der Methode des Eye Trackings sei anzumerken, dass das periphere ($>5^\circ$) sowie parafoveale ($2\text{--}5^\circ$) Sehen die Wahrnehmung von Buchstaben und Wörtern außerhalb des scharfen Sehbereichs ermöglicht. Die Wahrnehmungsspanne beim Lesen beträgt drei bis vier Buchstaben entgegen der Leserichtung sowie 14 bis 15 Buchstaben in Leserichtung. Diese Bereiche liegen außerhalb des Fixationspunktes und können mithilfe des Eye Trackings nicht erfasst werden (Rayner 1998). Um die Wahrnehmung sowie die Verarbeitung dieser Bereiche durch die Teilnehmenden dennoch zu erheben, wurde in der vorliegenden Studie eine Methodenkombination von Eye Tracking Daten und Verbalprotokollen durchgeführt. Verbalisierungen in Form von Verbalprotokollen bieten die Möglichkeit, wahrgenommene Merkmale des peripheren und des parafovealen Sehens zu erfassen. Die Methodenkombination von Eye Tracking Daten und Verbalprotokollen in der vorliegenden Studie ermöglicht darüber hinaus eine Erhöhung der Validität der In-

terpretation der Eye Tracking-Daten hinsichtlich der zugrundeliegenden kognitiven Prozesse.

Die mathematischen Fähigkeiten sowie die Lesefähigkeiten der Teilnehmenden wurden in der vorliegenden Studie nicht berücksichtigt. Es ist anzunehmen, dass beide Fähigkeiten einen Einfluss auf die Blickbewegungen sowie die Einschätzung der Aufgabenschwierigkeiten ausüben könnten. Die Berücksichtigung dieser Facetten in zukünftigen Studien könnte helfen, ausgleichende Parameter in Bezug auf die Stressinduktion zu identifizieren.

Die Validität der Operationalisierung von Stress in der vorliegenden Studie ist diskutabel. Bei der vorliegenden Studie handelt es sich um eine der ersten Studien, die den Einfluss von Stress im anwendungsbezogenen Schulkontext untersucht haben. Es liegen keine vergleichbaren Cortisolwerte von Lehrkräften während des Unterrichtsgeschehens vor. Aufgrund des methodisch anspruchsvollen Designs mit einem stationären Eye Tracker sowie des Zeitfensters (14–17 Uhr) in dem der Cortisolspiegel als stabil bewertet werden kann, war die Durchführung der Studie im Unterrichtsgeschehen nicht möglich. Auch wenn die Art der Stressinduktion nicht mit den Stressoren des Unterrichtsgeschehens vergleichbar ist, sind die physiologischen Prozesse, die durch die Stressinduktion im Labor-Setting angeregt werden, die gleichen Prozesse wie im Unterrichtsgeschehen.

6.2 Implikationen für Praxis sowie Forschung und Schlussfolgerung

Mit der Frage nach dem Einfluss von situativ aktivierten Personencharakteristika auf Prozesse beim Diagnostizieren ergeben sich Implikationen für die Praxis des Unterrichtsgeschehens und für die Forschung.

Für die Praxis bedeuten die vorliegenden Ergebnisse, dass Wahrnehmungs- und Interpretationsprozesse beim Diagnostizieren durch Stresserleben beeinträchtigt werden können. Insbesondere, weil Anspannungen und Stress im Unterrichtsgeschehen aufgrund der Interaktionsdichte nicht gänzlich vermieden werden können (Rothland und Terhart 2007; Scheuch und Knothe 1997) sollte die Integration der gewonnenen Erkenntnisse der vorliegenden Studie in die Qualifikation von Lehramtsstudierenden als auch von Lehrpersonen durch entsprechende Aus- und Weiterbildungsangebote erfolgen. Neben einer Sensibilisierung von Lehramtsstudierenden sowie Lehrkräften bezüglich des Einflusses von Stress auf Wahrnehmungs- sowie Interpretationsprozesse, können persönliche Ressourcen in Bezug auf Stress in Aus- und Weiterbildungsangeboten aufgebaut werden (von der Embse et al. 2019). Darüber hinaus wären Anschlussstudien wünschenswert, die fachwissenschaftliche und fachdidaktische Facetten integrieren, um ausgleichende Parameter im Hinblick auf Stresserleben von Lehrkräften untersuchen und identifizieren zu können. Insbesondere die Wissensfacette PCK stellt aufgrund bisheriger Forschungsergebnisse im Bereich der Mathematikdidaktik einen interessanten Ansatzpunkt dar. Die bereits zitierte Studie von Ostermann et al. (2017) konnte zeigen, dass durch die Vermittlung von spezifischem PCK die Einschätzung von Aufgabenschwierigkeiten zunimmt. Erstrebenswert wäre eine Anschlussstudie, die mit der Methode Eye Tracking den Einfluss der Vermittlung von PCK auf Wahrnehmungs- und Interpretationsprozesse untersucht. Die identifizierten Wissensfacetten in der Aus-

bildung von Lehramtsstudierenden stärker zu fokussieren und zu vermitteln, würde die Qualität von diagnostischen Prozessen sowie Entscheidungen während des Unterrichtsgeschehens weiter voranbringen.

Vor dem Hintergrund der vorliegenden Studie scheint es wichtig, dass sich Forschung im Bereich diagnostischer Prozesse von Lehrkräften stärker mit der Thematik Stress als Einflussfaktor auseinandersetzen. Es wäre erstrebenswert, dass zukünftige Studien im Bereich der diagnostischen Prozesse von Lehrkräften situativ aktivierte Personencharakteristika berücksichtigen.

Danksagung Die vorliegende Studie wurde im Rahmen des Forschungs- und Nachwuchskollegs „Diagnostische Kompetenzen von Lehrkräften: Einflüsse, Struktur und Förderung (DiaKom)“ vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert. Für die Umsetzung des aufwendigen Studiendesigns danken wir weiterhin der Pädagogischen Hochschule Heidelberg für die finanzielle Unterstützung.

Funding Open Access funding provided by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- al' Absi, M., Petersen, K. L., & Wittmers, L. E. (2002). Adrenocortical and hemodynamic predictors of pain perception in men and women. *Pain*, 96, 197–204.
- Birbaumer, N., & Schmidt, R. F. (2010). *Biologische Psychologie*. Berlin: Springer.
- Bromme, R., & Haag, L. (2008). Forschung zur Lehrerpersönlichkeit. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 803–819). Wiesbaden: VS.
- Brühwiler, C. (2014). *Adaptive Lehrkompetenz und schulisches Lernen: Effekte handlungssteuernder Kognitionen von Lehrpersonen auf Unterrichtsprozesse und Lernergebnisse der Schülerinnen und Schüler*. Zugl.: Koblenz-Landau, Univ., Diss. Münster: Waxmann.
- Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 215–234). Münster: Waxmann.
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Hrsg.), *Dual-process theories in social psychology* (S. 73–96). New York: Guilford.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N. J.: Erlbaum.
- Deubel, H. (2008). The time course of presaccadic attention shifts. *Psychological Research*, 72, 630–640.
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130, 355–391.
- Engbert, R., & Kliegl, R. (2001). Mathematical models of eye movements in reading: a possible role for autonomous saccades. *Biological Cybernetics*, 85, 77–87.
- Gürsoy, E., Benholz, C., Renk, N., Prediger, S., & Büchter, A. (2013). Erlös = Erlösung? – Sprachliche und konzeptuelle Hürden in Prüfungsaufgaben zur Mathematik. *Deutsch als Zweitsprache*, 1, 14–24.

- Heine, L. (2005). Lautes Denken als Forschungsinstrument in der Fremdsprachenforschung. *Zeitschrift für Fremdsprachenforschung*, 16, 163–185.
- Heine, L. (2014). Introspektion. In J. Settinieri, S. Demirkaya, A. Feldmeier, N. Gültekin-Karakoç, S. Im-mich & C. Riemer (Hrsg.), *Einführung in empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache* (S. 123–135). Stuttgart: UTB.
- Helmke, A., Hosenfeld, I., & Schrader, F.-W. (2004). Vergleichsarbeiten als Werkzeug für die Verbesserung der diagnostischen Kompetenz von Lehrkräften. In R. Arnold & C. Griesse (Hrsg.), *Schulleitung und Schulentwicklung. Voraussetzungen, Bedingungen und Erfahrungen* (S. 119–144). Baltmannsweiler: Schneider Hohengehren.
- Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., et al. (2018). Teachers' assessment competence: integrating knowledge-, process- and product-oriented approaches into a competence-oriented conceptual model. *Journal of Teaching and Teacher Education*, 76, 181–193.
- Het, S., Ramlow, G., & Wolf, O. T. (2005). A meta-analytic review of the effects of acute cortisol administration on human memory. *Psychoneuroendocrinology*, 30, 771–784.
- Hiebert, J., Gallimore, R. G., Garnier, H., Bogard, G. K., Hollingsworth, H., Jacobs, J. K., et al. (2003). *Teaching mathematics in seven countries*. Results from the TIMSS 1999 video study (NCES 2003 013). Washington, DC: U.S. Gov. Print. Off. <http://purl.access.gpo.gov/GPO/LPS31989>.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research*, 59, 297–313.
- 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: Oxford University Press.
- Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education*, 43, 358–390.
- Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, 23, 1587–1596.
- Joos, M., Rötting, M., & Velichkovsky, B. M. (2003). Die Bewegungen des menschlichen Auges: Fakten, Methoden, innovative Anwendungen. In G. Rickheit, T. Herrmann & W. Deutsch (Hrsg.), *Psycholinguistik/Psycholinguistics. Ein internationales Handbuch/ an international handbook* (S. 142–168). Berlin: De Gruyter.
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., & Kunter, M. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*. Berlin: Max-Planck-Inst. für Bildungsforschung.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, 87, 329–354.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23, 197–209.
- Kirschbaum, C., & Hellhammer, D. (Hrsg.). (1999). *Psychoendokrinologie und Psychoimmunologie*. Göttingen: Hogrefe.
- De Kloet, E. R., Joëls, M., & Holsboer, F. (2005). Stress and the brain: from adaptation to disease. *Nature Reviews Neuroscience*, 6, 463–475.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- 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* (S. 3–31). Berlin: Springer.
- 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.
- Maier, H., & Schweiger, F. (1999). *Mathematik und Sprache: Zum Verstehen und Verwenden von Fachsprache im Mathematikunterricht*. Wien: öbv & hpt.
- Merkley, R., & Ansari, D. (2010). Using eye tracking to study numerical cognition: the case of the ratio effect. *Experimental Brain Research*, 206, 455–460.
- 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, 579–605.
- Padberg, F., & Wartha, S. (2017). *Didaktik der Bruchrechnung*. Berlin: Springer.

- Philipp, K. (2018). Diagnostic competences of mathematics teachers with a view to processes and knowledge resources. In T. Leuders, K. Philipp & J. Leuders (Hrsg.), *Diagnostic competence of mathematics teachers* (S. 109–127). Cham: Springer.
- Praetorius, A.-K., & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (S. 13–18). Münster: Waxmann.
- Radach, R., Günther, T., & Huestegge, L. (2012). Blickbewegungen beim Lesen, Leseentwicklung und Legasthenie. *Lernen und Lernstörungen*, 1, 185–204.
- Rausch, T., Matthäi, J., & Artelt, C. (2015). Mit Wissen zu akkuraten Urteilen? Zum Zusammenhang von Wissensgrundlagen und Urteilsgüte im Bereich des Textverstehens. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47, 147–158.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *The psychology of reading*. New York: Psychology Press.
- Reichle, E.D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *Behavioral and Brain Sciences*, 26, 445–476.
- Rothland, M., & Terhart, E. (2007). Beruf: Lehrer-Arbeitsplatz: Schule. Charakteristika der Arbeitstätigkeit und Bedingungen der Berufssituation. In M. Rothland (Hrsg.), *Belastung und Beanspruchung im Lehrerberuf. Modelle, Befunde, Interventionen* (S. 11–31). Wiesbaden: VS.
- Rötting, M. (1999). Typen und Parameter von Augenbewegungen. In M. Rötting & K. Seifert (Hrsg.), *Blickbewegungen in der Mensch-Maschine-Systemtechnik*. ZMMS Spektrum, (Bd. 8, S. 1–18). Sinzheim: Pro-Universität.
- Rötting, M. (2001). *Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen*. Schriftenreihe Rationalisierung und Humanisierung, Bd. 34. Aachen: Shaker. Zugl.: Aachen, Techn. Hochsch., Diss.
- Scheuch, K., & Knothe, M. (1997). Psychophysische Beanspruchung von Lehrern in der Unterrichtstätigkeit. In S. Buchen, U. Carle, P. Döbrich, H.-D. Hoyer & H.-G. Schönwälder (Hrsg.), *Jahrbuch für Lehrerforschung* (Bd. 1, S. 285–299). Weinheim: Juventa.
- 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.
- Schoofs, D., Wolf, O. T., & Smeets, T. (2009). Cold pressor stress impairs performance on working memory tasks requiring executive functions in healthy young men. *Behavioral Neuroscience*, 123, 1066–1075.
- Van der Schoot, M., Bakker Arkema, A. H., Horsley, T. M., & van Lieshout, E. C. D. M. (2009). The consistency effect depends on markedness in less successful but not successful problem solvers: an eye movement study in primary school children. *Contemporary Educational Psychology*, 34, 58–66.
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31, 154–165.
- Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern. Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27–52.
- Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology*, 33, 890–895.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*. Zeitschrift für Pädagogik, (Bd. 56, S. 296–306). Weinheim: Beltz.
- Shulman, L. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15, 4–14.
- Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57, 1–23.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, 85–95.
- Steckler, T., Kalin, N. H., & Reul, J. M. H. M. (2005). *Handbook of stress and the brain part 2: stress: integrative and clinical aspects*. Amsterdam: Elsevier.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: a meta-analysis. *Journal of Educational Psychology*, 104, 743–762.

- Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, 3, 86–94.
- Velasco, M., Gómez, J., Blanco, M., & Rodriguez, I. (1997). The cold pressor test: pharmacological and therapeutic aspects. *American Journal of Therapeutics*, 4, 34–38.
- 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, 85–94.
- Von der Embse, N., Ryan, S. V., Gibbs, T., Mankin, A. (2019). Teacher stress interventions: a systematic review. *Psychology in the Schools*, 56, 1328–1343.
- Weinert, F. E. (2000). Lehren und Lernen für die Zukunft – Ansprüche an das Lernen in der Schule. In *Pädagogische Nachrichten Rheinland-Pfalz*, 2 – Schulleben Schulkultur, Sonderseiten (S. 1–16).