

Zur Evaluation von Cross-Age Peer Tutoring im Physikunterricht An Evaluation of Cross-Age Peer Tutoring in Physics Classes

Marianne Korner*, Martin Hopf*

*Universität Wien, AECC Physik, Porzellangasse 4/2/2, 1090 Wien,
marianne.korner@univie.ac.at und martin.hopf@univie.ac.at
(Eingegangen: 26.07.2016; Angenommen: 13.10.2016)

Kurzfassung

Cross-Age Peer Tutoring (CAPT) stellt eine Lernform dar, bei der ältere Schüler/innen Jüngere beim Lernen unterstützen. Bisher war die Frage offen, ob die in anderen Fächern und Altersstufen gefundenen positiven Zusammenhänge auch im Physikunterricht mit 10- bis 14-Jährigen bestehen, da Physiklernen aufgrund der aus der Fachdidaktik bekannten Rahmenbedingungen zum Conceptual Change spezielle Bedingungen benötigt. In der vorliegenden Studie wurden innerhalb von zwei Unterrichtsjahren Schüler/innen in Elektrizitätslehre und in Optik ($N_E = 172$, bzw. $N_O = 141$) aus den Schulstufen 5 bis 8 einem klassenweisen Tutoring, mehrheitlich in Dyaden, unterzogen und beforcht. Die Gestaltung der Intervention orientiert sich an Merkmalen konstruktivistischer Lernumgebungen. Um die Methode zu evaluieren wurde in einem Prätest – Posttest – Follow-up-Testdesign das Wissen der Schüler/innen dokumentiert. Die Ergebnisse zeigen, dass CAPT bei den durchgeführten Wissenstests zu einer Verbesserung mit Effektstärken von 0,46 bis 0,62 führt. Durch multiple lineare Regression konnten der Prätestscore, die aktive Rolle im Tutoringprozess und die Muttersprache als Prädiktoren für die Posttestergebnisse identifiziert werden. Hingegen konnten nur schwache Zusammenhänge zur Motivation gezeigt werden. Die Persistenz der Methode stellte sich durch Analysen der Follow-up-Tests als zufriedenstellend heraus. Zusammenfassend lässt sich CAPT als wirksame Methode für den Physikunterricht empfehlen, da bei allen Beteiligten Lernfortschritte nachgewiesen werden konnten.

Abstract

Cross-Age Peer Tutoring (CAPT) describes a learning process where older students help younger ones to learn. Though previous studies reveal evidence that CAPT works effectively in various contexts, the question is left open if these results can be transferred easily to Physics classes: Physics learning requires different learning settings to enhance the necessary conceptual change. This study was conducted within two years comprising $N_e = 172$ students (electricity) and $N_o = 141$ students (optics) from grade 5 to 8. Students underwent a class-wide tutoring, mostly in dyads. The intervention was designed to show features of constructivist learning environments. In order to evaluate CAPT, a pretest – posttest – follow-up test design was chosen to monitor the students' knowledge. The results show clearly that CAPT enhances the test results with effect sizes between 0.46 and 0.62. Multiple linear regression models reveal that the pretest score, the active role within the tutoring process and the first language are relevant parameters to estimate the posttest scores. On the other side, correlations to the motivational parameters were weak. Analyzes of the follow-up test indicate that CAPT is sufficiently persistent. Summing up, CAPT enhances knowledge for all participating learners and can therefore be recommended as a method for physics teaching.

1. Einleitung

Maßnahmen zu finden, den Physikunterricht zu verbessern, ist insofern von Interesse, da in den OECD-Studien wie PISA 2006 [1] und PISA 2012 [2] österreichische Schülerinnen und Schüler in den Naturwissenschaften lediglich durchschnittlich abschnitten. Damit lagen sie 2006, ähnlich wie Deutschland, im Mittelfeld, 2012 sogar signifikant unter den deutschen Ergebnissen. Zumindest in Österreich, aber auch in Deutschland, ist somit Handlungsbedarf gegeben, vergleicht man die bei-

den Länder mit den besten PISA-Staaten hinsichtlich des naturwissenschaftlichen Unterrichts.

Mit dem Blick auf PISA liegt der Fokus der hier präsentierten Studie hinsichtlich des Alters der Schülerinnen und Schüler auf der Sekundarstufe 1 (also auf den 10- bis 14-Jährigen), die in Deutschland wie in Österreich den Abschluss der Pflichtschulzeit darstellt.

Peer Tutoring stellt eine Unterrichtsform dar, im Rahmen derer Lernende einander beim Lernen unterstützen. Das kann unabhängig von der Jahr-

gangsstufe, vom Thema oder vom Altersunterschied und der Auswahl der Schülerinnen und Schüler geschehen. Der Erfolg der Methode basiert auf der sozialen Nähe der Peers [3] und führt zu positiven Effekten, sowohl in der Einstellung der Lernenden als auch in den Lernergebnissen. Das wurde bereits in einigen Kontexten wie z. B. Leseverständnis, Mathematiknachhilfe [4], Problemlösestrategien [5] oder Fertigkeiten am Computer [3] gezeigt. Auch Hattie [6] spricht in seiner Meta-Meta-Studie im Zusammenhang von Peer Tutoring mit Effektstärken von 0,55 von beachtlich großen Gesamteffekten. Im naturwissenschaftlichen Bereich gibt es Hinweise darauf, dass Peer Tutoring zu konzeptuellen Erfolgen führt [7, 8].

Ein Ziel der vorliegenden Studie ist es, herauszufinden, ob sich durch den Einsatz von Peer Tutoring auch der Erfolg von Physikunterricht verbessern lässt und es zu vergleichbaren Effekten wie in anderen Bereichen führt oder nicht. Diese Fragestellung scheint insofern von Belang zu sein, da naturwissenschaftlicher Unterricht immer Schülervorstellungen zu berücksichtigen hat und auf einen *Conceptual Change* ausgerichtet ist [9]. Es wäre daher denkbar, dass sich nicht alle Ergebnisse aus anderen Bereichen auf Physikunterricht übertragen lassen. Darüber hinaus zeigt schon die Studie von Cohen et al. [4], dass es in unterschiedlichen Anwendungsgebieten zu stark differierenden Ergebnissen kommt.

Ein weiterer Aspekt betrifft die Umsetzung konstruktivistisch orientierter Unterrichtsansätze als Qualitätsmerkmal für Physikunterricht. Empirische Untersuchungen haben ergeben, dass sich diese Unterrichtsansätze als effektiver erweisen als traditioneller Physikunterricht [10]. Gleichzeitig wird in einer Untersuchung über die Praxis des Physikunterrichts, die Teil der IPN-Videostudie ist, bemängelt, dass sich im beobachteten Unterricht selten konstruktivistische Merkmale festmachen lassen [11]. In diesem Sinne versteht sich der hier vorgestellte Unterrichtsansatz als eine Möglichkeit, zentrale konstruktivistische Aspekte wie das Anknüpfen an Schülervorstellungen und Fragestellungen, die zum Denken anregen, in den Unterricht zu implementieren und so die Kluft zwischen Theorie und Unterrichtspraxis zu verkleinern

2. Peer Tutoring und Cross-Age Peer Tutoring

2.1. Begriffsklärung

Peer Tutoring hat eine lange und wechselvolle Tradition, beginnend in der Antike [12], über einen Aufschwung im England der Industriellen Revolution [3] bis es schließlich Objekt moderner bildungswissenschaftlicher Forschung wurde [13]. Mit dem Paradigma konstruktivistischer Sichtweisen, das sich in den Didaktiken der Naturwissenschaften in den letzten Jahrzehnten durchgesetzt hat, ist eine zeitgemäße Definition von Peer Learning, bzw. Peer Tutoring (geregelter Rollenverteilung) nötig:

„[Peer Learning] involves people from similar social groupings who are not professional teachers helping each other to learn and learning themselves by doing so“ [14, S. 631]. Diese lässt auch bereits erkennen, dass in moderneren Studien der Forschungsfokus auf *allen* Beteiligten des Tutoringprozesses liegt, nicht nur auf denen, die das Tutoring empfangen (Tutees), sondern auch auf jenen, die es leiten (Tutoren). Von Cross-Age Peer Tutoring (CAPT) spricht man dann, wenn die Tutoren älter als die Tutees sind [15, S. 1]. Die Bezeichnung Peer Tutoring inkludiert demnach in vielen Fällen aber bereits das Cross-Age Peer Tutoring als eine der vielen möglichen Variationen erfolgreicher Programme. In diesem Zusammenhang sprechen Robinson et al. [13] davon, dass möglicherweise die bei CAPT zu erwartenden Effekte jene des Peer Tutorings übertreffen.

Was die Wirksamkeit speziell von CAPT angeht, so sprechen zahlreiche empirische Studien und Metastudien von positiven Effekten. Hattie ordnet Peer Tutoring, ohne den cross-age Einfluss, hinsichtlich des Lerneffekts mit mittleren Effektstärken von 0,55 in die „*zone of desired effects*“ ein [6, S. 16]. Er betont, dass Tutees *und* Tutoren einen Gewinn zeigen, sowohl in sozialer Hinsicht als auch im Wissenserwerb. Eine ältere Metastudie [4] ordnet Peer Tutoring mit 0,4 größere mittlere Effektstärken zu als herkömmlichem Unterricht, abhängig von der angewandten Art des Tutorings. Diesen Befund unterstützt auch die Metastudie von Robinson et al. [13], wo angegeben wird, dass kürzere, strukturiertere, den Unterricht ersetzende Programme auf cross-age Basis erfolgreicher ablaufen.

Einen Überblick über die Effekte auf Tutoren kann man erstmals bei Cohen [4] finden. Es wird von kognitiven Effektstärken in der Größe von 0,33 und von Effekten auf die Einstellung zum Lernen von 0,42 gesprochen. Diese positiven Effekte auf die Tutoren betonen auch Topping [14] und Robinson [13]. Sind die zu vermittelnden Inhalte auch für die Tutoren neu, so empfehlen zwei Publikationen [3, 13], vorab ein Tutorentaining (Mentoring) abzuhalten.

Das Studium der Literatur zu Peer Tutoring ergibt einen Überblick über die Bereiche, in denen diese Methode bereits erforscht ist. In großer Zahl sind es Studien, deren Thema Mathematik-Nachhilfe ist [3, 4, 13]. Aber auch Bereiche wie die Verbesserung der Fertigkeiten am Computer [3] oder beim Lesen [4] werden besprochen. Cohen et al. berichten hier von unterschiedlichen Effektstärken: So findet man in der Mathematik Effektstärken in der Größenordnung von 0,6, beim Lesen hingegen liegen sie unter 0,3. Hinsichtlich des naturwissenschaftlichen Bereichs lassen sich nur Einzelstudien finden. Es sind je eine Studie zum Thema Heiz- und Kühlprozesse [7] und eine zur Photosynthese [8] zu erwähnen. Beide Studien beschreiben allerdings *Peer Collabo-*

ration, was dem Peer Tutoring zwar ähnlich ist, aber nicht mit der strengen Rollenverteilung des Tutorings arbeitet. Eine jüngere Arbeit [16] beschäftigt sich mit diversen naturwissenschaftlichen Themen, die auf Basis eines dem Tutoring ähnlichen Unterrichtskonzepts (Lernen durch Lehren [17]) unterrichtet werden. Zinn arbeitete auf cross-age-Basis und zeigte in dieser Arbeit, dass das Interesse, speziell auch der Mädchen, gesteigert wird und Prozesswissen aktiviert wird. Er empfiehlt eine Validierung der Ergebnisse auf Basis konkreter Inhalte. Eine weitere Forschungsarbeit untersucht Verbesserungspotenziale speziell beim Tutorentaining des in dieser Arbeit als *Cross-Age Tutoring* bezeichneten Prozesses im Zusammenhang mit Physikunterricht [18].

Die beschriebenen Altersstufen betreffen zum Teil Grundschulkindern [19], zum Teil die Sekundarstufe 1 [4], während Topping [12] Studierende an Colleges beforcht. Lediglich Fogarty und Wang [3] beschreiben Tutoren der sechsten bis achten Schulstufe. Es entsteht das Bild, dass die Altersgruppe der 10- bis 14-Jährigen in den Studien bisher unterrepräsentiert ist.

Die Interaktion zwischen Tutoren und ihren Tutees basiert auf einem freundschaftlichen Umgang auf Augenhöhe [13]. Im Gegensatz zur hierarchischen Lehrer-Schüler-Interaktion handelt es sich bei den Peers um Gleichgesinnte, die einander wechselseitig geben und voneinander nehmen [3]. Diese soziale Nähe begründet den Erfolg der Methode, obwohl Tutoren keine ausgebildeten Lehrkräfte sind. Um sie nutzen zu können, soll der Altersabstand zwischen Tutoren und Tutees nicht zu groß sein. Empirischen Studien zufolge lassen sich optimale Ergebnisse erzielen, wenn der Altersabstand zwei bis vier Jahre beträgt [13], bei größerem Altersunterschied ist Vorsicht geboten. Tutoring auf der gleichen Altersstufe hingegen funktioniert mit etwas geringeren Effekten [6].

CAPT ist hinsichtlich einiger Themen, vor allem der Mathematik, bereits gut erforscht. Im naturwissenschaftlichen Bereich sind uns jedoch außer den oben angeführten Arbeiten keine weiteren bekannt. Die Erforschung einer möglichen Lernwirksamkeit von CAPT in naturwissenschaftlichen, speziell physikalischen Kontexten erscheint jedoch wichtig, da hier bereits eine lange Forschungstradition über Schülervorstellungen und Konzeptwechsel (*Conceptual Change*) existiert [20]. Deren Ergebnisse stehen im Zentrum naturwissenschaftlicher, konstruktivistischer Lerntheorien. Als selbstbestimmte Lernform lässt sich CAPT besser in das Anforderungsprofil konstruktivistisch orientierten Unterrichts einfügen als eine lehrerzentrierte Unterrichtspraxis [20] und soll somit imstande sein, einen Konzeptwechsel zu unterstützen. Ob und in welcher Weise sich nun Ergebnisse aus anderen Kontexten tatsächlich auf Physikunterricht *und* die wenig

beforschte Population der 10- bis 14-Jährigen übertragen lassen, ist jedoch offen.

2.2. Forschungsfragen

Im Zentrum der Fragestellungen steht daher die Wirksamkeit und Praxistauglichkeit von CAPT als Gesamtpaket in verschiedenen Inhaltsbereichen. Dazu wurde im ersten Unterrichtsjahr die Wirksamkeit der Methode an sich getestet, in einem zweiten Unterrichtsjahr wurden weitere Inhaltsbereiche für CAPT erschlossen. Gemessen wird diese Wirksamkeit am Lernerfolg (operationalisiert in Wissenstests) aller Schülerinnen und Schüler. Die Untersuchungshypothese ist dahingehend gerichtet, dass Erfolge zu erwarten sind. Eine weitere Frage untersucht, ob auch die Tutoren den aus der Literatur berichteten Lernerfolg erzielen. Das hat Einfluss auf die praktische Umsetzbarkeit im Unterricht. Es wird versucht, Prädiktoren für die Lernleistungen zu finden. Schließlich wird die Persistenz der Methode CAPT geprüft. Darüber hinaus wird der Fragestellung nachgegangen, ob CAPT die Motivation der Schüler/innen Physik zu lernen steigern kann.

3. Methodik

Explizites Ziel der Studie ist es, CAPT als Unterrichtsmethode auf thematisch breiter Basis zu evaluieren. In der Tradition der Evaluationsforschung werden Wirkungen (das ist der Prozess) und Wirksamkeiten (das sind die Folgen) einer Maßnahme dokumentiert und geprüft [21, 22]. Als Evaluationskriterium zählt in der aktuellen Studie die Qualität der Intervention, die anhand der Lernergebnisse der Schüler/innen beurteilt wird. In anderen Worten: Es wurde vor allem der Wissenserwerb der Schüler/innen beforcht. Dazu wurden die unten beschriebenen Interventionen entwickelt, durchgeführt und im Sinne einer summativen Evaluation evaluiert. Diese wurde in einem Prä-, Post- und Follow-up-Testdesign operationalisiert.

Die Evaluation erfolgte im Feld und in authentischen Lernsituationen. Das hatte den Nachteil, dass die Untersuchungsbedingungen weniger gut konstant gehalten werden konnten und die beforchten Schülergruppen hinsichtlich des Alters und der Schulform variierten. Es hatte aber den Vorteil, dass gezeigt werden konnte, wie robust die Methode gegenüber abweichenden Voraussetzungen ist. Somit können auch Implikationen für die Schulpraxis leichter diskutiert werden.

3.1. Themen

Die Datenerhebung zur vorliegenden empirischen Studie fand über zwei Unterrichtsjahre hinweg statt. Im ersten Jahr wurden Einheiten zu Themen aus der Elektrizitätslehre behandelt, im zweiten Jahr solche aus der Optik. Eine Fokussierung auf spezielle Themen ist nach Ergebnissen aus der Lernforschung [23] und der Conclusio der Arbeit von Zinn [16] sinnvoll. Die Auswahl dieser Themen begründete sich zunächst aus der Curriculumsvalidität. Zur

Wahl standen Bereiche, die in den Lehrplänen der Klassenstufen der Tutoren und der Tutees stehen. Andererseits ist es sinnvoll, Themen zu wählen, zu denen Schülervorstellungen gut erforscht und dokumentiert sind. Somit fiel die Wahl auf Elektrizitätslehre und Optik. Innerhalb der Elektrizitätslehre wiederum wurde auf grundlegende Konzepte [24-26] wie den geschlossenen Stromkreis, die Richtung des elektrischen Stroms, die Konstanz der Stromstärke und den Zusammenhang zwischen Stromstärke, elektrischem Widerstand und der Helligkeit gleichartiger Lämpchen fokussiert. In der Optik wurden die Teilbereiche Spiegel und Schatten als besonders passend hinsichtlich einer Streuung im Alter der Tutees identifiziert. Grundlegende Konzepte, die hier adressiert wurden, sind die Vermittlung einer korrekten Sehvorstellung, der Lichtausbreitung als Strömen von Licht, von Schatten als Lichtmangel, des Ortes und der Eigenschaften des Spiegelbildes sowie des Zusammenhangs zwischen Reflexionsgesetz und Spiegelbild [z. B. 27, 28-30]. Die Intervention wurde auf Basis dieser Konzepte designt und angelehnt an bereits erprobte Materialien [31, 32]. Die Umsetzung und Einbettung der Arbeitsaufträge in anschließende Reflexions- und Diskussionsprozesse erfolgte anhand ausgewählter Merkmale konstruktivistischer Lernumgebungen [11]. In der Durchführung wurde mit den Lehrkräften vereinbart, dass sie die zu vermittelnden Inhalte nicht vor oder zwischen den Wissenstests im Regelunterricht thematisierten und dass auch über die Testitems nicht in der Klasse diskutiert wurde.

3.2. Sample

Zu Beginn der Untersuchung waren in der Elektrizitätslehre $N = 172$ Schülerinnen und Schüler Gegenstand der Untersuchungen. Sie teilten sich auf 9 Schulklassen auf – mehrheitlich aus der Hauptschule, aber auch aus dem Gymnasium – und deckten die Schulstufen 6 bis 8 ab. Der Anteil der Jungen lag bei 63 %, der der Mädchen bei 37 %. Dieser Anteil an Jungen und Mädchen entspricht der österreichischen Verteilung in der Hauptschule [33].

Die Verteilung der Rollen ergab im beforschten Sample 55 % Tutoren, 25 % Tutees und 20 % in beiden Rollen. Tutees, die nicht aus der Sekundarstufe 1 stammten, wurden dabei nicht berücksichtigt, weshalb es deutlich mehr Tutoren als Tutees gab.

Da CAPT größtenteils auf verbaler Interaktion beruht, wurde auch erhoben, welche Sprache zu Hause gesprochen wurde. Dabei interessierte nicht die Sprache als solche, sondern ob es Deutsch war oder eine andere Sprache. Für 68 % der Schülerinnen und Schüler war die Muttersprache Deutsch, für 32 % war es eine andere Sprache.

An der zweiten Schleife, die Optik betreffend, waren $N = 141$ Lernende beteiligt, von denen etwa 62 % das Thema Schatten bearbeitete und 38 % das Thema Spiegel. Es wurden die gleichen Schulstufen

wie in der Elektrizitätslehre abgedeckt. Der Anteil der Jungen lag hier bei 59 %, der der Mädchen bei 41 %, also ähnlich wie beim ersten Sample. Der Anteil der Schülerinnen und Schüler mit deutscher Muttersprache lag bei 73 %, der mit nicht deutscher Herkunft bei 27 %. Klassenweise Analysen zeigten hier, dass die Verteilung jedoch sehr heterogen war. So schwankte der Anteil an Jugendlichen mit nicht deutscher Muttersprache zwischen 80 % und 0 %.

Es wurden im Rahmen der Studie jeweils ganze Klassen, nicht nur einzelne Schülerinnen und Schüler als Tutoren oder Tutees eingesetzt und beforcht. Somit liegt eine Klumpenstrichprobe vor. Um deren Nachteile zu minimieren, wurde versucht ein möglichst unverzerrtes Sample zu ziehen, indem vier unterschiedliche Schulstandorte und Schulformen in unterschiedlichen Gegenden gewählt wurden.

3.3. Motivation

Parallel zu den Wissenstests wurde die Motivation der Tutor/innen erhoben, um sie mit den Lernergebnissen zu verknüpfen. In diesem Zusammenhang erschien aufgrund der differenzierten Betrachtung der Motivation die *Self Determination Theory* nach Deci und Ryan [34] am geeignetsten. Zu dieser existiert ein Messinstrument in englischer Sprache, das *Intrinsic Motivation Inventory* [35], das teilweise in die deutsche Sprache übersetzt und durch altersgerechte Formulierungen für die Altersstufe angepasst wurde [36]. Es wurden die Tutoren getestet, um ihre Motivation für die CAPT-Intervention zu erfassen. Der Testzeitpunkt für diesen Test fiel mit dem Posttest, unmittelbar nach der Intervention, zusammen.

3.4. Mentoring und Tutoring

Die CAPT-Intervention gliederte sich in zwei Teile: Als Vorbereitung auf ihre Tutorenrolle erhielten die Tutorenklassen ein Mentoring, während das Tutoring die eigentliche Arbeit der Tutoren mit ihren Tutees bezeichnet.

Das Mentoring erfüllte mehrere Aufgaben: Zum einen diente es der inhaltlichen Klärung für die Tutoren, andererseits sollte es ihnen Gelegenheit geben, ihre eigenen (Schüler-)Vorstellungen kennenzulernen, zu reflektieren und weiterzuentwickeln. Indem die Tutoren dazu eingeladen wurden, Problemstellungen und Materialien für ihre Tutees auszusuchen, zu reihen oder zu ergänzen, wurden sie als Experten für das Lernen der Jüngeren angesprochen. Das sollte im Sinne der *Self Determination Theory* der Motivation [34], das Kompetenzerleben und die Autonomie steigern und somit die Motivation erhöhen.

Alle Mentorings wurden von derselben Person, nach dem gleichen Muster durchgeführt, um diesen Faktor konstant zu halten. Nach einer Begrüßung und Kurzinformation wurden die Prätests durchgeführt. Danach wurden Arbeitsblätter mit theoretischen und experimentellen Aufgabenstellungen

sowie das zugehörige Experimentiermaterial ausgeteilt. Experimentelle Aufgaben sollten nach der P-O-E-Strategie [37] behandelt werden: Diese stellt eine dreistufige Strategie zur Einbettung von Experimenten dar: Zuerst sollen – in diesem Fall – die Tutees durch die Tutoren aufgefordert werden vorherzusagen (*predict*), was passieren wird. Danach wird beobachtet (*observe*), was im Experiment passiert, und zum Schluss werden mögliche Erklärungen diskutiert (*explain*). *Predict* spielt hier eine zentrale Rolle. Es dient dazu, dass sich die Lernenden ihres mentalen Modells bewusst werden und dass auch die relevanten p-prims aktiviert werden [38]. Andererseits ermöglicht es den Tutoren, die Vorstellungen ihrer Tutees kennenzulernen, was die Basis für weitere Diskussionen darstellt.

Die Aufgaben waren einzeln oder in Partnerarbeit zu behandeln. Auf diese selbstgesteuerte Lernphase folgte eine moderierte Gruppendiskussion, in der korrekte mit inkorrekten Erklärungen verglichen wurden. Es wurde versucht, Gründe zu finden, weshalb inkorrekte Vorstellungen entwickelt worden waren, und es wurde besprochen, wie man diese in korrekte Vorstellungen überführen könnte. Abschließend wurden von den Tutoren im Klassensens Aufgaben ausgewählt, die sie für ihre speziellen Tutees als altersgemäß empfanden. Es stellte sich heraus, dass, angelehnt an die Aufgaben der Arbeitsblätter, die die Tutoren bearbeitet hatten, immer der gleiche Grundstock an Aufgaben gewählt wurde. Dazu ergänzend wählten die Gruppen einige wenige Zusatzaufgaben aus.

Die Dauer des Mentoring betrug zwischen 60 und 80 Minuten und es wurde eine bis zweieinhalb Wochen vor dem Tutoring durchgeführt.

Vorbereitend zum eigentlichen Tutoring konnten sich die Tutoren mit dem Experimentiermaterial und den Aufgaben noch einmal für 20 bis 30 Minuten beschäftigen, damit sie nach der Pause, die sich zwischen Mentoring und Tutoring ergeben hatte, wieder damit vertraut wurden. Da es sich im Sample um mehrheitlich leistungsschwächere Schülerinnen und Schüler handelte, wurden auch Hilfekärtchen ausgeteilt: Sie beinhalteten auf der Vorderseite die Aufgabenstellung und auf der Rückseite Lösungshinweise. Die Hilfekärtchen dienten dazu, den Tutoringprozess zu strukturieren, was bei der Arbeit mit leistungsschwächeren Jugendlichen sinnvoll ist [39]. Den Abschluss dieser Wiederholung bildeten Hinweise auf den Umgang mit den Tutees. Abgesehen von der Erinnerung an elementare Regeln der Höflichkeit betraf das die Umsetzung der P-O-E-Strategie beim Experimentieren.

Das eigentliche Tutoring begann mit dem Zusammentreffen der beiden Klassen, Tutorenklasse und Tuteeklasse. Die Zuordnung der einander fast immer unbekanntenen Schülerinnen und Schüler erfolgte zufällig. Es wurden nach Möglichkeit Tutor-Tutee-Dyaden gebildet, die zur intensiveren fachlichen

Auseinandersetzung führen [14]. Im Falle ungleicher Schülerzahlen wurden Dreiergruppen gebildet. Es wurde kein besonderer Wert auf gleichgeschlechtliche Paare gelegt, da die Praxistauglichkeit und nicht der Genderaspekt im Forschungsfokus stand. Desgleichen wurde keine Rücksicht auf die fachlichen Fähigkeiten der Teilnehmenden bei der Paarbildung genommen, sodass es vorkommen konnte, dass auch leistungsschwache Tutoren mit leistungsstarken Tutees zusammenarbeiteten. Die Schulstufe der Tutoren variierte zwischen 6. und 8. Schulstufe, die der Tutees zwischen 2. und 7. Schulstufe. Das Tutoring dauerte 30 bis 45 Minuten und endete, wenn die Unterrichtsstunde vorüber war oder die Beteiligten mit der Bearbeitung ihrer Aufgaben fertig waren.

Die Evaluation von CAPT als Gesamtpaket erfolgte gewollt in authentischen Lernsituationen, d. h. nicht unter strengen Laborbedingungen, sondern im Feld, damit möglichst breitgestreute, praxisnahe Anwendungssituationen erfasst werden konnten. Daher wurden auch Schülergruppen der Sekundarstufe 1 in unterschiedlichen Settings befohrt. Diese Settings und die damit verbundene Rollenverteilung der Schülerinnen und Schüler werden nun näher beschrieben.

Die Basisvariante des Tutoringprozesses zeigt Abb. 1. Klasse A erhielt ein Mentoring und alle Schülerinnen und Schüler dieser Klasse agierten als Tutoren. Klasse B wurde von Klasse A betreut und besteht daher aus Tutees.

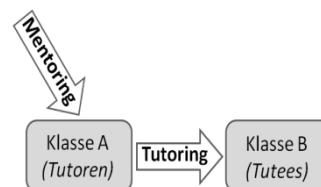


Abb. 1: Ablauf einer einfachen Mentoring-Tutoring-Sequenz

In manchen Fällen führte Klasse A ein weiteres Mal ein Tutoring mit einer zweiten Tutorenklasse (B') durch (Abb. 2).

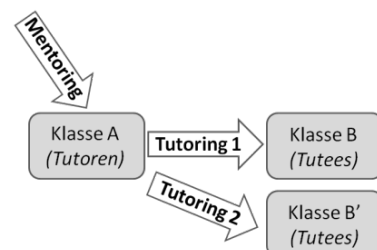


Abb. 2: Tutoring-Sequenz, bei der Klasse A zwei Tutorings durchführt.

Die dritte Möglichkeit einer CAPT-Intervention im Rahmen dieser Studie zeigt Abb. 3. Hier erhielt eine zuvor getutorte Klasse (D) ein eigenes Mentoring (Mentoring 2) und arbeitete ihrerseits mit Klasse E. Schülerinnen und Schüler der Klasse D

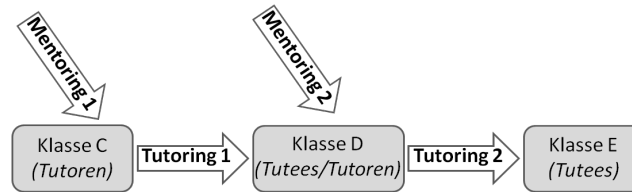


Abb. 3: Mentoring-Tutoring-Sequenz mit Klasse D in der Doppelrolle

3.5. Messinstrumente

Der Lernfortschritt wurde im ersten Studienjahr, in der Elektrizitätslehre, mit einem Testinstrument gemessen, das sich an Basiskonzepten der Elektrizitätslehre orientiert [40]. Aus diesem Instrument wurden fünf konzeptuell zur Intervention passende Items ausgewählt, die bei summativer Zählweise im getesteten Sample über eine Reliabilität von $\alpha = 0,78$ (Prätest) bzw. $\alpha = 0,85$ (Posttest) verfügten. Die maximal erreichbare Punktezahl waren hier 9 Punkte.

Im zweiten Studienjahr, in der Optik, stand für die Bereiche Spiegel und Schatten zum Zeitpunkt der Untersuchung kein psychometrisch valides und reliables Testinstrument zur Verfügung. Auf Basis von Vorarbeiten in der Forschergruppe konnte zum Thema Spiegel zumindest auf erprobte Items zurückgegriffen werden [41, 42]. Für die Anzahl von fünf Items ergab sich eine Reliabilität von $\alpha = 0,63$. Die maximal erreichbaren Punktezahlen waren 11 (Spiegel) bzw. 12 Punkte (Schatten).

Zum Thema Schatten existiert bis dato keine derartige Untersuchung, weshalb einige passend erscheinende Items der Literatur [z. B. 43] entnommen, innerhalb der Forschergruppe expertenvalidiert und in einer kleinen Vorstudie an $N = 9$ Schülerinnen und Schülern der 9. Schulstufe getestet wurden. Für dieses Testinstrument, das aus acht Items bestand, ergab sich eine Reliabilität von $\alpha = 0,66$, was in diesem Zusammenhang als zufriedenstellend bewertet werden kann.

Da es Ziel der Studie war, CAPT in unterschiedlichen Themenbereichen zu untersuchen und da die vorhandenen Testinstrumente so unterschiedlich waren, unterschieden sich auch die Forschungsfoci in beiden Studienjahren: Im ersten Studienjahr wurde anhand der Elektrizitätslehre die Wirksamkeit der Methode an sich und der Einfluss der Rollen untersucht. Im zweiten Studienjahr wurde auf Basis der Ergebnisse des ersten Jahres die Erweiterung auf andere Themen untersucht.

waren somit in der Doppelrolle: zuerst als Tutees, danach als Tutoren.

Somit gab es aufgrund des Untersuchungsdesigns drei mögliche Rollen, die die Schülerinnen und Schüler innehaben konnten: die der Tutoren, die der Tutees und die Doppelrolle als Tutees/Tutoren.

4. Ergebnisse

4.1. Prätests und Posttests

Der erste Schritt der Analyse bestand in einem Vergleich der Prätests mit den Posttests über das gesamte Sample hinweg, jedoch nach Themen getrennt. Vergleicht man für die Elektrizitätslehre Prä- und Posttests, so erhält man eine hochsignifikante Steigerung ($t_{163} = 5,826$ und $p < 0,001$, vgl. Tab. 1) bei einer Effektstärke nach Cohen von $d_z = 0,46$. Die gleiche Analyse liefert für die Optik im Bereich Schatten ($t_{74} = 4,271$ mit $p < 0,001$, vgl. Tab. 1) und im Bereich Spiegel ($t_{45} = 4,205$ mit $p < 0,001$, vgl. Tab. 1) ebenfalls hochsignifikante Steigerungen. Die dazugehörigen Effektstärken betragen $d_{z, \text{Schatten}} = 0,49$ und $d_{z, \text{Spiegel}} = 0,62$. Eine Beurteilung dieser Effekte im bildungswissenschaftlichen Zusammenhang kann über Hatties [6] Klassifizierung erfolgen: Er bezeichnet Effektstärken ab 0,4 als *zone of desired effects*, also als solche, die über dem Median aller von ihm erfassten Maßnahmen liegen. Eine Effektstärke von 0,62, wie beim Themenbereich Spiegel, charakterisiert er sogar als großen Effekt.

Posttest – Prätest	Elektrizitäts- lehre	Spiegel	Schatten
Mittelwert	1,42	1,07	0,92
SD	3,12	1,72	1,87
T	5,826	4,205	4,271
df	163	45	74
Sig. (2-seitig)	< 0,001	< 0,001	< 0,001

Tab. 1: Überblick über die Mittelwertverschiebungen zwischen Prätest und Posttest (Statistik bei gepaarten Stichproben) für die einzelnen Themenbereiche

4.2. Rollen in der Elektrizitätslehre

Eine ANOVA der Prätests aus der Elektrizitätslehre legte offen, dass die Prätests hier sehr homogen hinsichtlich der unterschiedlichen Klassen, Schulformen und Altersstufen waren. Mit anderen Worten: Trotz der doch relativ unterschiedlichen Parameter präsentierte sich die Stichprobe so, dass sie dieselbe Population repräsentiert. Obendrein stand ein elaboriertes Messinstrument zur Verfügung, sodass es möglich war, den Einfluss der Rolle an diesen Daten zu erforschen.

Zunächst standen die Veränderungen bei den Tutoren im Forschungsfokus. Hier wurden Daten von Schülerinnen und Schülern ausgewertet, die ausschließlich die Tutorenrolle eingenommen hatten. Es zeigte sich, dass sich die Tutoren hochsignifikant mit einer Effektstärke von $d_z = 0,42$ verbesserten ($t_{90} = 3,976$, $p < 0,001$, vgl. Tab. 2). Dieser Befund ist im Einklang mit den Berichten aus vorangegangenen Studien [3, 4, 12].

Posttest – Prätest	Elektrizitätslehre nur Tutoren
Mittelwert	1,24
SD	2,98
T	3,976
df	90
Sig. (2-seitig)	< 0,001

Tab. 2: Prätest-Posttest-Vergleiche für die Tutoren in der Elektrizitätslehre (gepaarte Stichproben)

Betrachtet man nun das gesamte Sample in der Elektrizitätslehre, so können, wie im vorigen Abschnitt beschrieben, drei unterschiedliche Rollen im Lernprozess identifiziert werden: die der Tutees, die der Tutoren und die der Tutees/Tutoren, also jener Schüler/innen, die einmal ein Tutoring empfangen haben und danach eines in der Rolle der Tutoren selbst durchgeführt haben. Ausgehend von in den getesteten Parametern homogenen Prätests, zeigen sich in den Posttests unterschiedliche Ergebnisse (Tab. 3).

Posttest Elektrizitätslehre	N	Mittelwert	SD
Tutoren	92	5,99	2,70
Tutees/Tutoren	38	6,76	2,51
Tutees	35	4,46	3,09

Tab. 3: Ergebnisse im Posttest der Elektrizitätslehre, nach Rollen aufgespalten

Eine ANOVA dieser Daten lässt den Schluss zu, dass sich die Lernergebnisse der beteiligten Schülerinnen und Schüler in Abhängigkeit von der Rolle hochsignifikant unterscheiden ($F_{2,162} = 6,716$, $p = 0,002$). Eine weitere Analyse des systemati-

schen Varianzanteils wurde mittels geplanter Kontraste durchgeführt. Diese Vorgehensweise bietet sich eher an als post-hoc-Tests, da aufgrund der Forschungsergebnisse, die in der Literatur berichtet wurden [4, 13, 14], Unterschiede zwischen den Rollen zu erwarten waren. Somit ist ein systematisches, theoriegeleitetes Vorgehen vorzuziehen. Es wurden alle möglichen Kontraste getestet. Dabei ergab sich, dass der bedeutendste Unterschied zwischen der *aktiven Rolle* im Tutoringprozess (Tutoren und jene in der Doppelrolle) und der *passiven Rolle* (Tutees) lag ($t_{49,83} = 3,320$, $p = 0,002$). Dabei fasst die Variable aktive Rolle alle Schülerinnen und Schüler zusammen, die während des Tutoringprozesses zumindest einmal als Tutoren (aktiv) erklären mussten. Den Tutees, die die Erklärungen empfangen, wird hier eine *passive Rolle* zugewiesen. Darüber hinaus konnten keine signifikanten Unterschiede innerhalb der aktiven Rolle identifiziert werden, wie dies wegen der unterschiedlichen Interventionszeit hätte möglich sein können ($t_{74,072} = -1,564$, $p = 0,122$).

4.3. Multilineare Regression

Die Ergebnisse aus der Elektrizitätslehre sowie der Einfluss verschiedener Prädiktoren wurden nach sorgfältiger Prüfung der Voraussetzungen in multiplen linearen Regressionsmodellen (MLR) zusammengefasst, wobei die Kriteriumsvariable das jeweilige Posttestergebnis war. Als Prädiktoren wurden die erhobenen Parameter theoriebasiert und entsprechend ihrer Korrelationen zum Posttestergebnis ausgewählt und getestet. Tab. 4 stellt für vier dieser Modelle die Regressionskoeffizienten und die standardisierten Betas gemeinsam mit ihren Signifikanzen dar. Dabei bezeichnet die Variable *Ist Tutor* alle Schüler/innen, die zumindest einmal in der aktiven Rolle als Tutoren waren (Tutoren und jene in der Doppelrolle).

Tabelle 5 gibt darüber hinaus unter anderem den korrigierten Determinationskoeffizienten (R^2), der als die durch das Modell aufgeklärte Varianz interpretiert werden kann, und die Effektstärken an.

Alle Modelle können, was die Prädiktoren betrifft, gut gedeutet werden. Modell 3, das als Prädiktoren den Prätest, die aktive Rolle und die Muttersprache beinhaltet ($F_{\text{empirisch}} = 17,004$, $p < 0,001$ bei einer Effektstärke von 0,37) wird jedoch präferiert, da alle Prädiktoren zumindest auf einem Niveau von 0,01 signifikant sind und die durch dieses Modell aufgeklärte Varianz den größten Wert von 25,4 % beträgt. Interessant an Modell 4 ist die Tatsache, dass weder das Geschlecht noch die Muttersprache signifikanten Einfluss auf die Posttestergebnisse haben.

Mit Modell 3 ist es nun möglich, die Posttestergebnisse auf Basis der Prätests, der Rolle und der Muttersprache zu schätzen. So erwartet man für einen Probanden, der im Prätest einen Punkt erzielt, Deutsch als Muttersprache hat und Tutee war, einen Posttestscore von 3,4 Punkten. Wäre der Proband

hingegen bei gleichen Voraussetzungen Tutor gewesen, so sagt das Modell einen Posttestscore von 5,8 Punkten voraus. Für einen Tutor, der ebenfalls

einen Punkt im Prätest erzielte, für den Deutsch aber nicht die Muttersprache ist, ergibt sich ein geschätzter Posttestscore von 5,0 Punkten.

Modell	Regressionskoeffizient b	Standardfehler	Standardisiertes Beta	T	Sig.
1 (Konstante)	3,808	0,430		8,862	< 0,001
Prätest	0,456	0,082	0,426	5,565	< 0,001
2 (Konstante)	2,327	0,639		3,643	< 0,001
Prätest	0,406	0,081	0,379	4,996	< 0,001
Ist Tutor	1,956	0,639	0,232	3,063	0,003
3 (Konstante)	1,428	0,712		2,006	0,047
Prätest	0,385	0,080	0,360	4,827	< 0,001
Ist Tutor	2,215	0,633	0,263	3,499	0,001
Muttersprache	1,155	0,437	0,195	2,643	0,009
4 (Konstante)	2,234	0,930		2,401	0,018
Prätest	0,374	0,080	0,349	4,656	< 0,001
Ist Tutor	2,367	0,644	0,281	3,676	< 0,001
Muttersprache	0,836	0,502	0,141	1,666	0,098
Geschlecht	-0,274	0,421	-0,048	-,650	0,517
Note	-0,275	0,231	-0,103	-1,194	0,235

Tab. 4: Multiple lineare Regressionsmodelle im Überblick

dell	R	R ²	Korrigiertes R ²	Standardfehler d. Schätzers	F _{empirisch}	Effektstärke f ²
1	0,426	0,181	0,175	2,554	30,970	0,22
2	0,483	0,233	0,222	2,481	21,104	0,30
3	0,520	0,270	0,254	2,429	17,004	0,37
4	0,529	0,280	0,253	2,430	10,574	0,39

Tab. 5: (Korrigierte) Determinationskoeffizienten (R²) und Effektstärken zu den MLR-Modellen

4.4. Follow-up-Analysen

Um die Persistenz der CAPT-Intervention beurteilen zu können, wurden nach Themengebieten sortiert je zwei Testzeitpunkte (Prätest, Posttest, Follow-up-Test) miteinander durch t-Tests verglichen. Auf eine einfaktorische Varianzanalyse mit Messwiederholung wurde hier verzichtet, da eine der Voraussetzungen dafür, die Balanciertheit des Designs, nur unter erheblichem Datenverlust zu erfüllen gewesen wäre. Abb. 4 zeigt im Überblick eine grafische Darstellung der Testergebnisse zu allen drei Testzeitpunkten und für alle drei behandelten Themenbereiche. Aufgrund der unterschiedlichen Testinstrumente sind hier allerdings die absoluten Punktezahlen nicht miteinander vergleichbar und nur die relativen Änderungen interpretierbar. Diese lassen sich Tab. 6 entnehmen. Während die Grafik Trends in den Testergebnissen erkennen lässt, zeigt die zugehörige Tabelle auch die Ergebnisse der paarweisen t-Tests.

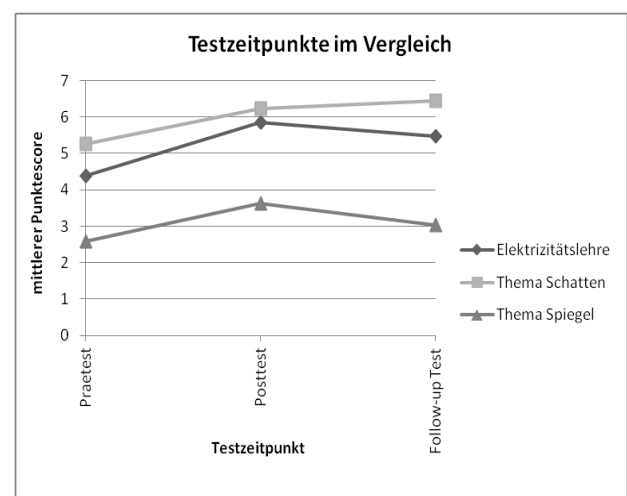


Abb. 4: Mittlere Punktescores für alle drei Testzeitpunkte (Praetest, Posttest, Follow-up Test) und die Themenbereiche aus der Elektrizitätslehre und der Optik

Thema	Vergleich	Mittelwert	T	Sign. (2-seitig)
Elektrizitätslehre	Prä-Post	-1,45	-6,086	< 0,001
	Prä-Follow-up	-1,08	-4,488	< 0,001
	Post-Follow-up	0,37	1,982	0,048
Schatten	Prä-Post	-0,92	-4,271	< 0,001
	Prä-Follow-up	-1,175	-4,665	< 0,001
	Post-Follow-up	-0,225	-0,927	0,354
Spiegel	Prä-Post	-1,07	-4,205	< 0,001
	Prä-Follow-up	-0,462	-1,297	0,196
	Post-Follow-up	0,577	1,897	0,059

Tab. 6: Überblick über die Vergleiche von jeweils zwei Testzeitpunkten, geordnet nach Themengebieten

Wie zu Beginn bereits beschrieben, sind für alle Themen die Ergebnisse der Posttests hochsignifikant besser als jene der Prätests. Ein weiterer Vergleich betrifft nun die Ergebnisse der Posttests mit jenen der Follow-up-Tests. Ein Rückgang im mittleren Punktescore wird hier als „Vergessen“ der zuvor in der Intervention erarbeiteten Inhalte interpretiert. In keinem der Fälle ist dieser Rückgang hochsignifikant. Interessanterweise gibt es für die Daten aus dem Themenbereich Schatten sogar noch eine Steigerung im Follow-up-Test im Vergleich zum Posttest. Diese ist allerdings ebenfalls nicht signifikant ($p = 0,354$). Für den Themenbereich Spiegel ist der Post-Follow-up-Vergleich tendenziell signifikant ($p = 0,059$), für die Elektrizitätslehre fällt der Rückgang auf einem 5%-Niveau signifikant aus ($p = 0,048$).

Vergleicht man die Prätests mit den Follow-up-Tests, so zeigt sich, dass für die Elektrizitätslehre dennoch eine hochsignifikante Verbesserung verbleibt ($p < 0,001$), der einer Effektstärke von 0,36 entspricht. Ein ähnlicher Befund ergibt sich für den Themenbereich Schatten: Die Steigerung ist hochsignifikant mit einer Effektstärke von 0,49. Lediglich für den Themenbereich Spiegel fällt der gleiche Vergleich nicht signifikant ($p = 0,196$) mit einer Effektstärke von 0,32 aus.

Zusammenfassend lassen die Vergleiche vorangegangener Tests mit den Follow-up Tests den Schluss zu, dass CAPT in allen getesteten Bereichen zu einem Wissenszuwachs führt. Dieses Wissen ist auch nach einigen Wochen statistisch gesehen in dem Maße verfügbar, dass Inferenzen auf die Population möglich sind und auch hier nachhaltige Effekte zu erwarten sind.

5. Fazit und Ausblick

Vom Standpunkt konstruktivistischer Unterrichtstheorien aus bietet CAPT per se die Möglichkeit, einige zentrale Aspekte konstruktivistischer Lernumgebungen umzusetzen [44]. Im Zentrum stehen die Lernenden. Das Tutoring als Lernumgebung ermöglicht es, Wissen in der sozialen Auseinander-

setzung mit dem jeweiligen Lernpartner eigenständig und individuell zu konstruieren.

Wie in einer parallel durchgeführten Studie gezeigt werden konnte [45], führt das intensive eins-zu-eins-Setting (ein Tutor mit einem Tutee) zu Aushandlungsprozessen über plausible Erklärungen und leitet somit eine Reflexion über die eigenen Vorstellungen sowie über jene der Lernpartner ein. Indem die Tutoren als Experten für das Lernen der Jüngeren angesprochen wurden und innerhalb gewisser Vorgaben ein wenig Wahlfreiheit über die Umsetzung der Inhalte hatten, wurden Eigenständigkeit und Selbstverantwortung angesprochen. Die zur Verfügung gestellten Materialien orientierten sich bestmöglich an der Lebenswelt der Schülerinnen und Schüler, indem z. B. Schatten und Spiegelphänomene, die aus dem Alltag bekannt sind, besprochen wurden. Das Tutoring deckt somit einen Teil der Eigenschaften konstruktivistischer Lernumgebungen ab, wie sie von den oben genannten Autoren charakterisiert wurden.

Indem sich die CAPT-Intervention an den Vorstellungen der Lernenden orientierte, wurde ein weiterer wichtiger, wenn nicht sogar der zentrale Aspekt konstruktivistischer Lernumgebungen umgesetzt. Obwohl die eigentliche Intervention als kurz zu bezeichnen ist, mag die Orientierung an Schülervorstellungen einen der Gründe darstellen, dass die gefundenen Lernwirksamkeiten von CAPT im Prä-Post-Vergleich deutlich über jenen traditionellen Unterrichts liegen [6, S. 17].

Aus der beforschten Unterrichtspraxis scheinen sich gewisse Gelingensbedingungen für CAPT abzuzeichnen. In zwei Parallelstudien wurden die Tutor-Tutее-Interaktionen genauer beleuchtet. Diese Studien arbeiteten nach dem Mixed Methods Approach [46] und verbanden Fragebogendaten mit Interviews. Es zeigte sich, dass die Tutees ihre Tutoren nicht so sehr als Lehrpersonen, sondern auf freundschaftlicher Basis wahrnahmen [47], was die Befunde aus der Literatur unterstreicht [3]. Tutees scheinen darüber hinaus das Rollenverhalten ihrer

Tutoren zu übernehmen [48]. Das führt dazu, dass das Kompetenzerleben der Tutees steigt, je besser die Tutoren auf Schülervorstellungen eingehen können.

Die quantitativen Analysen im Rahmen dieser Studie legen offen, dass CAPT in der Sekundarstufe 1 zu Lernerfolgen führt, was aufgrund älterer Studien, die sich zum Großteil mit ganz anderen Altersstufen beschäftigten, nicht a priori klar war. Darüber hinaus konnte gezeigt werden, dass diese Lernerfolge in Teilbereichen des Faches Physik beobachtbar sind, obwohl oder gerade weil hier der Konzeptwechsel eine zentrale Position einnimmt. Ältere Studien aus der Literatur lassen Untersuchungen zu Peer- und Cross-Age-Peer-Tutoring-Interventionen in Verbindung mit Schülervorstellungen und Conceptual Change vermissen. Die verbesserten Testergebnisse der aktuellen Studie können dahingehend interpretiert werden, dass CAPT allein durch sein Design zu belastbareren, wissenschaftlich anschlussfähigen Konzepten führt und einen Konzeptwechsel zumindest initiiert. Aufgrund des Untersuchungsdesigns sind jedoch keine verlässlichen Aussagen über detaillierte konzeptuelle Veränderungen möglich und auch der Mechanismus des Konzeptwechsels konnte nicht beleuchtet werden. Diesem könnte im Rahmen einer eigenen Untersuchung nachgegangen werden.

Die Wirksamkeit von CAPT *insgesamt* konnte im Rahmen dieser Studie belegt werden, die einzelnen Komponenten des Lernerfolges konnten jedoch nicht identifiziert werden. Insbesondere ist der Einfluss des Mentorings selbst nicht untersucht worden. Da im Zentrum der Forschungsfragen stand, ob CAPT im physikalischen Kontext *überhaupt* wirkt, wurden die einzelnen Komponenten, die zum Erfolg führen können, nicht extra untersucht. Die Frage, ob das bei einer derart komplexen Intervention, die so viele Details in sich vereint, überhaupt möglich ist, sei dahingestellt. Möglicherweise erklären einzelne Komponenten nur unbefriedigend wenig Varianz, obwohl das Gesamtpaket wirkt. Trotzdem sollte ein Punkt nicht außer Acht gelassen werden, dessen Einfluss zu diskutieren ist: das Mentoring. Bei kritischer Betrachtung könnte man behaupten, dass der Erfolg von CAPT, vor allem bei den Tutoren, allein darauf basiert. Aufgrund der auftretenden hohen Effektstärken zwischen 0,46 und 0,62 ist aber davon auszugehen, dass zur Wirksamkeit weit mehr beiträgt als das Mentoring. In der Hattie-Studie [6, S. 17] werden für die Wirksamkeit von Lehrer/inne/n, die man ohne besondere Maßnahme vor die Klasse stellt, Effektstärken von 0,20 bis 0,40 angegeben. Da die Effektstärken von CAPT deutlich darüber liegen, kann daraus geschlossen werden, dass sie sich aus der Summe der Maßnahmen ergeben. Dennoch stellt eine Untersuchung zum Einfluss des Mentorings eine lohnenswerte Aufgabe für eine Folgestudie dar.

Das gewählte Untersuchungsdesign mit Prä-Post- und Follow-up-Tests verzichtet auf die Einbindung von Kontrollgruppen. Der Hauptgrund dafür lag darin, dass sich das Forschungsprojekt über die Evaluierung hinaus mit der Entwicklung der Unterrichtsinterventionen befasste. Eine Bereitstellung weiterer randomisierter Gruppen hätte den Rahmen gesprengt und die schulpraktische Durchführung infrage gestellt. Da aber Studien [z. B. 6] über die Lernwirksamkeit unterschiedlicher Stile und Methoden existieren, lässt sich jene von CAPT auch ohne eine Kontrollgruppe einordnen und die gefundenen Effektstärken lassen sich bewerten.

Die MLR-Modelle legten offen, dass die Posttestergebnisse mit denen der Prätests und der Rolle der Schüler/innen signifikant zusammenhängen. Das war aufgrund des Studiums der Literatur und nach den Ergebnissen der ANOVAs zur Elektrizitätslehre auch zu erwarten. Der Zusammenhang des dritten Prädiktors, der Muttersprache, aus Modell 3 mit dem Posttest ist jedoch komplexer und kann lediglich auf Klassenebene erfasst und interpretiert werden [49].

Obwohl nicht im Forschungsfokus stehend, ergab sich darüber hinaus aus dem MLR-Modell, dass das Geschlecht der Probanden lediglich einen sehr geringen und nicht signifikanten Einfluss auf das Posttestergebnis hat. Das kann vorsichtig dahingehend interpretiert werden, dass mit CAPT eine Unterrichtsform für das Fach Physik gefunden wurde, die geschlechtsunabhängig wirkt. Dieser Befund ist vor dem Hintergrund der Tatsache, dass die Mädchen im Physikunterricht hinter den Jungen immer mehr nachhinken, je länger der Schulbesuch währt [z. B. 50], interessant. Daraus ergibt sich eine lohnenswerte weitere Fragestellung für Folgestudien, die allerdings in einem Sample, das gleich viele Jungen und Mädchen beinhaltet, abgesichert werden sollte.

Ein Aspekt, der im Rahmen dieser Studie nicht zufriedenstellend abgedeckt werden konnte, betrifft die affektiven Komponenten des Lernens. Hier wäre es z. B. von Bedeutung gewesen, Aussagen über die Motivation der Tutoren und der Tutees zu erhalten und diese mit den Lernerfolgen verknüpfen zu können. Es wurde zwar versucht diese zu erfassen, aber es stellte sich heraus, dass bei der Adaption des Intrinsic Motivation Inventory unerwartete Schwierigkeiten auftraten [36]. Somit stand zum Zeitpunkt der Untersuchung kein Testinstrument in deutscher Sprache zur Verfügung, das in einem Sample von mehrheitlich Hauptschülern der Sekundarstufe 1 reliable Daten geliefert hat, die mit den Ergebnissen der Wissenstests schlüssig interpretierbar waren. Ergänzend wurde in einer Parallelstudie ein Teilbereich der Motivation der Schülerinnen und Schüler (die Herausforderung) in Interviews erhoben und mit den Ergebnissen der Wissenstests verknüpft [51]. Die Ergebnisse dieser Untersuchung waren allerdings sehr heterogen und lassen eher darauf schließen, dass die kognitiven

Fähigkeiten der Schülerinnen und Schüler moderierend wirken. Sobald ein psychometrisch valides Testinstrument für die Motivation zur Verfügung steht, ist auch dieser Aspekt von CAPT unbedingt zu berücksichtigen.

Dennoch scheinen sich gewisse Gelingensbedingungen für CAPT abzuzeichnen. In zwei weiteren Parallelstudien wurden die Tutor-Tutee-Interaktionen genauer beleuchtet. Diese Studien arbeiteten nach dem Mixed Methods Approach [46] und verbanden Fragebogendaten mit Interviews. Es zeigte sich, dass die Tutees ihre Tutoren nicht so sehr als Lehrpersonen, sondern auf freundschaftlicher Basis wahrnehmen [47], was die Befunde aus der Literatur unterstreicht [3]. Tutees scheinen darüber hinaus das Rollenverhalten ihrer Tutoren zu übernehmen [48]. Das führt dazu, dass das Kompetenzerleben der Tutees steigt, je besser die Tutoren auf Schülervorstellungen eingehen können.

Für die Implementierung von CAPT in den Regelunterricht sind zwei Dinge von Interesse: Erstens der Befund, dass CAPT in jedem Fall positive Effekte zeigt, die jene des Regelunterrichts übersteigen. Das gilt unabhängig von der individuellen kognitiven Leistungsfähigkeit der Tutoren und bedeutet somit, dass keine aufwändigen Diagnoseverfahren nötig sind, um geeignete Tutoren oder Tutorenklassen zu identifizieren. Dennoch erscheint eine sorgfältige Vorbereitung der Tutoren im Rahmen des Mentorings, vor allem im Hinblick auf Schülervorstellungen und damit verbundenen Diagnosekompetenzen, unbedingt empfehlenswert. Da die Tutoren keine ausgebildeten Lehrer/innen sind, ist der sozialen Komponente der Interaktion Rechnung zu tragen, indem ein eins-zu-eins-Setting gewählt wird und die Altersabstände zwischen den Gruppen nicht zu groß sind. Die Vorgangsweise, ganze Klassen miteinander arbeiten zu lassen, mag zwar wenig differenziert erscheinen, ist aber erfolgsträchtig und praktisch leicht anwendbar.

Zweitens profitieren nicht nur, sondern gerade die Tutoren von der CAPT-Intervention. Das lässt spezielle Belohnungsprogramme für diese Schülergruppe überflüssig erscheinen. Denn auch wenn in den Interventionen teilweise „alte“ Inhalte, die bereits gelernt wurden und „verstanden“ hätten sein sollen, Thema waren, konnte eine Verbesserung in der Testleistung beobachtet werden. Auch wenn die Mechanismen der Konzeptentwicklung nicht geklärt werden konnten, kann CAPT als eine den Regelunterricht ergänzende, relevante Methode uneingeschränkt empfohlen werden.

6. Literatur

- [1] A. GRAFENDORFER, in: C. Schreiner, U. Schwantner (Eds.), PISA 2006, Leykam, Graz, 2009.
- [2] B. TOFERER, in: U. Schwantner, B. Toferer, C. Schreiner (Eds.), PISA 2012, Leykam, Graz, 2013.
- [3] J.L. FOGARTY, M.C. WANG: An Investigation of the Cross-Age Peer Tutoring Process: Some Implications for Instructional Design and Motivation. In: The Elementary School Journal 82 (1982), S. 451-469.
- [4] P.A. COHEN, J.A. KULIK, C.L.C. KULIK: Educational Outcomes of Tutoring – A Meta-Analysis of Findings. In: Am. Educ. Res. J. 19 (1982), S. 237-248.
- [5] K.J. TOPPING, A. BRYCE: Cross-age Peer Tutoring of Reading and Thinking: Influence on thinking skills. In: Educational Psychology 24 (2004), S. 595-621.
- [6] J.A.C. HATTIE: Visible Learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge: London, New York, 2009.
- [7] C. HOWE, A. TOLMIE, K. GREER, M. MACKENZIE: Peer collaboration and conceptual growth in physics: Task influences on children's understanding of heating and cooling. In: Cognition and Instruction 13 (1995), S. 483-503.
- [8] A.T. LUMPE, J.R. STAVER: Peer Collaboration and Concept Development: Learning about Photosynthesis. In: Journal of Research in Science Teaching 32 (1995), S. 71-98.
- [9] R. DUIT, D.F. TREGUST: Conceptual Change: A powerful framework for improving science teaching and learning. In: International Journal of Science Education 25 (2003), S. 671-688.
- [10] J.H. WANDERSEE, J.J. MINTZES, J.D. NOVAK: Research on alternative conceptions in science. Macmillan: New York, 1994.
- [11] A. WIDODO, R. DUIT: Konstruktivistische Sichtweisen vom Lernen und Lehren und die Praxis des Physikunterrichts. In: Zeitschrift für Didaktik der Naturwissenschaften 10 (2004), S. 233-255.
- [12] K.J. TOPPING: The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. In: Higher Education 32 (1996), S. 321-345.
- [13] D.R. ROBINSON, J.W. SCHOFIELD, K.L. STEERS-WENTZELL: Peer and Cross-Age Tutoring in Math: Outcomes and Their Design Implications. In: Educational Psychology Review 17 (2005), S. 327-362.
- [14] K.J. TOPPING: Trends in Peer Learning. In: Educational Psychology 25 (2005), S. 631-645.
- [15] J. GAUSTAD: Peer and Cross-Age Tutoring. In: Digest 79 (1993).
- [16] B. ZINN: Ergebnisse einer Pilotuntersuchung zur Unterrichtsmethode "Lernen durch

- Lehren". In: Zeitschrift für Didaktik der Naturwissenschaften Jg. 15 (2009), S. 325-329.
- [17] J.P. MARTIN, Das Projekt "Lernen durch Lehren" – fachdidaktische Forschung im Spannungsfeld von Theorie und selbsterlebter Praxis, in: M. Liedtke (Ed.), *Gymnasium – Neue Formen des Unterrichts und der Erziehung*, Klinkhardt, Bad Heilbrunn, 1998, S. 151-166.
- [18] M. MÜLLER, R. BERGER, M. HÄNZE, Entwicklung von Trainings zur Verbesserung der Unterstützung im Cross-Age Tutoring, in: S. Bernhold (Ed.), *Jahrestagung der GDCP 2013*, IPN Kiel, München, 2014, S. 282-284.
- [19] C.A. ROHRBECK, M.D. GINSBURG-BLOCK, J.W. FANTUZZO, T.R. MILLER: Peer-Assisted Learning Interventions With Elementary School Students: A Meta-Analytic Review. In: *Journal of Educational Psychology* 95 (2003), S. 240-257.
- [20] R. DUIT, D.F. TREGUST, A. WIDODO: *Teaching Science for Conceptual Change: Theory and Practice*. Taylor & Francis: New York, 2008.
- [21] J. BORTZ, N. DÖRING: *Forschungsmethoden und Evaluation*. Springer: Berlin, 2003.
- [22] M. GOLLWITZER, R.S. JÄGER: *Evaluation kompakt*. Beltz: Weinheim, 2009.
- [23] S. PREDIGER, M. LINK, in: H. Bayrhuber, U. Harms, B. Muszynski, B. Ralle, M. Rothgangel, L.-H. Schön, H. Vollmer, H.-G. Weigand (Eds.), *Formate fachdidaktischer Forschung – Empirische Projekte – historische Analysen – theoretische Grundlagen*, Waxmann, Münster, 2012.
- [24] R. DUIT, C. RHÖNECK: Learning and understanding key concepts of electricity. In: *Connecting research in physics education with teacher education* (1998), S. 55-62.
- [25] D.M. SHIPSTONE: A study of children's understanding of electricity in simple DC circuits. In: *European Journal of Science Education* 6 (1984), S. 185-198.
- [26] H. WIESNER, in: R. Müller, R. Wodzinski, M. Hopf (Eds.), *Schülervorstellungen in der Physik*, Aulis Verlag Deubner, Köln, 2004, S. 53-65.
- [27] B. ANDERSSON, C. KÄRRQUIST: How Swedish pupils, aged 12-15 years, understand light and its properties. In: *International Journal of Science Education* 5 (1983), S. 387-402.
- [28] I. GALILI, A. HAZAN: Learners' knowledge in optics: interpretation, structure and analysis. In: *International Journal of Science Education* 22 (2000), S. 57-88.
- [29] H. WIESNER: Schülervorstellungen und Lernschwierigkeiten mit dem Spiegelbild. In: *Naturwissenschaften im Unterricht – Physik 3* (1992), S. 16-18.
- [30] H. WIESNER, in: R. Müller, R. Wodzinski, M. Hopf (Eds.), *Schülervorstellungen in der Physik*, Aulis Verlag Deubner, Köln, 2004, S. 71-79.
- [31] P.S. SHAFFER, L.C. MCDERMOTT: Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of instructional strategies. In: *American Journal of Physics* 60 (1992), S. 994-1013.
- [32] C. HAAGEN-SCHÜTZENHÖFER, M. HOPF, Development of a two-tier test-instrument for geometrical optics, in: C. Constantinou (Ed.), *E-Book Proceedings of the ESERA 2014 Conference: Science Education Research for Evidence-based Teaching and Coherent Learning*, Cyprus, 2014.
- [33] STATISTIK-AUSTRIA: Schulen, Schulbesuch. URL http://www.statistik.at/web_de/static/schulbesuch_der_5_schulstufe_200607_besuch_der_allgemeinbildenden_hoehere_035638.gif. (Stand: 01.06.2014)
- [34] E.L. DECI, R.M. RYAN: Self-Determination Theory: A Macrotheory of Human Motivation, Development, and Health. In: *Canadian Psychology* 49 (2008), S. 182-185.
- [35] E.L. DECI, R.M. RYAN: Intrinsic Motivation Inventory. URL <http://selfdeterminationtheory.org/edu/scales/category/5-intrinsic-motivation-inventory>. (Stand: 23.12.2013)
- [36] M. KORNER, H. URBAN-WOLDRON, M. HOPF, Entwicklung eines Messinstrumentes zur Motivation, in: S. Bernhold (Ed.), *GDCP-Jahrestagung – Konzepte fachdidaktischer Strukturierung für den Unterricht*, LIT Verlag, Oldenburg, 2012.
- [37] R. WHITE, R. GUNSTONE: *Probing Understanding*. RoutledgeFalmer: London, New York, 1992.
- [38] D. HAMMER: Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions. In: *The Journal of the Learning Sciences* 5 (1996), S. 97-127.
- [39] A. HELMKE, F.E. WEINERT (Hrsg.) *Bedingungsfaktoren schulischer Leistungen*. 71-176. Hogrefe: Göttingen, 1997.
- [40] H. URBAN-WOLDRON, M. HOPF: Testinstrument zum Verständnis in der Elektrizitätslehre. In: *Zeitschrift für Didaktik der Naturwissenschaften* Jg. 18 (2012).

- [41] C. HAAGEN-SCHÜTZENHÖFER, M. HOPF, Entwicklung eines Testinstruments zur geometrischen Optik, in: S. Bernholt (Ed.), Jahrestagung der GDCP, LIT, Oldenburg, 2011.
- [42] C. HAAGEN-SCHÜTZENHÖFER, J. ROTTENSTEINER, M. HOPF, Akzeptanzbefragung zu Optikunterrichtsmaterialien, in: S. Bernholt (Ed.), Jahrestagung der GDCP, IPN, Hannover, 2012.
- [43] H. WIESNER, in: R. Müller, R. Wodzinski, M. Hopf (Eds.), Schülervorstellungen in der Physik, Aulis Verlag Deubner, Köln, 2004.
- [44] A. WIDODO: Constructivist Oriented Lessons. The Learning Environments and the Teaching Sequences. Peter Lang GmbH: Frankfurt/Main, 2004.
- [45] C. TRINKL: Lernprozesse zum Thema Schatten und Lichtausbreitung. Universität Wien, AECC Physik, 2012.
- [46] J.W. CRESWELL, A.L. GARRETT: The "movement" of mixed methods research and the role of educators. In: South African Journal of Education 28 (2008), S. 321-333.
- [47] M. ZIEGLER: Lernprozesse und Rollenverständnis beim Cross Age Peer Tutoring. Universität Wien, AECC Physik, 2012.
- [48] B. HIMMER, M. KORNER, H. URBAN-WOLDRON, M. HOPF, Peer Tutoring: Rollenverständnis und Lernprozesse, in: S. Bernholt (Ed.), GDCP-Jahrestagung – Konzepte fachdidaktischer Strukturierung für den Unterricht, LIT Verlag, Oldenburg, 2011.
- [49] M. KORNER, Cross-Age Peer Tutoring in Physik. Evaluation einer Unterrichtsmethode. In: Studien zum Physik- und Chemielernen, Niedderer, H., Fischler, H., Sumfleth, E. (Hrsg), Logos: Berlin, 2015.
- [50] P. HÄUBLER, W. BÜNDER, R. DUIT, W. GRÄBER, J. MAYER: Naturwissenschaftsdidaktische Forschung: Perspektiven für die Unterrichtspraxis. Institut für die Pädagogik der Naturwissenschaften: Kiel, 1998. ISBN 3890881246
- [51] D. ABRAHAM, M. KORNER, H. URBAN-WOLDRON, M. HOPF, Motivation und Wissenszuwachs (Poster), GDCP-Jahrestagung – Konzepte fachdidaktischer Strukturierung für den Unterricht, Oldenburg, 2011.

7. Anhang

Einige Beispielmateriale zum Cross-Age Peer Tutoring mit Hinweisen zur Verwendung im Unterricht sind in der Zeitschrift Plus Lucis, Ausgabe 1-2 (2013), S. 11-15 des Vereins zur Förderung des Chemischen und Physikalischen Unterrichts mit Sitz in Wien publiziert. Die Zeitschrift ist neben einer Printversion online verfügbar unter: <http://pluslucis.univie.ac.at/PlusLucis/131/s11.pdf> (Stand: 21.12.2016).

Danksagung

Diese Studie stellt einen Teil des Sparkling Science Projektes „Cross-Age Peer Tutoring in Physik“ (CAPT) dar, das als Forschungs-Bildungs-kooperation zwischen Schulen und Universitäten vom ehemaligen Österreichischen Bundesministerium für Wissenschaft, Forschung und Wirtschaft (bmfw), derzeit Bundesministerium für Bildung und Frauen (bmbf) gefördert wurde.