

Danksagung

An erster Stelle möchte ich mich bei meinem Doktorvater Prof. Dr. Gunnar Friege herzlich bedanken. Ihre Unterstützung und das stets sehr gute Betreuungsverhältnis haben maßgeblich zum Gelingen dieser Dissertation beigetragen. Ich bin besonders dankbar für die vielen schönen Momente auf Konferenzen, bei denen ich die Gelegenheit hatte, meine Arbeit vorzustellen. Sie waren eine entscheidende Unterstützung, und nicht nur die vielen Gespräche über Fußball werden mir immer in guter Erinnerung bleiben.

Ein herzlicher Dank gilt auch Jun.-Prof. Dr. Peter Wulff und Prof. Dr. Eirini Ntoutsis für die Begutachtung meiner Arbeit sowie die anregenden Diskussionen darüber. Prof. Dr. Siegfried Raasch danke ich für die Übernahme des Prüfungsvorsitzes und die angenehme Disputation.

Mein Dank richtet sich ebenfalls an alle ehemaligen und aktuellen Kolleg:innen der Arbeitsgruppe Physikdidaktik der Leibniz Universität Hannover. Die vielen wertvollen fachdidaktischen Diskussionen und der Austausch über mein Projekt haben mir meine Zeit sehr angenehm gemacht. Mein Dank gilt auch den studentischen Mitarbeiter:innen, vor allem Hannes und Antonia, die mich in vielen Dingen stark entlastet haben.

Dirk danke ich für die unzähligen schönen und lustigen Momente, auch nach Feierabend, sei es beim Bier trinken oder Bier brauen, sowie für die inspirierenden Gespräche, die mir noch lange positiv in Erinnerung bleiben werden.

Ein besonderer Dank geht an meinen Bürokollegen André. Du hast mir in schwierigen Phasen nicht nur mit Deinem Optimismus, sondern auch mit deinem einzigartigen Humor geholfen. Die gemeinsamen Kaffeepausen waren wahre Oasen im Arbeitsalltag, die mir geholfen haben, körperlich und seelisch durchzuhalten. Deine Unterstützung, auch außerhalb der Arbeit, war unbezahlbar, sei es bei spontanen oder geplanten Feierabendbieren. Ich werde unsere Gespräche über das Leben, die Arbeit und alles dazwischen immer in guter Erinnerung behalten und

freue mich auf viele weitere lustige Momente mit dir.

Ich möchte mich auch bei den vielen Mitgliedern des Graduiertenkollegs Lern-MINT bedanken. Die gemeinsamen Diskussionen haben meine Arbeit bereichert und ich konnte viel Neues lernen. Mein besonderer Dank gilt Wolfgang, der immer ein offenes Ohr für mich hatte und mich mit den richtigen Hinweisen unterstützt hat.

Ein großer Dank geht an alle Schüler:innen und Physiklehrer:innen, die sich bereit erklärt haben, an meiner empirischen Untersuchung teilzunehmen. Ohne Eure Hilfe und Kooperation wäre die vorliegende Arbeit nicht möglich gewesen.

Danke auch an die Kolleg:innen des ALTA Institute der Cambridge University für den tollen Forschungsaufenthalt in England, der es mir ermöglichte, meine Methoden aus einem neuen Blickwinkel zu betrachten und meine Arbeit zu verbessern.

Von ganzem Herzen danke ich meiner Familie, insbesondere meinen Eltern sowie meinen Freunden, die mich in meinem Vorhaben bestärkt und unterstützt haben. Ohne euch wäre das alles nicht möglich gewesen.

Abschließend möchte ich mich bei meiner Freundin Ann-Christin bedanken. Du hast mich durch deine Unterstützung, dein Verständnis und deinen Zuspruch gerade in schwierigen Phasen immer wieder motiviert, mein Ziel konsequent zu verfolgen. Du warst und bist meine größte Stütze und ohne dich hätte ich diese Reise nicht geschafft. Ich bin dankbar, dich an meiner Seite zu haben.

Zusammenfassung

Formatives Assessment hat sich in vielen Studien als eine lernförderliche Maßnahme herausgestellt. Im Gegensatz zum summativen Assessment werden beim formativen Assessment diagnostische Informationen während des Lernprozesses erhoben. Auf der Grundlage dieser Informationen soll der weitere Lehr-Lern-Prozess adaptiert werden. Im besten Fall wird das Lernen auf einer individualisierten Ebene optimiert. Dazu sind eine Vielzahl von Daten nötig, die zudem aufbereitet und ausgewertet werden müssen. Im regulären Physikunterricht mit großen Klassen ist dies oft zu zeitintensiv, sodass formative Assessments nur eingeschränkt möglich sind. Dies zeigt sich beispielsweise beim Einsatz von Concept Maps. Durch Concept Maps können Zusammenhänge zwischen verschiedenen Begriffen der Physik dargestellt werden, wodurch wertvolle Informationen u. a. über Wissenslücken und Missverständnisse gewonnen werden. Der Auswertungsprozess kann jedoch sehr zeitaufwendig sein, insbesondere wenn die Concept Maps umfangreich sind und eine große Anzahl von Begriffen enthalten.

In der vorliegenden Arbeit erfolgt daher eine automatische Auswertung von Concept Maps durch Techniken des maschinellen Lernens, um Lehrkräfte und Lernenden im Lehr-Lernprozess zu unterstützen. In Studie 1 wurde eine Concept Map zum Thema Mechanik entwickelt und im Physikunterricht eingesetzt. Sie wurden durch zwei menschliche Rater bewertet und mehrere Machine-Learning-Modelle wurden mit den Daten trainiert und getestet. Die Ergebnisse zeigen eine sehr gute Übereinstimmung zwischen der menschlichen Bewertung und den verschiedenen Machine-Learning-Modellen. Basierend auf diesen Ergebnissen und dem besten geeigneten Machine-Learning-Modell wurde in Studie 2 ein Feedback-Tool entwickelt. Ziel war es, Einblicke in die Integration der automatischen Auswertung in den alltäglichen Physikunterricht zu gewinnen. Zu diesem Zweck wurde das Feedback-Tool an zwei Zeitpunkten in verschiedenen Klassen eingesetzt, um Lehrkräften und Lernenden automatische Rückmeldungen bereitzustellen. Die Auswertung von Fragebögen und Interviews ergab, dass die Vorteile der automatischen Auswertung von den Lehrkräften erkannt wird, sie jedoch das vollständige Potenzial nicht ausschöpfen. Die Ergebnisse zeigten, dass Lehrkräfte vor allem

Rückmeldungen auf Klassenebene und nicht auf individueller Lernenden-Ebene nutzen.

Insgesamt konnte in der vorliegenden Arbeit aufgezeigt werden, dass Systeme, die auf maschinellem Lernen basieren, im Bildungsbereich eine Vielzahl von Chancen bieten. Die Kombination von menschlicher Lehrkraft und computergestützter Hilfe ermöglicht eine individuelle Unterstützung von Lernenden, die in der Regel sonst nicht möglich wäre. Allerdings konnten auch Anregungen für künftige Forschungen erarbeitet werden, die eine stärkere Zusammenarbeit zwischen Informatik, Fachdidaktik und Lehrkräften erfordern.

Abstract

Formative assessment has been shown in many studies to be a method that promotes learning. In contrast to summative assessment, formative assessment is designed to collect diagnostic information during the learning process. Based on this information, the further teaching-learning process should be adapted. In the best case, learning is optimized on an individual level. This requires a large amount of data to be processed and evaluated. In regular physics classes with large numbers of students, this is often too time-consuming, so that formative assessments are only possible to a limited extent. An example of this is the use of concept maps. Concept maps can be used to illustrate relationships between different concepts in physics, providing valuable information about gaps in knowledge and misunderstandings that can be used as part of formative assessment. However, the assessment process can be time-consuming, especially when concept maps are complex and contain numerous concepts.

In this thesis, concept maps are automatically evaluated using machine learning techniques to support teachers and learners in the teaching-learning process. In Study 1, a concept map on the topic of mechanics was developed and used in physics classes. It was evaluated by two human raters and several machine learning models were trained and tested on the data. The results show a very good agreement between the human evaluation and the machine learning models. Based on these results and the best machine learning model, a feedback tool was developed in Study 2. The goal was to gain insight into the integration of automatic evaluation into everyday physics teaching. To this end, the feedback tool was used twice in different classes to provide automatic feedback to teachers and learners. The evaluation of questionnaires and interviews showed that the advantages of automatic evaluation are recognized by teachers, but that they have not exploited its full potential. The results indicated that teachers mainly used feedback at the class level rather than at the individual learner level.

Overall, this thesis suggests that machine learning-based systems offer a wide range of opportunities in education. The combination of human teachers and computer-based assistance allows for individualized support of learners that would otherwise not be possible. However, suggestions for further research could also

be developed, which require a stronger cooperation between computer science, science education and teachers.

1 Einleitung

„AI [Artificial Intelligence] has the potential to revolutionize the education sector by enhancing learning experiences, supporting teachers and offering more personalized learning opportunities for students. We must equip teachers with the knowledge and strategies they will need to use this new technology to improve and streamline everyday processes as well as classroom implementation“ (Bojorquez & Vega, 2023).

Obiges Zitat hebt die Chancen von künstlicher Intelligenz (KI) für den Bildungsbereich hervor: KI hat das Potenzial, den Bildungsbereich zu revolutionieren, indem Lernerfahrungen verbessert, Lehrkräfte unterstützt und individualisierte Lernmöglichkeiten für Lernende geschaffen werden können. Ebenso wird darauf hingewiesen, dass es wichtig ist, Lehrkräfte mit dem nötigen Wissen und den Strategien auszustatten, um diese neue Technologie effektiv einzusetzen. Dies umfasst sowohl die Optimierung der Lehr-Lern-Prozesse als auch die Integration von KI in den alltäglichen Unterricht.

Trotz des Potenzials muss der Einsatz von KI im Unterricht auch immer kritisch reflektiert werden. So thematisieren Studien, die KI-Anwendungen im Bildungsbereich untersucht haben, Probleme wie eine unreflektierte Akzeptanz von KI-basierten Ergebnissen oder die Benachteiligung bestimmter Lernenden-Gruppen durch KI-basierte Systeme (z. B. Krupp et al., 2023; Yao et al., 2020). Zusätzlich sind KI-basierte Systeme oftmals so komplex, dass die Kontrollierbarkeit und Erklärbarkeit der Ergebnisse eine große Herausforderung darstellen (Steinert et al., 2023).

Für den Einsatz im Unterricht muss daher sichergestellt werden, dass KI-basierte Systeme Informationen von Lernenden und Lehrenden zuverlässig, fachlich korrekt und ethisch unbedenklich auswerten können. Dies erfordert eine domänenübergreifende Zusammenarbeit von Expert:innen aus Fachdidaktik und Informatik, um die Chancen und Risiken von KI für die Optimierung des Lehr-Lern-Prozesses zu erforschen.

Im Bildungsbereich gibt es eine Vielzahl von Anwendungen für KI wie eine automatische Rechtschreibkontrolle oder die automatische Übersetzung von Texten. Der Vorteil von KI-basierten Systemen ist, dass sie große Datenmengen in sehr kurzer Zeit analysieren. Dadurch können Einblicke in den Lernprozess der Lernenden gewonnen und zum Beispiel im Rahmen eines formativen Assessments genutzt werden.

Formatives Assessment gilt seit langer Zeit als eine pädagogische Maßnahme mit einem hohen lernförderlichen Effekt (z. B. Hattie, 2009). Dabei besteht formatives Assessment aus einer Abfolge von drei Handlungsschritten (Souvignier & Hasselhorn, 2018):

1. Erfassung des individuellen Lernstandes
2. Rückmeldung für Lehrkräfte und Lernende
3. Optimierung des weiteren Lehr-Lern-Prozesses, aufgrund der gesammelten diagnostischen Informationen

Aus dieser Abfolge wird deutlich, dass die diagnostischen Informationen, die im ersten Schritt erhoben werden, die Grundlagen für die weiteren Handlungen im formativen Assessment schaffen (Schütze et al., 2018). Im Schulalltag werden diese Informationen oft durch Fragen im Unterrichtsgespräch erhoben, wobei eine Vielzahl von unterschiedlichen Methoden dazu existiert (Wiliam, 2010).

Eine dieser Methoden stellen Concept Maps dar. Concept Maps sind grafische Hilfsmittel zur Darstellung und Organisation von Inhalten und Wissen (Novak & Cañas, 2008; Ryssel, 2018). Lernende müssen bei der Erstellung einer Concept Map die Zusammenhänge zwischen verschiedenen Begriffen organisieren und visualisieren und so ihr Wissen über die gelernten Inhalte reflektieren. Daher können wichtige Informationen über Wissenslücken und Missverständnisse gesammelt werden, die im Rahmen eines formativen Assessments zur Optimierung des Lehr-Lern-Prozesses genutzt werden können (Novak & Cañas, 2006).

Aber trotz der lernförderlichen Effekte werden formative Assessments im Unterricht kaum genutzt (Black & Wiliam, 1998). Die Ursachen hierfür sind vielfältig. In der Praxis werden häufig der zu hohe Zeitaufwand für die menschliche Auswertung der gesammelten Informationen sowie die große Anzahl an Lernenden in einer Klasse, die eine individuelle Rückmeldung erschwert, als Gründe genannt (Bennett, 2011; Black & Wiliam, 1998; Hunt & Pellegrino, 2002).

Maps, die für ein formatives Assessment eingesetzt werden, zu erheben. Durch die automatische Auswertung kann der hohe Zeitaufwand, der sonst bei einer menschlichen Auswertung der Concept Maps entsteht, reduziert werden. Zudem kann die automatische Auswertung auch für eine große Anzahl an Lernenden ohne zusätzlichen Aufwand durchgeführt werden, was eine deutliche Unterstützung im Lehr-Lern-Prozess darstellen kann. Es wird überprüft, wie hoch die Übereinstimmung zwischen menschlicher und automatischer Auswertung ist, um Aussagen über die Leistungsfähigkeit der KI-basierten Auswertung treffen zu können. Zudem wird analysiert, wie zuverlässig die automatische Auswertung ist und ob die Auswertung unabhängig von Eigenschaften der Lernenden ist. So kann überprüft werden, ob eine gewisse Kontrollierbarkeit und Erklärbarkeit vorliegt, was für den Einsatz in Schulen wichtige Aspekte sind.

Die automatische Auswertung der Concept Map wird aber nicht nur auf einer theoretischen Ebene betrachtet. Ein weiteres Ziel der Arbeit ist die Untersuchung der KI-basierten Concept-Map-Auswertung im regulären Physikunterricht. Damit Aussagen zur Integration der automatischen Auswertung im alltäglichen Physikunterricht gemacht werden können, wird daher eine Studie an verschiedenen Gymnasien durchgeführt. In dieser Studie wird die automatische Auswertung an zwei unterschiedlichen Zeitpunkten des Physikunterrichts den Lehrkräften und Lernenden ein Feedback zur Verfügung stellen, welches auf der automatischen Auswertung von Concept Maps basiert. Anschließend werden über Interviews und Fragebögen Informationen zum Umgang mit der KI-basierten Auswertung erhoben und analysiert. Damit vielfältige Einblicke gesammelt werden können, wird der dritte Schritt des formativen Assessments *Optimierung des weiteren Lehr-Lern-Prozesses, aufgrund der gesammelten diagnostischen Informationen* nicht festgelegt.

Durch die vorliegende Arbeit können vielfältige Ergebnisse erzielt werden. Es werden nicht nur neue Erkenntnisse bezüglich der Leistung KI-basierter Auswertungen, die für ein formatives Assessment genutzt werden, erzielt, sondern wichtige Einblicke in die Integration von KI in den alltäglichen Physikunterricht ermittelt. Dies ermöglicht, neue Perspektiven einzunehmen und Strategien für einen effektiven Umgang mit KI-basierten Technologien zu ermitteln. Es können so wichtige Grundbausteine für einen zukünftigen Physikunterricht gelegt werden, denn KI wird in den nächsten Jahren das Lehren und Lernen prägen, oder wie es Seldon (2018) formulierte: „*AI the biggest thing to happen in education for 500 years*“.

Aufbau der vorliegenden Arbeit

Im Folgenden wird ein Überblick über die Inhalte der einzelnen Kapitel gegeben, welcher der Orientierung dienen soll:

In *Kapitel 2* wird zunächst das Thema formatives Assessment behandelt. Dabei wird versucht, den Begriff des formativen vom summativen Assessment abzugrenzen. Es wird außerdem auf die verschiedenen Schlüsselstrategien von Wiliam und Thompson (2008) eingegangen, wobei ein besonderes Augenmerk auf die lernförderliche Rückmeldung gelegt wird. Anschließend werden Befunde der Lehr-Lern-Forschung betrachtet und eingeordnet.

In *Kapitel 3* werden Concept Maps eingeführt, da sie, wie oben beschrieben, zur Erhebung des Lernstandes eingesetzt werden. Das Kapitel wird verschiedenen Concept-Map-Formate und Auswertungsmöglichkeiten diskutieren und einordnen.

In *Kapitel 4* wird das maschinelle Lernen (Machine Learning) beleuchtet, was ein Teilgebiet der künstlichen Intelligenz ist. In diesem Kapitel wird der Fokus vor allem auf das überwachte maschinelle Lernen gelegt, da die automatische Auswertung der Concept Maps durch ein Modell aus diesem Bereich des maschinellen Lernens erzeugt wird. Um das entwickelte Machine-Learning-Modell analysieren und vergleichen zu können, werden in Kapitel 4 aktuelle Forschungen aus dem Bildungsbereich untersucht.

In *Kapitel 5* werden die genaue Zielsetzung und das Erkenntnisinteresse der vorliegenden Arbeit beschrieben. Um die oben genannten Ziele der Arbeit zu erreichen, wird der empirische Teil in zwei Teilstudien unterteilt.

In *Kapitel 6* wird der Fokus auf der Entwicklung der Concept Map und die automatische Auswertung durch das Machine-Learning-Modell liegen. Zunächst wird daher die Konzipierung der Concept Map, die zur Erhebung des Lernstandes eingesetzt wird, beschrieben. Es wird außerdem die Entwicklung des Bewertungsschemas, welches die Grundlage für die lernförderliche Rückmeldung ist, dargelegt. Anschließend wird die erste Erhebungsphase beschrieben, bei der die entwickelte Concept Map in verschiedenen Klassen im Physikunterricht eingesetzt und von menschlichen Bewertern ausgewertet wird. Danach wird auf die Entwicklung des Machine-Learning-Modells eingegangen. Nach der Entwicklung werden

die Ergebnisse bezüglich der Übereinstimmung zwischen der menschlichen und automatischen Auswertung dargestellt und für den Anwendungsfall im regulären Physikunterricht kritisch diskutiert.

In *Kapitel 7* liegt der Fokus auf dem Einsatz des entwickelten Machine-Learning-Modells als Feedback-Tool. Daher wird zunächst auf der Grundlage der automatischen Auswertung ein lernförderliches Feedback konzipiert, welches in einer zweiten Erhebung in verschiedenen Klassen im Physikunterricht eingesetzt wird. Daher wird anschließend die Stichprobe und die Methodik dieser Erhebung erläutert. Abschließend werden auch in *Kapitel 7* die Ergebnisse aus den Interviews und Fragebögen dargestellt und diskutiert.

In *Kapitel 8* werden die Ergebnisse aus den beiden Teilstudien zusammenfassend diskutiert.

In *Kapitel 9* endet die Arbeit mit einem kurzen Fazit und einem Ausblick.