

Untersuchung und Operationalisierung von Performance-Messungen für adaptive Assistenzsysteme

MASTERARBEIT

KIT – KARLSRUHER INSTITUT FÜR TECHNOLOGIE
FRAUNHOFER IOSB – FRAUNHOFER-INSTITUT FÜR OPTRONIK,
SYSTEMTECHNIK UND BILDAUSWERTUNG

German Pustovojtovskij

15. Mai 2023

Verantwortlicher Betreuer:	Prof. Dr.-Ing. Jürgen Beyerer
Verantwortlicher Zweitbetreuer:	Prof. Dr.-Ing. Thomas Längle
Betreuender Mitarbeiter:	Dipl.-Inf. Alexander Streicher

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die Arbeit selbständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht habe und die Satzung des Karlsruher Instituts für Technologie zur Sicherung guter wissenschaftlicher Praxis in der gültigen Fassung beachtet habe.

Karlsruhe, den 15. Mai 2023

(German Pustovojtovskij)

In der vorliegenden Arbeit wird aus Gründen der besseren Lesbarkeit das generische Maskulinum verwendet. Weibliche und anderweitige Geschlechteridentitäten werden dabei ausdrücklich eingeschlossen, soweit es für die Aussage erforderlich ist.

Zusammenfassung

Im Zuge der COVID-19 Pandemie konnte die Wichtigkeit der Bereiche E-Learning und adaptive Assistenzsysteme besonders deutlich aufgezeigt werden. Unter E-Learning versteht man den Einsatz von elektronischen Medien, um Lerninhalte zu präsentieren und zu vermitteln. Adaptive Assistenzsysteme wiederum beziehen sich auf Technologien, welche sich an die individuellen Bedürfnisse sowie den Lernstil der Nutzer anpassen und ihnen personalisierte Unterstützung oder Feedback bieten.

Adaptivität kann die Wissensvermittlung für einen Nutzer beim E-Learning sowohl effektiver als auch effizienter gestalten. Damit adaptives Verhalten des Systems gewährleistet werden kann, muss der Lernfortschritt des Nutzers - auch Performance genannt - gemessen werden. Dabei wird die Performance durch die erfassten Interaktionen des Nutzers mit einem E-Learning System beschrieben.

Die vorliegende Masterarbeit untersucht, wie sich die Performance für adaptive Assistenzsysteme operationalisieren lässt. Hierfür werden die Modelle Knowledge Tracing (KT) und Performance Factor Analysis (PFA) aufbauend auf dem Stand der Forschung und Technik im Hinblick auf ihre Einsetzbarkeit für die Berechnung der Performance untersucht. Die Untersuchung zeigt, dass sich Graph-basierte Lernpfade sowie die Graph Performance Factor Analysis (GPFA) aufgrund ihrer gleichzeitigen Betrachtung aller Interaktionsdaten besonders für die Performance-Messungen eignen. Aufbauend auf dem entwickelten Konzept, wird eine Modellierung der GPFA für adaptive Assistenzsysteme erstellt. Diese wird im Anschluss für das Quizlernspiel Spot-X implementiert und der Lösungsansatz konzeptionell für die E-Learning-Plattformen ILLIAS sowie Moodle diskutiert.

Anhand einer Nutzerstudie wurde das Konzept und die implementierte Adaptivität evaluiert, Verbesserungsvorschläge identifiziert, und erste Ergebnisse ($n = 4$) weisen auf eine Verbesserung der Performance hin.

Abstract

In the course of the COVID-19 pandemic, the importance of the fields of e-learning and intelligent tutoring systems (ITSs) was clearly demonstrated. E-learning is the use of media to present and deliver learning content. Intelligent tutoring systems, on the other hand, are technologies that adapt to the user's individual needs and learning style and provide personalized support or feedback. The purpose of adaptivity is to make e-learning more effective and also more efficient in delivering knowledge to a user. To enable adaptive behavior, the user's learning progress - also called performance - must be measured. Performance is described by the recorded interactions of the user with an e-learning system. This thesis investigates how performance can be operationalized for ITSs. For this purpose, Knowledge Tracing (KT) and Performance Factor Analysis (PFA) are investigated with respect to their applicability for the performance calculation. The investigation shows that graph-based learning paths as well as Graph Performance Factor Analysis (GPFA) are suitable for performance measurement due to their simultaneous interpretation of all interaction data. Based on the theoretical concept, a GPFA model for ITSs is developed and implemented for the serious game Spot-X. This approach will also be discussed conceptually for other e-learning platforms such as ILIAS and Moodle. Based on a user study, the concept and implemented adaptivity were evaluated, suggestions for improvement were identified, and initial results ($n = 4$) indicate an improvement in performance.

Inhaltsverzeichnis

Zusammenfassung	v
Abstract	vii
1 Einführung	1
1.1 Problemstellung	2
1.2 Zielsetzung	2
1.3 Lösungsansatz	3
1.4 Aufbau und Gliederung der Arbeit	4
2 Grundlagen	5
2.1 Messung & Operationalisierung	5
2.2 Adaptive Assistenzsysteme und Serious Games	6
2.2.1 Adaptivität und adaptive Assistenzsysteme	7
2.2.2 Digital Games und Serious Games	11
2.3 Beispiel für Serious Games: SpotX	12
2.4 Beispiel für E-Learning Plattformen: ILIAS und Moodle	14
2.5 E-Learning A.I. (ELAI)	17
2.6 Beobachtungsdaten durch Experience API (xAPI)	20
2.7 Performance & Lernfortschrittsoptimierung	22
2.8 Progress-Performance mit Graph-basierten Ansätzen	24
2.8.1 Graphentheorie	24
2.8.2 Graph-basierte Modellierung von Lernpfaden	26
2.8.3 Graph Performance Service (GPS)	30
2.9 Result-Performance	30
2.9.1 Wissenskomponenten	31
2.9.2 Knowledge Tracing (KT)	32
2.9.3 Logistische Regressionsmodelle	34
2.9.4 Learning Factor Analysis (LFA)	35
2.9.5 Performance Factor Analysis	38

3	Stand der Forschung und Technik	43
3.1	Adaptive Assistenzsysteme und E-Learning	43
3.2	Lernfortschrittsmodellierung bei adaptiven Assistenzsystemen	45
4	Konzept für Performance-Messungen für adaptive Assistenzsysteme	55
4.1	Erweiterung des GPS um eine kombinierte Performance	56
4.2	Graph Performance Factor Analysis (GPFA)	57
4.3	Generisches Performance-Modell für Assistenzsysteme	60
5	Operationalisierung der Performance-Messungen	65
5.1	Festlegung der Hypothesen	65
5.2	Implementierung des Performance-Modells für Graphen	66
5.3	Implementierung der Result- und Progress-Performance	67
5.4	Erweiterung des adaptiven Assistenten für Spot-X	70
5.5	Vorstellung der Anwendungsszenarien	71
5.6	Korrektur des GPS nach der Nutzerstudie	74
5.7	Diskussion	77
6	Nutzerstudie: Evaluation der Performance-Messung	79
6.1	Experimentelles Verfahren	79
6.2	Ergebnisse	81
6.3	Diskussion	86
7	Fazit und Ausblick	89
	Literatur	93
	Abkürzungsverzeichnis	101
	Anhang	102

1 Einführung

Die Covid-19 Pandemie verdeutlichte die Wichtigkeit von digitalem Lernen für Ausbildungsinstitutionen, wie z.B. Schulen und Universitäten [Radha u. a. 2020]. Unabhängig hiervon stellt E-Learning einen wichtigen Bestandteil für Lehrmethoden in verschiedenen Fachbereichen dar, wie z.B. der medizinischen Ausbildung [Sharma u. a. 2017].

Eine dauerhafte Motivation des Lernenden sowie eine effektive und effiziente Hinführung zum Lernerfolg kann jedoch nur durch eine dynamische Anpassung der E-Learning-Systeme an das Nutzungs- bzw. Lernverhalten erreicht werden [Alshammari u. a. 2014]. Zum Beispiel können Effektivität und Effizienz sich auf die Geschwindigkeit und das Verständnis des Lerninhalts durch den Nutzer beziehen. Unter dynamischer Anpassung versteht man die beschriebene Unterstützung des adaptiven Assistenten. Im Kontext dieser Arbeit wird Adaptivität als die dynamische Anpassung eines Systems an die Bedürfnisse der Nutzer definiert. Beispiele für solche Anpassungen basierend auf Interaktionsmechanismen sind Inhalte oder Empfehlungen, welche von einem adaptiven Assistenzsystem personalisiert werden. Hierfür müssen adaptive Assistenzsysteme das Verhalten des Nutzers aufzeichnen. In dieser Arbeit erfolgt die Adaption an das Nutzungsverhalten mittels eines intelligenten Assistenten in Form einer dynamischen Anpassung des Schwierigkeitsgrads. Nutzungsverhalten wird hier verstanden als die messbaren Interaktionen des Nutzers mit einem E-Learning System. Um herauszufinden, zu welchem Zeitpunkt ein Assistent eingreifen muss, werden Techniken des maschinellen Lernens verwendet, um solche Interaktionsdaten zu analysieren und den bestmöglichen Einfluss auf die Personalisierung zu bewirken.

Damit ein passender Adaptivitätszeitpunkt oder -strategie gewählt werden kann, muss der Lernfortschritt des Nutzers durch eine Metrik messbar gemacht werden [Frutos-Pascual u. a. 2017]. In dieser Arbeit wird der Lernfortschritt auch als Performance des Nutzers bezeichnet. Mögliche Methoden, um die Performance zu messen, sind Graph-basierte Lernpfade [Delaunay 2022] sowie Regressionsmodelle. Mit Lernpfaden werden hier vordefinierte Nutzungsmuster innerhalb eines Kontexts in E-Learning Plattformen bezeichnet [Streicher, Schönbein u. a. 2021]. Modelle, die häufig zur Datenanalyse in E-Learning Systemen verwendet werden, sind Knowledge Tracing und Performance Factor Analysis [Pavlik Jr u. a. 2009]. Diese Arbeit konzentriert sich auf die Adaption durch die Wahl geeigneter Performance-Messungen. Dabei wird

untersucht, wie sich verschiedene Performance-Messungen aus den Benutzungsdaten auf den „Performance-Score“ auswirken. Benutzungsdaten sind im Kontext dieser Arbeit xAPI Daten, die aus den Interaktionen des Nutzers mit dem System entstehen. Zur Operationalisierung wird das Performance-Modell für das Quizlernspiel Spot-X eingesetzt. Weiterhin wird eine allgemeine Schnittstelle für die Kommunikation zwischen E-Learning-Systeme, wie ILIAS und Moodle, und dem Performance-Modell formal definiert.

1.1 Problemstellung

Die wissenschaftliche Fragestellung untersucht, wie die Adaptivität von Assistenzsystemen für die Lernfortschrittsoptimierung verbessert werden kann. Konkreter wird in dieser Thesis eine Variante des PFA Modells angewendet und mit den Graph-basierten Lernpfaden kombiniert, um die aktuelle Performance des Nutzers messen zu können. Pavlik Jr u. a. 2009; Gong, Joseph E Beck u. a. 2010 stellen Performance Factor Analysis (PFA) als eine Verbesserung der Performance-Messung vor. Hierbei handelt es sich um ein logistisches Regressionsmodell zur Berechnung der Performance eines bestimmten Nutzers. Dörner u. a. 2016 beschreiben, dass sich Performance aus dem Fortschritt und den Ergebnissen des Nutzers zusammensetzt. Dies wirft folgende Fragen auf, welche durch die vorliegende Arbeit beantwortet werden sollen:

- Wie kann die kombinierte Performance mittels PFA für Graphen und Graph-basierten Lernpfaden modelliert werden?
 - Wie sieht so eine Graph-basierte Modellierung für PFA aus?
 - Wie lassen sich die beiden Ansätze kombinieren?
 - Wie sieht so eine Modellierung für Assistenzsysteme allgemein aus?
- Wie sieht so eine Anwendung für E-Learning-Plattformen und Serious Games aus?
- Wie sieht so eine Anwendung für ein konkretes Anwendungsbeispiel aus?

Das Anwendungsbeispiel dieser Arbeit ist das am Fraunhofer IOSB entwickelte Quizlernspiel Spot-X. Das Spiel dient sowohl der Unterhaltung, als auch dem Training der Nutzer im Bereich der Allgemeinbildung.

1.2 Zielsetzung

Zunächst werden Modelle, wie z.B. Knowledge Tracing (KT) und PFA, anhand einer theoretischen Analyse verglichen, um das Beste als Basis für die Modellierung auszuwählen. Hierzu

werden die Modelle zunächst untersucht und deren Vor- und Nachteile erläutert. Die Wahl fiel auf das PFA Modell, da es sich - im Gegensatz zum KT - aufgrund seiner gleichzeitigen Betrachtung aller Interaktionsdaten besonders für die Performance-Messungen eignet. Anschließend wird ein Graph-basiertes Modell auf der Basis von PFA konzipiert werden, um sowohl die Gegebenheiten der Graphstruktur also auch die Vorteile der PFA nutzen zu können. Zum Beispiel bietet ein Graph-basierter Ansatz bessere Interpretierbarkeit der Vorhersagen im Vergleich zu den anderen Methoden [Nakagawa u. a. 2019]. Weiterhin soll dieses neu entstandene Modell mit dem Ansatz der Graph-basierten Lernpfade von Delaunay 2022 kombiniert werden. Es ist zu klären, inwieweit die kombinierte Performance für adaptive Assistenzsysteme allgemein und ein Fraunhofer IOSB Quizlernspiel Spot-X angewendet werden kann. Zuletzt soll die Wirksamkeit des entwickelten Performance-Modells im Serious Game Spot-X untersucht und in Form einer Nutzerstudie evaluiert werden.

1.3 Lösungsansatz

Zur Ermittlung der Performance wird die auf der Performance Factor Analysis [Pavlik Jr u. a. 2009] basierte Graph Performance Factor Analysis genutzt werden. Um aus dem PFA das GPFA Modell zu konstruieren, ist eine Extraktion sämtlicher benötigter Parameter für die Performance-Berechnung aus den Interaktionsdaten des Nutzers notwendig. Sodann wird auf Basis von Dörner u. a. 2016 die Performance zwischen Progress-Performance ($Perf^P$) und Result-Performance ($Perf^R$) differenziert, um eine kombinierte Messung aus Fortschritt und Ergebnis zu erhalten (s. Abschnitt 2.7). Die $Perf^P$ ist für den Fortschritt des Nutzers im Lernprozess zuständig und wird mithilfe von Lernpfade gemessen. Mit der $Perf^R$ werden die Ergebnisse des Nutzers durch den GPFA berechnet. Aus der Kombination der beiden Performance-Messungen lässt sich vorhersagen, wie der Nutzer die gestellte Frage zu einem bestimmten Themenbereich beantworten wird. Ein adaptiver Assistent kann aufgrund dieser Messung eine mögliche Anpassung auswählen, um das E-Learning System an die Bedürfnisse des Nutzers anzupassen (s. Abschnitt 2.2.1). Außerdem kann durch Beobachtungsdaten aus den Interaktionen des Nutzers mit dem System (s. Abschnitt 2.6) ein Adaptivitätszeitpunkt bestimmt werden, um das System zu aktivieren. Um die Mehrwerte der Performance-Messungen zu bewerten, werden in der Evaluation Korrelationen zwischen Selbsteinschätzung und den verschiedenen Performance-Messungen untersucht sowie die Performance-Messungen mit und ohne adaptive Assistenz verglichen. Natürlich gibt es auch weitere Faktoren, welche die Performance beeinflussen können, beispielsweise die Zeit, die zwischen den Lerneinheiten vergeht. Dabei kann ein Nutzer sein Wissen nach bestimmter Zeit wieder vergessen. Diese Masterarbeit beschränkt sich jedoch auf die zuvor genannten Faktoren, Fortschritt und Ergebnisse. Die hier nicht

berücksichtigten Faktoren, wie die Zeit zwischen Lerneinheiten, könnten Themen zukünftiger Arbeiten werden.

1.4 Aufbau und Gliederung der Arbeit

Im Folgenden wird der Aufbau der Arbeit erläutert. Die Arbeit beginnt mit der Darstellung der Motivation für die Behandlung der „Untersuchung und Operationalisierung von Performance-Messungen für adaptive Assistenzsysteme“. Daraufhin werden Problemstellung, Zielsetzung und Lösungsansatz dieser Masterarbeit vorgestellt. Anschließend folgt das Grundlagenkapitel, welches der Einführung und Definition einiger grundlegender Begriffe und Konzepte dient. Dazu gehören insbesondere adaptiven Assistenzsystemen, E-Learning-Plattformen, Serious Games, das Fraunhofer IOSB ELAI-Framework, Experience API (xAPI) sowie die Performance-Messungen Progress-Performance ($Perf^P$) und Result-Performance ($Perf^R$). Das nächste Kapitel widmet sich dem aktuellen Stand der Forschung und Technik zu den Themen adaptive Assistenzsysteme und Performance-Messungen. Kapitel 4 stellt die Modellierung einer kombinierten Performance vor und beschreibt ein Konzept für die Graph Performance Factor Analysis mit möglicher Parameterbelegungen. Darüber hinaus wird ein generisches Performance-Modell für Assistenzsysteme dargestellt. Sodann folgt ein Kapitel zur Operationalisierung der kombinierten Performance sowie der GPFA. Weiterhin werden aus dem vorangegangenen Kapitel Konzepte realisiert. Abschließend werden die benötigten Anpassungen am GPS zur Gewährleistung der Performance Berechnung diskutiert. In der Evaluation wird das experimentelle Verfahren der Nutzerstudie vorgestellt und anschließend deren Ergebnisse ausgewertet. Zuletzt schließt das Kapitel „Fazit und Ausblick“ mit einem Überblick über die Ergebnisse dieser Arbeit sowie einem Ausblick auf zukünftige Arbeiten aus Anlass der hier gewonnen Erkenntnisse.

2 Grundlagen

Das folgende Kapitel widmet sich verschiedenen, bereits entwickelten Konzepten und Technologien. Die hier vorgestellten Themen sind von großer Relevanz für den theoretischen sowie praktischen Teil der vorliegenden Arbeit. Kapitel 3 greift diese im Anschluss auf, um diese Arbeit vom Stand der Technik abzugrenzen und die Erweiterung des aktuellen Forschungsstandards zu verdeutlichen.

Zunächst wird das Prinzip der Adaptivität erläutert und in einen Zusammenhang mit dem Konzept von adaptiven Assistenzsystemen sowie Serious Games gestellt. Im Anschluss erfolgt eine Vorstellung der vom Fraunhofer IOSB entwickelten intelligenten Tutor Komponente ELAI sowie des Services, mit welchem die Nutzerdaten gesichert werden. Anschließend werden die Konzepte aus der Graphentheorie und der Graph Analytics mit den Themen dieser Arbeit in einen gemeinsamen Kontext gesetzt. Zuletzt folgt eine Auseinandersetzung mit möglichen Methoden der Lernfortschrittsoptimierung, wie z.B. Knowledge Tracing (KT), Learning Factor Analysis (LFA) oder Performance Factor Analysis (PFA) sowie den Modellen, welche darauf basieren. Teile dieser Arbeit sind der Bachelorarbeit von Pustovojtovskij 2021 entnommen, da diese die Basis der vorliegenden Thesis bildet.

2.1 Messung & Operationalisierung

Diese Arbeit nutzt die quantitative Methode der Operationalisierung zur Auswertung von Hypothesen in Bezug auf Performance-Messungen für adaptive Assistenzsysteme. Operationalisierung bedeutet, eine Variable mittels Indikatoren messbar zu machen [Wirth u. a. 2008]. Dabei sind Indikatoren manifeste Merkmale, welche beispielsweise durch eine Frage erhoben werden können [Heiss u. a. 2018]. Heiss u. a. 2018 beschreiben, dass theoretische Konzepte erst im Prozess der Operationalisierung und Messung präzise definiert werden, um sie anschließend mittels Indikatoren messbar zu machen. Hierbei wird der Prozess in vier Schritte eingeteilt (s. Abbildung 2.1). Zunächst müssen die theoretischen Konzepte genau definiert werden. Beispielsweise könnte die Performance eines Nutzers in einem SG als theoretisches Konzept betrachtet werden. Erst wenn ein Konzept vollständig spezifiziert ist, kann dieses im Umfang eingeschränkt werden. In Schritt zwei müssen Indikatoren für die Messung festge-

legt werden. Um das Konzept aus dem vorherigen Beispiel aufzugreifen, könnte die erreichte Punktzahl einen Indikator für die Messung darstellen. Anschließend findet die Datenakquise statt, mit welcher die Güte des Konstrukts bestimmt wird. Heiss u. a. 2018 definieren Güte als die Reliabilität und Validität des Konstrukts. Nur mit validen und reliablen Indikatoren können in Schritt vier diese in Skalen und Indizes gebündelt werden, um komplexere theoretische Konzepte empirisch erfassen zu können.

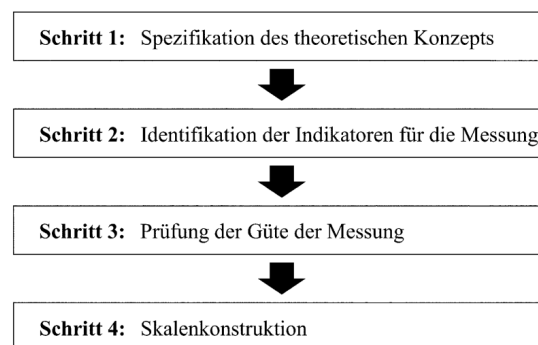


Abbildung 2.1: Prozess der Operationalisierung und Messung [Heiss u. a. 2018]

Für die Durchführung einer quantitativen Forschung muss zu Beginn eine Hypothese festlegen, welche quantitativ überprüft werden muss. Daraus wird anschließend eine Forschungsfrage abgeleitet, welche eine Annahme über einen möglichen Zusammenhang des Sachverhalts formuliert. Diese wird in der Statistik als Alternativhypothese H_1 bezeichnet. Um etwas statistisch zu testen, ist es notwendig, eine gegenteilige Hypothese, die Nullhypothese H_0 , zu formulieren [Albers u. a. 2009].

Um eine Hypothese zu überprüfen, kann die Falsifizierbarkeit durch statistische Signifikanz nachgewiesen werden. Für diesen Zweck wird ein Signifikanztest eingesetzt, welcher zwei komplementäre Hypothesen H_1 und H_0 überprüft [Albers u. a. 2009]. Das Signifikanzniveau gibt die Wahrscheinlichkeit an, mit welcher H_0 abgelehnt werden kann. Die Nullhypothese wird nicht abgelehnt, wenn erhobene Daten nicht im signifikanten Widerspruch zu H_0 stehen. Dies bedeutet nicht zwingend, dass H_0 zutrifft, sondern nur, dass mit den vorliegenden Daten H_0 nicht mit einer gewissen Wahrscheinlichkeit abgelehnt werden kann [Bamberg u. a. 2012; Albers u. a. 2009].

2.2 Adaptive Assistenzsysteme und Serious Games

Adaptive Assistenzsysteme - im Bereich des E-Learnings auch bekannt als Intelligent Tutoring Systems (ITSs) - haben in den letzten Jahren große Fortschritte erzielt [Gross u. a. 2015]. Das

übergeordnete Ziel dieser Systeme besteht darin, dem Nutzer eine personalisierte Unterstützung zu bieten, welche sich an dessen Bedürfnisse und Vorlieben anpasst. Im Vordergrund steht dabei die Verbesserung des Lernerfolgs und die Steigerung der Motivation [Gross u. a. 2015]. Serious Games (SG) stellen ein Teilgebiet im E-Learning dar, bei dem ein digitales Spiel sowohl der Unterhaltung als auch dem Erreichen eines konkreten Ziels dienen soll. Beim E-Learning besteht das konkrete Ziel darin, dem Nutzer etwas beizubringen [Dörner u. a. 2016]. Auch Serious Games profitieren stark von der stetigen Weiterentwicklung von adaptiven Assistenzsystemen. Durch die Nutzung von adaptiven Assistenzsystemen in SGs können Anpassungen am Spielgeschehen, wie beispielsweise ein Hinweis oder eine Veränderung der Schwierigkeitsstufe, getroffen werden, um den Lernfortschritt zu optimieren [S. Göbel u. a. 2016]. Im Folgenden wird der Lernfortschritt auch als Performance des Lernenden bezeichnet. Eine ausführliche Beschreibung der Performance ist in Unterkapitel 2.7 enthalten.

2.2.1 Adaptivität und adaptive Assistenzsysteme

Im folgenden Abschnitt erfolgt eine nähere Betrachtung von adaptiven Assistenzsystemen und Serious Games. Dabei soll die Motivation hinter diesen Themenfeldern erläutert und ein Zusammenhang zwischen beiden Bereichen herausgearbeitet werden. Zum Schluss werden einige konkrete Beispiele für SGs und weitere E-Learning Plattformen vorgestellt.

Anhand der voranschreitenden Entwicklung der E-Learning-Systeme wurde festgestellt, dass die „one-size-fits-all“ Lösung nicht die effektivste Lernmethode ist. Bei dem „one-size-fits-all“ Ansatz werden alle Lernenden auf dieselbe Art und Weise unterrichtet, was bedeutet, dass alle denselben Lernprozess durchlaufen. Dabei werden die unterschiedlichen Grundlagen und Fundamente der Lernenden nicht berücksichtigt [V. Shute u. a. 2018]. Damit die Effektivität des Lernprozesses maximiert werden kann, sollte das zugrunde liegende System von den unterschiedlichen Charakteristiken, welche der Lernende dem System liefert, profitieren können. Die gezeigten Inhalte sollten sich daher an jene individuellen Grundvoraussetzungen anpassen. Kurz gesagt, wäre eine fortlaufende Anpassung des Systems an den Lernfortschritt des jeweiligen Nutzers erforderlich. In diesem Fall spricht man von einem adaptiven System [V. Shute u. a. 2018]. In der vorliegenden Arbeit beschreibt Adaptivität das stetige Anpassen eines Systems an die individuellen Bedürfnisse des Nutzers [Streicher und Smeddinck 2016]. Folglich ist festzustellen, dass sich adaptive Systeme durch ihren dynamischen Charakter auszeichnen, welcher eine stetige nutzerbezogene Veränderung ermöglicht [Streicher und Smeddinck 2016].

Bereits in [Brusilovsky 1996] wird festgestellt, dass Adaptivität aus zwei Gründen bedeutend für den Bereich des E-Learning ist. Zunächst einmal können Lernsysteme von Anwendern, welche unterschiedliche Zielsetzungen, Lernmethoden, Präferenzen oder Wissen besitzen,

genutzt werden. Zweitens kann ein System dem Nutzer dienen, die Lerninhalte besser zu verstehen, indem es nutzerspezifische Pfade bietet [Hauger u. a. 2007]. Beispielsweise können bei einer Interaktion des Systems mit dem Nutzer, Inhalte und Empfehlungen angepasst werden. Mithilfe von Methoden der Datenanalyse und des maschinellen Lernens wird versucht, den positiven Einfluss durch die Interaktion zu maximieren [Streicher und Roller 2015]. Im Kontext von Serious Games oder E-Learning Plattformen ist dieser positive Einfluss die Erzielung eines Lerneffekts.

Zusammenfassend zeigt sich, dass adaptive Assistenzsysteme in der Lage sind, auf Grundlage von Nutzerdaten- und -präferenzen, angepasste Hilfestellungen zu ermöglichen, um Lernprozesse zu optimieren [Katsaris u. a. 2021]. Diese Art von System kann in verschiedenen Anwendungsbereichen eingesetzt werden, einschließlich des E-Learnings. Ein Beispiel für Intelligent Tutoring Systems stellt das E-Learning Artificial Intelligence (ELAI) dar, welches im Unterkapitel 2.5 näher beschrieben wird.

Um adaptives Verhalten eines Systems zu ermöglichen, müssen Auslöser für solche Modifikationen fest definiert sein. Diese Auslöser können Ereignisse oder verbindliche Schwellenwerte konkreter Metriken bei der Analyse sein [Sampayo-Vargas u. a. 2013]. Bei einem SG kann die Anzahl an Spielminuten oder eine Analyse des Spielverhaltens anhand verschiedener Metriken als derartiges Ereignis betrachtet werden [Streicher und Roller 2015]. Bei der Analyse des Spielverhaltens könnte man beispielsweise die Anzahl an korrekten bzw. falschen Antworten verwenden. Mithilfe dieser Auslöser ist es möglich, einen Zeitpunkt zu bestimmen, um den Spielverlauf an die Bedürfnisse des Nutzers anzupassen. Ein solches anpassungsfähiges Spiel nennt man adaptives Lernspiel [Streicher und Roller 2015].

Flow

Weshalb Adaptivität eine solch wichtige Rolle für die Lernfortschrittsoptimierung spielt, kann anhand des Flow-Konzepts begründet werden. Der Begriff Flow wurde von Mihály Csíkszentmihályi in [Mirvis 1991] erstmals im psychologischen Sinne geprägt. Mihály Csíkszentmihályi beschreibt Flow als einen psychologischen Zustand, in dem die gesamte Aufmerksamkeit eines Menschen auf eine Aufgabe gerichtet wird und man sich vollständig in dieser Aktivität verliert [Csíkszentmihályi 2009]. Jener Zustand tritt auf, wenn eine Person eine Tätigkeit ausübt, welche auf deren Fähigkeiten abgestimmt ist und folglich eine angemessene Herausforderung darstellt. Um in diesem Zustand zu bleiben, muss der Mensch sich im sogenannten Flow-Kanal befinden. Diese Zone lässt sich zwischen Zuständen der Überforderung sowie der Langeweile verorten [Csíkszentmihályi 2009]. Diese Charakteristik des Flow-Kanals ist der Abbildung 2.2 zu entnehmen. Unterschiedliche Grund- und Vorkenntnisse erschweren es, den jeweiligen

Flow Kanal eines Menschen zu bestimmen. Anhand von Informationen über den Wissensstand des Menschen ist es möglich, Anpassungen an der gestellten Aufgabe vorzunehmen, um jeden Menschen individuell in dessen Flowzone zu führen. Csikszentmihályi beschreibt den Flow-Zustand als harmonisches Zusammenspiel von Körper und Geist. Der Mensch ist während des Flow-Zustands dazu in der Lage, Höchstleistung zu erbringen und sein gesamtes Potenzial zu entfalten [Csikszentmihályi 2009].

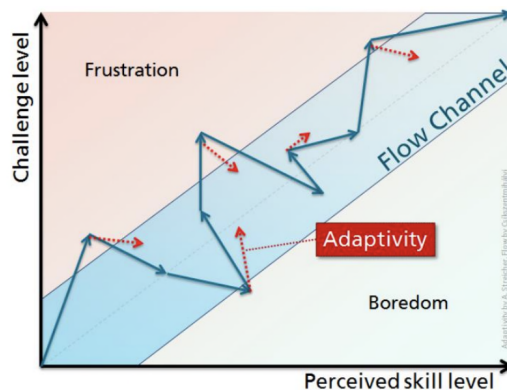


Abbildung 2.2: Flow-Kanal, basierend auf Csikszentmihályi 2009. Mithilfe von Adaptivität (gestrichelte Pfeile) kann der Nutzer innerhalb des Flow-Kanals zwischen Überforderung und Unterforderung bleiben.

Um den Lernfortschritt zu optimieren, ist es erforderlich, den Nutzer in dieser Flowzone zu halten. Das heißt, dass sich das System anpassen muss, um den Nutzer nicht zu über- oder unterfordern. Diese Anpassungen kann mittels Adaptivität erreicht werden [Streicher und Smeddinck 2016]. Es stellt sich hierbei die Frage, woran eine Über- bzw. Unterforderung erkannt werden kann. Ein Indikator für die Überforderung eines Nutzers könnte beispielsweise dessen falsche Antwort im Rahmen eines SG sein. Eine Reihe von richtigen Antworten weist möglicherweise auf eine Unterforderung hin.

Adaptivitätszyklus

Im folgenden Abschnitt wird eine Möglichkeit erläutert, wie Adaptivität in ein Assistenzsystem integriert werden kann, um den Nutzer mithilfe der Adaptivität im Flow-Kanal zu halten und somit dessen Lernfortschritt zu optimieren. Es wurde verdeutlicht, dass Adaptivität einen wichtigen Faktor im Rahmen der Lernfortschrittsoptimierung darstellt. Daraus ergeben sich zentrale Fragen und Probleme, die es zu klären gibt. Zentral ist dabei, wie und wann Anpassungen vorgenommen werden sollen, also wann ein adaptives System aktiv wird. Um einen passenden Zeitpunkt bzw. eine Strategie zu wählen, muss der Lernfortschritt des Lernenden

messbar gemacht werden [Frutos-Pascual u. a. 2017]. Der folgende Abschnitt erläutert eine Methode, mithilfe derer sich Adaptivität in ein System integrieren lässt.

V. J. Shute u. a. 2012 stellen einen Adaptivitätszyklus bestehend aus vier Phasen vor. Abbildung 2.3 zeigt die von Streicher und Smeddinck 2016 überarbeitete Darstellung des Adaptivitätszyklus. Der vierphasige Zyklus beschreibt ein iteratives Verfahren, um den Lernvorgang an die Bedürfnisse des Nutzers anzupassen. Dabei besteht das Modell aus vier verschiedenen Phasen, die im Folgenden aufgelistet werden:

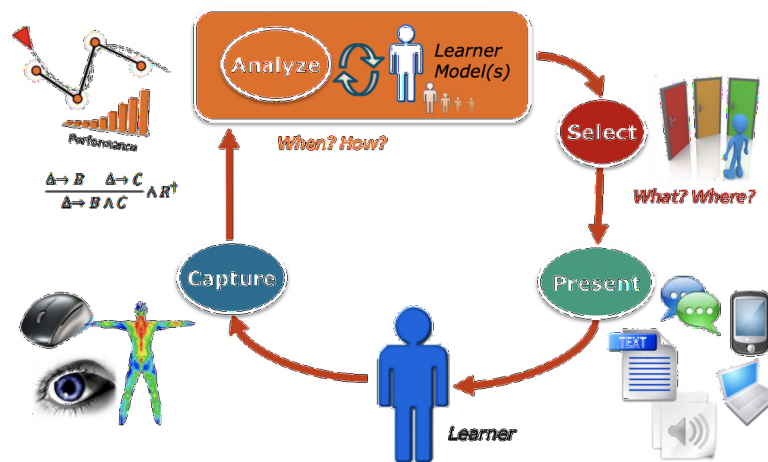


Abbildung 2.3: Der vierphasige Adaptivitätszyklus [Streicher und Smeddinck 2016] (basierend auf [V. J. Shute u. a. 2012])

- Erfassung („Capture“): In dieser Phase werden während des Lernprozesses Informationen über den Nutzer gesammelt. Die Datenakquise kann von System zu Systemen variieren. Zum Beispiel könnten für SGs die Daten Bewegungen der Maus sowie die erzielten Ergebnisse beinhalten. Falls das System in einem Umfeld eingesetzt wird, in dem weitere Sensoren - beispielsweise Eyetracking - verwendet werden, können diese auch für die Datenbeschaffung von Nutzen sein [Streicher und Smeddinck 2016].
- Analyse („Analyze“): Hier werden die gesammelten Daten mittels intelligenter Analysetechnik ausgewertet und anhand der Ergebnisse Entscheidungen über den Lernfortschritt getroffen. In dieser Phase können unterschiedliche Techniken des maschinellen Lernens oder Data-Minings eingesetzt werden, um auf den aktuellen Lernfortschritt des Nutzers zu schließen [Streicher und Smeddinck 2016]. Derartige Techniken werden in Abschnitt 2.7 im Detail vorgestellt.

- Selektion („Select“): Nachdem die Datenanalyse abgeschlossen ist, werden die gewonnenen Ergebnisse gespeichert. In diesem Schritt werden der Zeitpunkt der Anpassung sowie die Methode, mit welcher die Adaptivität realisiert wird, gewählt. Anhand der aus der vorherigen Phase gewonnenen Ergebnisse wird der Inhalt gewählt, welcher für den Nutzer am geeignetsten scheint [Streicher und Smeddinck 2016]. Der Inhalt kann in unterschiedlichen Variationen auftreten. Bei einer E-Learning-Plattform kann dem Assistenzsystem beispielsweise ein Fragenpool zur Verfügung stehen, bei welchem sich der Inhalt zwischen schweren und leichten Fragen unterscheidet. Bei Serious Games könnten die gestellten Aufgaben durch ein zusätzliches Hindernis oder einen Hinweis angepasst werden.
- Darstellung („Present“): Die ausgewählten Anpassungen werden dem Nutzer im Kontext des Lernprozesses vorgestellt.

Im Anschluss an die Darstellungsphase beginnt der gesamte Zyklus wieder von vorne. Der Grund hierfür liegt im dynamischen Charakter eines adaptiven Systems, welches sich typischerweise mit der Zeit immer weiterentwickelt [Gunderson u. a. 2002; V. J. Shute u. a. 2012]. Im nächsten Zyklus werden stetig Informationen erfasst, da der Nutzer sein Verhalten an das Dargestellte anpasst. Adaptivität ist als Konzept interessant, da das Ziel verfolgt wird, für jeden Lernenden den maximalen Fortschritt zu erzielen und Adaptivität dafür am besten geeignet scheint. Dies wird anhand von Lernfortschritts-Messungen - in der vorliegenden Arbeit auch Performance-Messungen genannt - realisiert, welche in Kapitel 2.7 detailliert aufgeführt werden. Um Adaptivität zu integrieren, muss jedoch zunächst bestimmt werden, ob der Nutzer eine Anpassung benötigt. Der Kernteil der Arbeit befasst sich mit der Analysephase des zuvor dargestellten Adaptivitätszyklus. Dort wird anhand der erfassten Daten mittels einer Variante der Performance Factor Analysis (PFA) bestimmt, ob eine Anpassung im Spielverhalten notwendig ist.

2.2.2 Digital Games und Serious Games

Im Rahmen der vorliegenden Arbeit werden SG und E-Learning-Plattformen behandelt, da sie einen konkreten Anwendungsbereich für adaptive Anpassungen in einem System bilden. Aus diesem Grund geht der nächste Abschnitt näher auf Serious Games und deren Charakteristik ein.

Wie bereits in der Arbeit von Pustovojtovskij 2021 beschrieben, mögen es Menschen, unterhalten zu werden und Spiele zu spielen [Dörner u. a. 2016, S.2]. Spiele wie Backgammon und Go werden heute immer noch gespielt, obwohl sie bereits 5.000 Jahre alt sind. Nachdem der

Mensch begonnen hat, Gegenstände wie Bälle, Karten und Würfel zum Spielen zu verwenden, ist es nicht überraschend, dass der Mensch auch den Computer zu spielerischen Zwecken verwendet. Daraus sind Digital Games entstanden, welche sich als Spiele beschreiben lassen, die eine bestimmte Art von Computer nutzen - z.B. Smartphone, Laptop, Konsole. Diese sind sehr erfolgreich und besitzen ein breites Spektrum an Nutzerdemografien [Dörner u. a. 2016].

Aufgrund des Erfolges werden sie auch für andere Zwecke, wie z.B. zum Lernen oder für physische Aktivitäten, eingesetzt. Auch im Bereich des Fernsehens, welches ebenfalls der Unterhaltung dient, wurden Fortschritte im Bereich des Lernens festgestellt. Ein Beispiel für eine Kindersendung, die sowohl der Unterhaltung als auch dem Lernfortschritt dienen soll, ist die Sesamstraße [Dörner u. a. 2016, S.2]. Dasselbe Ziel wird auch in einem Teilbereich der Digital Games angestrebt. Somit wird auf die Definition von Serious Games geschlossen. Ein SG ist ein Digital Game, das neben der Unterhaltung ein weiteres zusätzliches Ziel verfolgt (z.B. Gesundheit oder Lerneffekt) [Dörner u. a. 2016, S.2]. Diese zusätzlichen Ziele nennt man charakterisierende Ziele. Serious Games sind dazu in der Lage, den Lernprozess schwieriger Inhalte zu einer unterhaltsamen, fesselnden Erfahrung zu machen und verwandeln somit anspruchsvolle Lernprozesse in Spielerfahrung für den Nutzer. [Ritterfeld u. a. 2009]. Das Anwendungsgebiet von SGs ist breit gefächert und wird häufig für akademische Zwecke genutzt, um bspw. beim Auswendiglernen oder Wiedergeben von Fakten Unterstützung zu bieten [Ritterfeld u. a. 2009]. Abbildung 2.4 ist ein Schaubild zu entnehmen, welches die bereits definierten Begriffe vereint und welches kenntlich macht, dass SG die Schnittstelle zwischen dem Lernen, der Simulation sowie dem Spielen bilden.

Das charakterisierende Ziel von SG im Allgemeinen stellt das Lernen dar [Ning u. a. 2021]. In der vorliegenden Arbeit besteht das Ziel darin, einen Lerneffekt zu erzielen. Abbildung 2.5 zeigt eine Bildschirmaufnahme eines SGs, in dem das charakterisierende Ziel die Vermittlung eines Lernfortschritts in einem konkreten Themenbereich ist. Um einen positiven Lerneffekt zu erzielen, sind Entwurf und Design des Spielverlaufs entscheidend [Streicher und Smeddinck 2016]. Mithilfe des bereits erläuterten Adaptivitätszyklus lässt sich Adaptivität in ein SG integrieren. Die vorliegende Arbeit zielt darauf ab, das Konzept der Adaptivität in das Quizlernspiel Spot-X zu integrieren, welches im folgenden Kapitel zunächst näher beleuchtet wird.

2.3 Beispiel für Serious Games: SpotX

Das Quizlernspiel Spot-X ist ein adaptives SG und wird in dieser Arbeit als Anwendungsbeispiel zur Operationalisierung von Performance-Messungen betrachtet. Spot-X profitiert von der Einfachheit der Nutzeroberfläche und ermöglicht es, sich spielerisch Wissen anzueignen. Ein Spieldurchlauf besteht aus insgesamt zehn Aufgaben. Nach jedem Spieldurchlauf

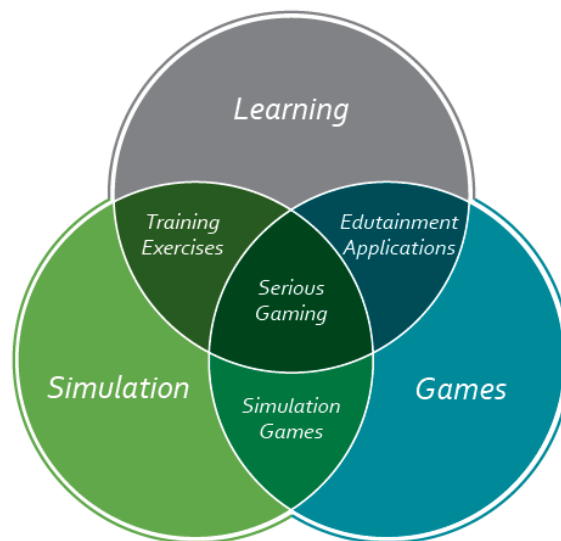


Abbildung 2.4: Ein Schaubild zum Thema SG (Bildquelle: *Serious Games* 2016)

wird dem Nutzer eine Übersicht bestehend aus allen abgeschlossenen Aufgaben sowie eine Gesamtpunktzahl angezeigt. In einem Aufgabendurchlauf besteht das Ziel des Spielers, die richtige Antwort innerhalb eines vordefinierten Zeitfensters auszuwählen. Abbildung 2.6 zeigt eine Multiple-Choice-Aufgabe, in der die abgebildete Stadt bestimmt werden soll.

Spot-X besitzt sowohl unterschiedliche Aufgabentypen, als auch verschiedene Aufgabekategorien. Zum Beispiel bietet Spot-X neben Multiple-Choice-Fragen auch Aufgaben mit Freitextfeldern und Wimmelbildern. Zu den Aufgabekategorien zählen unter anderem die Themen Städte (s. Abbildung 2.6) sowie Wildkatzen (s. Abbildung 2.5).

Außerdem bietet Spot-X die sogenannte Schwierigkeitskategorie (SK). Die SK dient dazu, den Schwierigkeitsgrad der verschiedenen Aufgaben zu ermitteln. Jede Aufgabe besitzt entweder die SK „leicht“, „mittel“ oder „schwer“. In Abbildung 2.7 sind zwei Aufgaben a und b, beide aus der Aufgabekategorie Mathematik mit den Schwierigkeitskategorien $SK(a) = leicht$ und $SK(b) = schwer$.

Neben der Schwierigkeitskategorie existiert zusätzlich die sogenannte dynamische Schwierigkeitsanpassung. Diese wird durch den Schwierigkeitsgrad $diff \in [0,1]$ reguliert. Dabei ist 0 die leichteste und 1 die schwerste dynamische Schwierigkeitsanpassung. Diese beeinflusst jedoch nicht die Aufgabenstellung, sondern lediglich die Anzahl der Antwortmöglichkeiten oder die Zeitlimitierung. Abbildung 2.8 zeigt dieselbe Aufgabe - zum einen mit dem Schwierigkeitsgrad

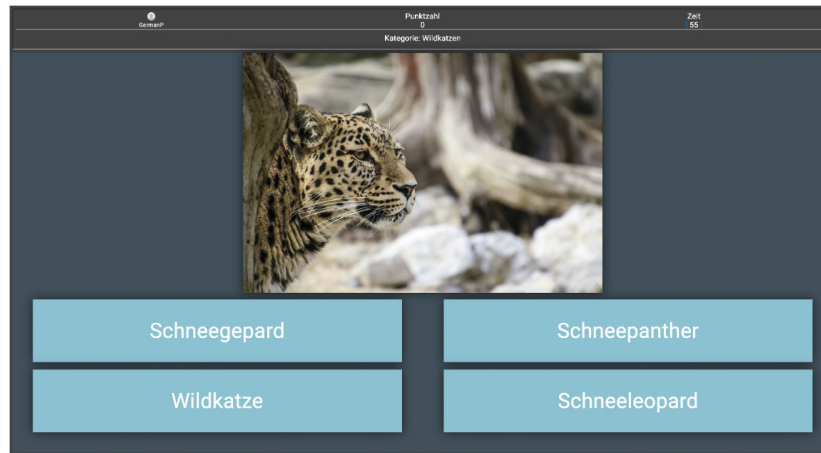


Abbildung 2.5: Eine Beispielfrage aus dem Serious Game Spot-X

$diff(a) = 0.5$, zum anderen mit $diff(b) = 1$. Durch die dynamische Schwierigkeitsanpassung werden dem Nutzer bei dieser Aufgabe mehr Antwortmöglichkeiten angezeigt. In dieser Arbeit wird sich auf die Schwierigkeitskategorien SK beschränkt.

2.4 Beispiel für E-Learning Plattformen: ILIAS und Moodle

Wie in [K. Göbel u. a. 2023] beschrieben, ist seit Ausbruch der Corona-Pandemie im Frühjahr 2020 und den damit verbundenen Schließungen der Universitäten, die Digitalisierung der Hochschullehre nicht mehr aufzuhalten. Die Anzahl an E-Learning Plattformen, auch Learning Management Systems (LMSs) genannt, ist in den letzten Jahren konstant gestiegen, da zahlreiche Universitäten sowie Unternehmen diese Plattformen seither für den gemeinsamen und persönlichen Gebrauch nutzen [Hauger u. a. 2007].

Nach [Piotrowski 2010] stellt eine E-Learning-Plattform ein System dar, welches eine integrierte Unterstützung für folgende sechs Aktivitäten bietet: Erstellung, Organisation, Bereitstellung, Kommunikation, Zusammenarbeit und Bewertung. Zu den Funktionen von E-Learning-Plattformen gehören in der Regel die Bereitstellung von Lerninhalten, Tests und Kommunikationswerkzeuge. Darüber hinaus bieten die Plattformen eine Möglichkeit zur Zusammenarbeit der Nutzer sowie Kursverwaltungs- und Bewertungsfunktionen für Lehrkräfte. E-Learning-Plattformen können auch Verwaltungsfunktionen oder Schnittstellen zu Verwaltungssystemen (engl. „campus management systems“) für die Zulassung und Einschreibung von Nutzern enthalten [Piotrowski 2010]. Aufgrund der Vielzahl der LMSs konzentriert sich die Arbeit auf die E-Learning-Plattformen ILIAS und Moodle, die stellvertretend vorgestellt

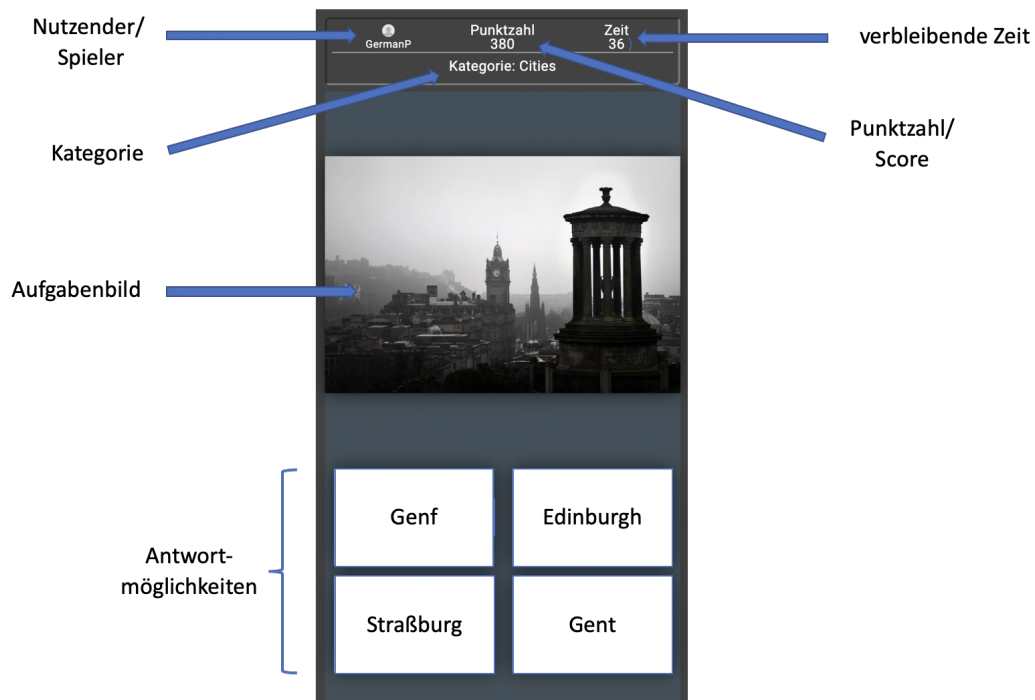
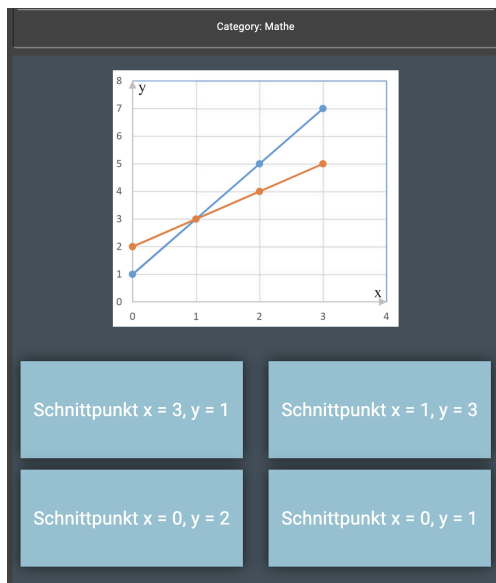


Abbildung 2.6: Eigenschaften und Attribute einer Multiple-Choice-Aufgabe in Spot-X

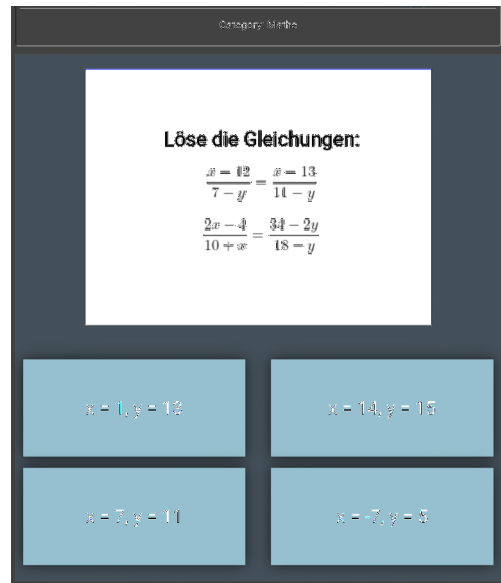
werden. Diese beiden Systeme werden anschließend in Kapitel 4.3 und 4.3 konzeptionell für die Performance-Messungen eines Nutzers durch einen adaptiven Assistenten verwendet.

ILIAS

ILIAS [ILIAS 2023] ist ein Open Source LMS für das Entwickeln und Realisieren von webbasiertem E-Learning und ist eine Abkürzung für Integriertes Lern-, Informations- und Arbeitskooperationssystem. Der erste Prototyp für ILIAS wurde 1998 an der Universität Köln entwickelt [Hauger u. a. 2007]. Als E-Learning-Plattform bietet ILIAS sowohl Kurs-Management als auch Lernmodule, Tests, Prüfungen und vieles mehr an. Ein Beispiel für eine ILIAS-Übersicht mit einem Kurs und einer Gruppe ist in Abbildung 2.9 zu sehen. Ein weiterer Vorteil, den diese Lernplattform bietet, ist, dass ILIAS standardkonform ist. Es unterstützt SCORM 1.2 sowie SCORM 2004-Lernmodule (Abkürzung für „Sharable Content Object Reference Model“), welche einen internationalen Standard für die Erstellung von E-Learning-Modulen besitzen [ILIAS 2023].

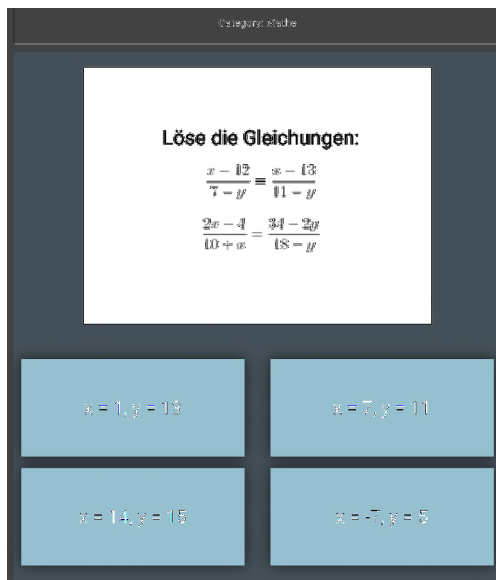


(a)

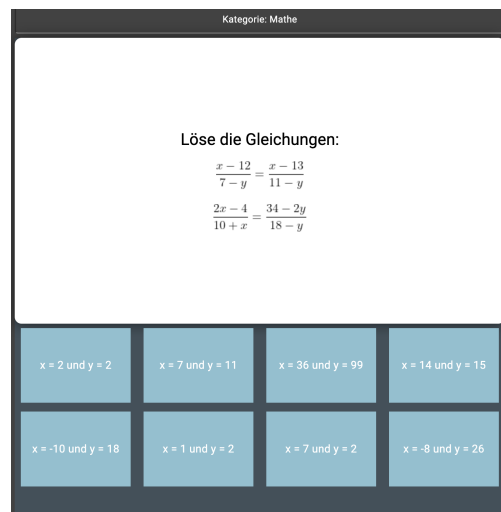


(b)

Abbildung 2.7: Zwei Aufgaben aus dem Quizlernspiel Spot-X mit den Schwierigkeitskategorien leicht und schwer



(a)



(b)

Abbildung 2.8: Eine Aufgabe aus dem Quizlernspiel Spot-X mit unterschiedlichen Schwierigkeitsanpassungen $diff(a) = 0.5$ und $diff(b) = 1$

The screenshot shows the ILIAS eLearning dashboard for 'abc Schule eLearning'. The interface includes a top navigation bar with the ILIAS logo and a search icon. A left sidebar contains icons for Dashboard, Magazin, Persönlicher Arbeitsraum, Lernerfolge, Kommunikation, and Administration. The main content area is divided into several sections: 'Favoriten' (Favorites) with instructions on how to add favorites; 'Meine Kurse und Gruppen' (My Courses and Groups) showing a list of 'Gruppe Beispiel' and 'Kurs Beispiel' with an 'Anmeldung' (Registration) button; a 'Kalender' (Calendar) widget for February 2021; and a 'Mail' section showing 'Keine Einträge' (No entries). The footer indicates the system is powered by ILIAS (v6.7 2021-02-05) and provides links for Impressum and Info Barrierefreiheit.

Abbildung 2.9: Open Source Learning Management System ILIAS Übersicht mit einer Auswahl an Gruppen und Kursen (Bildquelle: ILIAS 2022)

Moodle

Moodle [Moodle LMS 2023] wurde 2003 gegründet [Hauger u. a. 2007] und stellt ebenfalls ein beliebtes Open Source LMS dar, das an Universitäten, Schulen und medizinischen Einrichtungen verwendet wird [Athaya u. a. 2021]. Moodle ermöglicht es Lehrenden, Aufgaben sowie Unterrichtsmaterialien und Dokumente mit den Studierenden der Plattform zu teilen. Dabei ist Moodle leicht zu benutzen und bietet eine hohe Qualität an Lernmöglichkeiten bietet [Al-Ajlan u. a. 2008]. Zusätzlich stellt Moodle dem Nutzer eine Vielzahl an Kommunikationsmöglichkeiten zur Verfügung, wie z.B. Chats, Foren und Nachrichtensysteme [Oproiu 2015]. Diese können für Gruppenarbeiten oder bei Rückfragen verwendet werden. Ähnlich wie ILIAS hat Moodle eine ausführliche Dokumentation und unterstützt die SCORM Standards [Al-Ajlan u. a. 2008].

2.5 E-Learning A.I. (ELAI)

Wie bereits in Abschnitt 2.2.1 behandelt, analysiert ein adaptiver Assistent die Interaktionen des Nutzers mit der E-Learning-Plattform. Diese Arbeit baut auf dem am Fraunhofer IOSB von

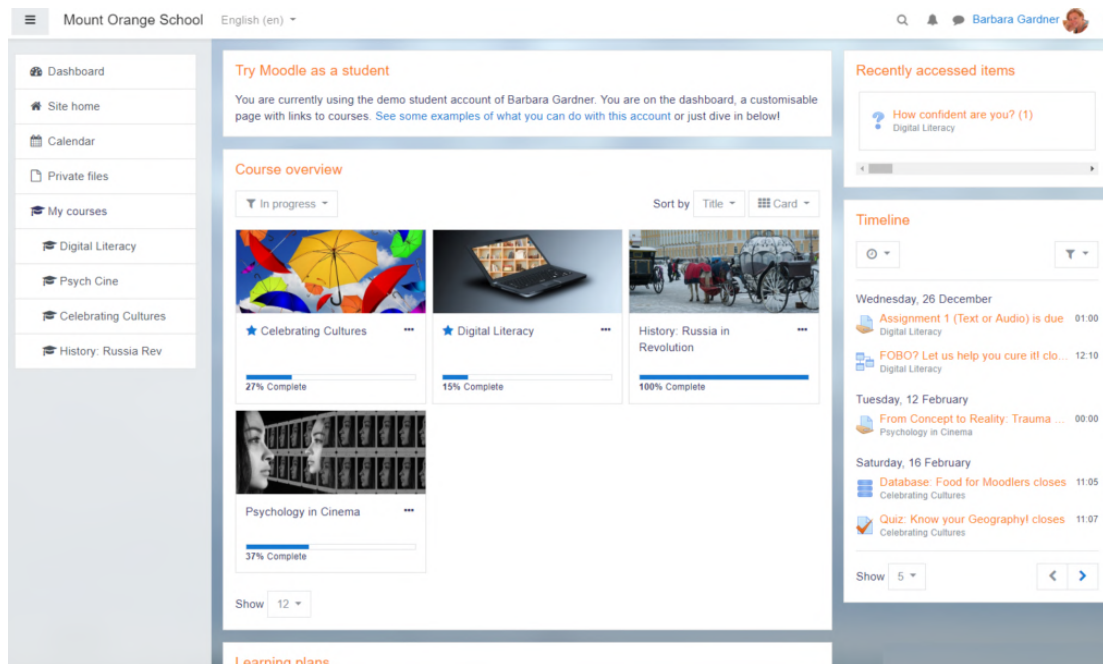


Abbildung 2.10: Open Source Kurs Management System Moodle mit einer Übersicht von Kursen und angefangenen Aktivitäten (Bildquelle: Moodle 2018)

Streicher und Roller 2015 entstanden Framework E-Learning Artificial Intelligence (ELAI) auf. ELAI ist als eigenständige Software entwickelt und dient der Entkopplung von Serious Games und dem adaptiven Verhalten. Spiele, wie beispielsweise Spot-X und Lost Earth 2307, verwenden intern ELAI, um adaptive Lernmethoden zu nutzen [Streicher und Roller 2015].

In Abbildung 2.11 ist die zugrunde liegende Softwarearchitektur dargestellt. Das SG ist dabei mit der ELAI Schnittstelle verbunden. Über eine Kommunikationsschicht werden Experience API (xAPI) Statements mit ELAI geteilt. Mithilfe von xAPI wird das Verhalten des Nutzers erfasst und die so gewonnenen Daten anschließend in einem Learning Record Store (LRS) gespeichert. Diese Technologien werden im nächsten Abschnitt genauer erläutert. Aufgrund der generischen Implementierung kann ELAI in Kombination mit unterschiedlichen Plattformen verwendet werden [Streicher, Roller und Biegemeier 2017]. Darüber hinaus lässt sich die ELAI Komponente in zwei eigenständige Teile trennen [Streicher, Roller und Biegemeier 2017]:

- „Interpretation Engine“: Ist für die Analyse der erfassten Daten im LRS zuständig. Um die Daten für ELAI in interpretierbare Informationen umzuwandeln, können verschiedene Methoden der Datenanalyse für Performance-Messungen, wie z.B. KT oder PFA, verwendet werden.

- „Influence Engine“: Ist dafür zuständig, falls nötig, eine Adaptionstrategie auszuwählen (z.B. ein Hinweis oder eine Änderung des Schwierigkeitsgrades).

Die verschiedenen Phasen aus dem Adaptivitätszyklus nach [V. J. Shute u. a. 2012] werden hier erkennbar. Die Kommunikationsschicht realisiert dabei die Erfassungsphase durch das Sichern der xAPI Statements und die Analysephase wird von der „Interpretation Engine“ ausgeführt. In der Selektionsphase ist die „Influence Engine“ dafür zuständig, entsprechend einer konkreten Adaptivitätsstrategie in den Lernprozess einzugreifen und den Verlauf an die Bedürfnisse des Nutzers anzupassen. In der Präsentationsphase wird eine mögliche Anpassung an das E-Learning System weitergeleitet und anschließend durch integrierte Funktionen vorgestellt.

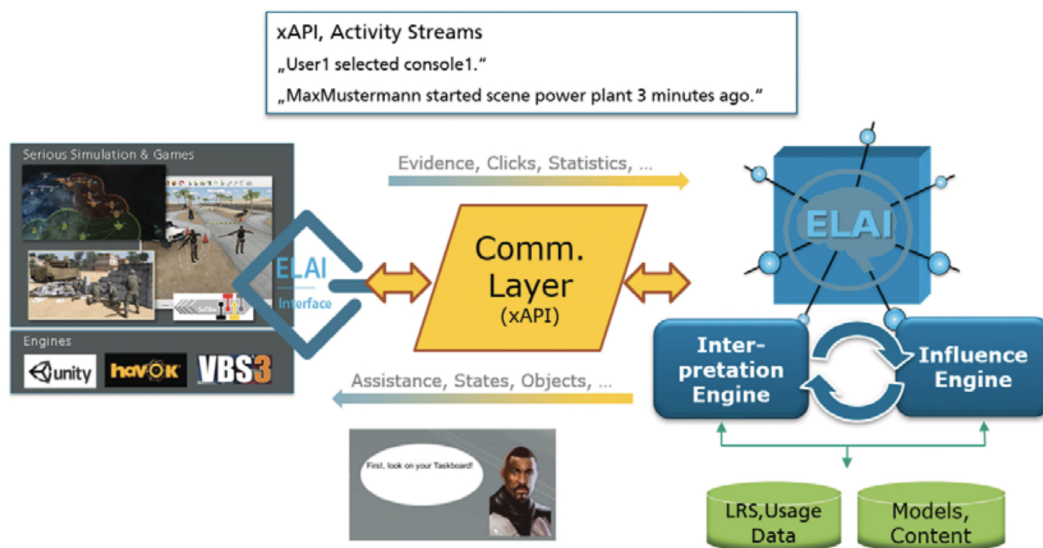


Abbildung 2.11: E-Learning Artificial Intelligence (ELAI) Framework Architektur mit Anbindung an ein SG [Streicher, Roller und Biegemeier 2017]

In dieser Arbeit wird das ELAI Framework um eine Performance-Messung erweitert. Weiterhin wird konzeptionell gezeigt, dass sowohl ILLIAS als auch Moodle, ELAI für eine Lernfortschrittsmodellierung nutzen können (s. Kapitel 4). Abschließend wird im Rahmen einer Nutzerstudie eine Performance-Messung für das Quizlernspiel Spot-X implementiert und anschließend ausgewertet (s. Kapitel 6).

2.6 Beobachtungsdaten durch Experience API (xAPI)

Wie bereits in den vorherigen Abschnitten beschrieben, werden für die Lernfortschrittsoptimierung durch Adaptivität Daten des Nutzers benötigt. In dieser Arbeit wird die Experience API (xAPI) Spezifikation [xAPI 2023] verwendet, um die Datenakquirierung in einem einheitlichen Format sicherzustellen. Diese Spezifikation dient dem effektiven und flexiblen Datenaustausch zwischen verschiedenen unabhängigen Systemen, wie z. B. zwischen Lernmanagementsystemen (LMS) und Serious Games [Streicher und Smeddinck 2016]. Mit einer solchen Standardisierung des Formats wird xAPI durch den Austausch von Lernerfahrungen des Nutzers zu einem nützlichen Instrument der Interoperabilität für E-Learning-Plattformen [xAPI 2023]. Die erfassten Daten werden anschließend in einem festgelegten Format an einen Learning Record Store (LRS) übermittelt, welches die Daten mit anderen LRSs teilen kann. Ein LRS kann auch in ein LMS integriert sein [xAPI 2023].

Das festgelegte Format sind sogenannte Statements, die aus einem Triple aus Akteur, Verb und Objekt bestehen [Kevan u. a. 2016]. Das Statement wird im JavaScript Object Notation (JSON) Format gespeichert. Dabei ist Akteur eine eindeutige, plattformunabhängige Information, welche das Subjekt der Aktivität beschreibt, wie z.B. ein Student oder ein Nutzer. Das Verb repräsentiert eine Aktivität, die der Akteur durchführt, z.B. spielen, lesen oder springen. Die Objekteigenschaft ist zumeist ein Nebenereignis, das zu der Aktivität gehört [Kevan u. a. 2016]. Beispielsweise könnte ein Statement für einen Akteur „Bob“, der soeben einen Mathematik-Kurs „Pythagoras“ begonnen hat, wie folgt lauten: „Bob hat Mathematik-Kurs „Pythagoras“ begonnen.“

Auflistung 2.1 zeigt ein xAPI-Statement, das aus einem LRS mit Spot-X Daten stammt. Dem Statement ist zu entnehmen, dass ein Akteur namens „German“ eine Bundeslandaufgabe in Form einer HotSpot-Suche abgeschlossen hat.

```
1 {  
2   "actor": {  
3     "objectType": "Agent",  
4     "mbox": "mailto:germanp@user.de",  
5     "name": "German"  
6   },  
7   "verb": {  
8     "id": "http://adlnet.gov/expapi/verbs/completed",  
9     "display": {  
10      "en": "task completed"  
11    }  
12  },
```

```

13     "object": {
14         "id": "https://www.iosb.fraunhofer.de/spotx/game/play/55",
15         "objectType": "Activity",
16         "definition": {
17             "type": "http://adlnet.gov/expapi/activities/question",
18             "name": {
19                 "en": "Deutschland/Bundesland"
20             },
21             "description": {
22                 "en": "Played a HotSpot about Deutschland/Bundesland and
23                     answered it correct."
24             }
25         }
26     }

```

Auflistung 2.1: Ein Beispiel für ein xAPI Statement aus dem Quizlernspiel Spot-X

Statements können noch weitere Informationen enthalten, wie z.B. einen Kontext oder Ergebnisse der aktuellen Aktivität [Kevan u. a. 2016]. Diese Ergänzungen können dem xAPI Statement mit zusätzlichen Feldern hinzugefügt werden. In Auflistung 2.2 ist eine Erweiterung des vorherigen Statements zu sehen. Dieses Statement wurde um ein Ergebnisfeld ergänzt, in welchem zu erkennen ist, dass der Nutzer „German“ die korrekte Antwortmöglichkeit „Bayern“ nach 23 Sekunden ausgewählt hat und eine Punktzahl von 79/100 Punkten erzielt hat. Diese zusätzlichen Felder bieten weitere Möglichkeiten für detailliertere Performance-Messungen. Wie diese Felder für die Performance-Messungen zum Einsatz kommen, wird in Kapitel 4 genauer erläutert.

```

1     "result": {
2         "success": true,
3         "duration": "PT23S",
4         "response": "Bayern(correct)",
5         "score": {
6             "raw": 79,
7             "min": 0,
8             "max": 100
9         },
10        "completion": true
11    }

```

Auflistung 2.2: xAPI Statement mit einer Ergebniserweiterung

Nachdem das Format vorgestellt wurde, mit welchem die Datenmengen gespeichert werden, erfolgt im anschließenden Unterkapitel die Vorstellung von Methoden und Modellen der Datenanalyse.

2.7 Performance & Lernfortschrittsoptimierung

Wie bereits in den vorherigen Kapiteln beschrieben, gibt es Architekturen, wie z.B. den Adaptivitätszyklus von V. J. Shute u. a. 2012, die es ermöglichen, sich an den Nutzer anzupassen. Dazu wird eine Variante von Performance-Messungen verwendet [Pavlik Jr u. a. 2009]. In dieser Arbeit wird der Lernfortschritt auch als Performance des Lernenden bezeichnet. Im Rahmen dieser Arbeit wird der Begriff Performance im Kontext von adaptiven Assistenten und E-Learning-Plattformen bzw. Serious Games betrachtet. Nach Schmucker u. a. 2022 definiert sich die Performance eines Nutzers als Fähigkeit, verschiedene Fragen zu beantworten. Die Performance wird anhand der sequentiell erfassten Daten, welche die Interaktionen mit dem System beschreiben, berechnet [Schmucker u. a. 2022]. In Kapitel 2.6 wurde eine Möglichkeit erläutert, diese Daten mithilfe von xAPI-Statements zu erfassen.

In [Wiemeyer u. a. 2016] wird beschrieben, dass Performance für Serious Games zwei Komponenten umfasst. Die erste Komponente ist der Fortschritt (engl. progress), der sich mit der Zeit entwickelt. Es ist zum Beispiel notwendig, ein Problem zunächst zu betrachten, um es lösen zu können. Anschließend überlegt man sich mögliche Lösungsansätze, um letztlich einen davon auszuwählen und zu realisieren. Die zweite Komponente stellen die Beobachtungen sowie messbaren Ergebnisse (engl. result) dar, welche die Performance hervorbringt. Damit beinhaltet Performance sowohl den Fortschritt als auch die Ergebnisse, die der Nutzer durch Aktionen mit dem Spiel entwickelt [Wiemeyer u. a. 2016]. Dabei kann es durchaus vorkommen, dass eine identische Performance durch unterschiedliche Ereignisse gemessen wird.

In der vorliegenden Arbeit werden für die beiden Performance-Komponenten die Begriffe Progress-Performance ($Perf^P$) und Result-Performance ($Perf^R$) verwendet. Eine Möglichkeit, die $Perf^P$ zu berechnen, wurde in den Arbeiten von Delaunay 2022 und Baur 2022 untersucht. Baur 2022 verwendet den Begriff $Perf^P$, um zu beschreiben, wie effektiv ein Nutzer lernt. Im Allgemeinen wird damit eine Fortschrittsgeschwindigkeit gemessen. In Delaunay 2022 wird $Perf^P$ als die Effizienz der Zielerreichung beschrieben. Diese Arbeit orientiert sich hinsichtlich der Progress-Performance an der begrifflichen Definition von Delaunay 2022. In Kapitel 2.8 wird im Detail erläutert, wie die Progress-Performance mithilfe sogenannter Ideal Path Models (IPMs) berechnet wird. Für die Berechnung der $Perf^R$ bestehen unterschiedliche Methoden. In Pustovojtovskij 2021 wurde das logistische Regressionsmodell Performance Factor Analysis (PFA) untersucht und auf ein SG angewendet. Eine weitere Möglichkeit die Result-Performance

vorherzusagen, besteht anhand des Graph Knowledge Tracing (GKT) Modell, welches in Delaunay 2022 verwendet wurde. In Kapitel 2.9 werden die Methoden zur Berechnung von $Perf^R$ genauer veranschaulicht.

Im Bereich E-Learning werden zwei Methoden zur Verbesserung der Lernfortschrittsmodellierung eingesetzt [Cristobal Romero u. a. 2020]:

- Educational Data-Mining (EDM): Methoden, um Daten zu untersuchen, die aus einer Bildungsumgebung stammen. Dabei werden Techniken aus dem Bereich des Data-Minings verwendet [C. Romero u. a. 2007]. Die Motivation für EDM und ein EDM Modell LFA wird in Unterkapitel 2.9.4 im Detail erläutert.
- Learning Analytics (LA): Modelle, die sich mit dem Messen, Akquirieren und Analysieren von Daten aus einer Bildungsumgebung beschäftigen. Darauf folgt eine Aktion, welche mit dem Lernenden oder seiner Umgebung zu tun hat, um den Lernprozess zu verstehen und zu optimieren [Lang u. a. 2017]. Bei dieser Methode sind drei Komponenten ausschlaggebend: Daten, Analyse und Aktion [Siemens 2013].

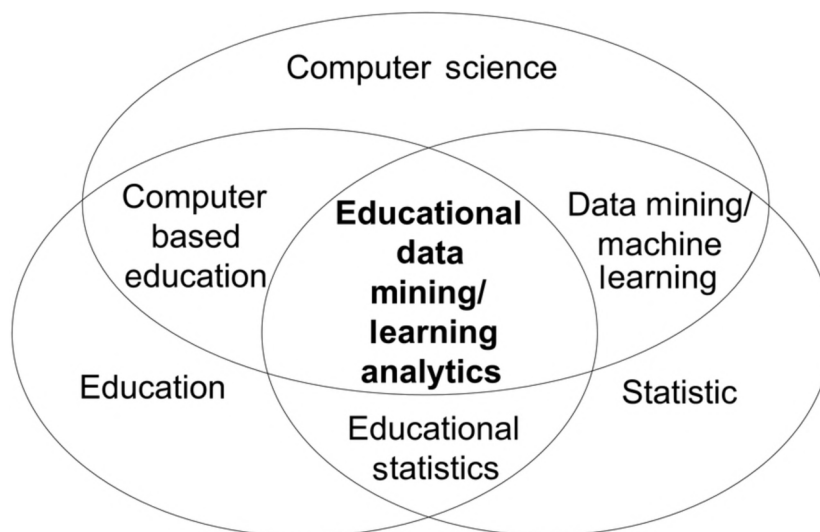


Abbildung 2.12: Hauptbereiche, die mit Educational Data-Mining und Learning Analytics verwandt sind [Cristobal Romero u. a. 2020].

Wie sich die beiden Methoden im Bereich der Statistik, Bildung und Informatik eingliedern lassen, ist in Abbildung 2.12 zu sehen. Die drei Komponenten der Learning Analytics nach Siemens 2013 können auch als eine alternative Darstellung des Vier-Phasen Adaptivitätszyklus (s. Abbildung 2.3) nach Streicher und Smeddinck 2016 betrachtet werden, wenn man die

Phasen Select und Present als eine Aktion betrachtet. Aufgrund des iterativen Prozesses der Performance-Messung und der adaptiven Anpassungen des Systems mit Bezug auf einen konkreten Nutzer liegt der Fokus dieser Arbeit auf den Learning Analytics [Lang u. a. 2017]. Bis auf die Learning Factor Analysis (LFA) lassen sich alle im Abschnitt 2.9 erläuterten Modelle der LA Methode zuordnen [Pavlik Jr u. a. 2009]. Die LFA wird in diesem Kapitel dennoch aufgeführt, da die Performance Factor Analysis (PFA) auf der LFA aufbaut.

2.8 Progress-Performance mit Graph-basierten Ansätzen

In diesem Abschnitt werden grundlegende Modelle für die Berechnung der $Perf^P$ definiert, welche in dieser Arbeit verwendet werden. Zunächst werden die verschiedenen Begriffe aus der Graphentheorie definiert, um Graph-basierte Lernpfade anzuwenden. Anschließend werden einige Algorithmen vorgestellt, die auf Graphen eingesetzt werden können, um kürzeste Pfade zu berechnen. Auf diesen Techniken aufbauend wird die Modellierung der Graph-basierten Lernpfade vorgestellt. Zum Schluss wird der Begriff der $Perf^P$ im Kontext von Lernpfaden für adaptive Assistenzsysteme definiert und geklärt, wie die Berechnung von $Perf^P$ mithilfe der Lernpfade erfolgt.

2.8.1 Graphentheorie

Ein Graph ist ein Tupel $G = (V, E)$ bestehend aus einer Knotenmenge V und einer Kantenmenge E . Häufig werden Knoten als Punkte und Kanten als Linien zwischen den Punkten dargestellt [Diestel 2017]. Die Kantenmenge ist disjunkt, wobei $E \subseteq |V|^2$. Eine Kante ist somit eine zweielementige Teilmenge der Knoten V . In Abbildung 2.13 ist ein Graph G_1 mit der Knotenmenge $V = \{A, B, C\}$ und der Kantenmenge $E = \{(A, B), (B, C), (A, C)\}$ zu sehen.

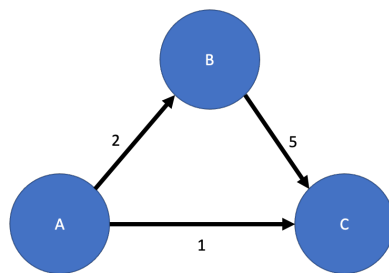
Ein gerichteter Graph besteht aus einer Menge von Knoten V und einer Menge von Kanten E , wobei $E \subseteq V \times V$ gilt [Briskorn u. a. 2020]. In Abbildung 2.13 ist links ein gerichteter Graph abgebildet. Rechts ist ein ungerichteter Graph abgebildet. Für den Graphen G_2 gilt $E = \{(A, B), (B, A), (B, C), (C, B), (A, C), (C, A)\} = \{\{A, B\}, \{B, C\}, \{A, C\}\}$.

Zusätzlich dazu gibt es gewichtete Graphen $G = (V, E, \omega)$ mit einer Gewichtsfunktion $\omega : E \rightarrow \mathbb{R}$, wobei je nach Anwendung ein anderer Wertebereich für die Funktion ω gewählt werden kann (z.B. \mathbb{R} oder \mathbb{N}_0) [Briskorn u. a. 2020]. In Abbildung 2.14 ist ein gewichteter Graph G_3 gegeben. Das Kantengewicht wird meistens über der jeweiligen Kante dargestellt. Zum Beispiel ist $\omega((A, B)) = 2$.

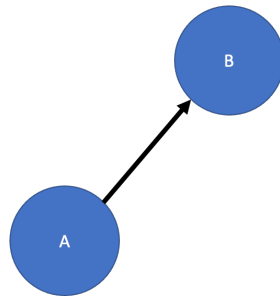
Ein Teilgraph $H \subseteq G$ ist ein Graph H mit der Knotenmenge $V(H) \subseteq V(G)$ und der Kantenmenge $E(H) \subseteq E(G)$, wobei man mit $V(G)$ und $E(G)$ die Knoten- und Kantenmenge von

gerichteter Graph G_1 ungerichteter Graph G_2

Abbildung 2.13: Beispiel eines gerichteten sowie eines ungerichteten Graphen

Abbildung 2.14: Ein Beispiel für einen gewichteten Graphen G_3

G bezeichnet [Diestel 2017]. In Abbildung 2.15 ist ein Teilgraph H_1 von dem Graphen G_1 in Abbildung 2.13 zu sehen.

Abbildung 2.15: Ein Beispiel für einen Teilgraphen $H_1 \subseteq G_1$

In dieser Arbeit werden für die $Perf^P$ hauptsächlich Pfade untersucht. Pfade sind ebenfalls Graphen mit $V = \{v_1, v_2, \dots, v_n\}$ und $E = \{(v_1, v_2), (v_2, v_3), \dots, (v_{n-1}, v_n)\}$ mit $n \in \mathbb{N}$ [Diestel 2017]. In Abbildung 2.16 ist ein Pfad P mit $V = \{A, B, C\}$ und $E = \{(A, B), (B, C)\}$ zu sehen.

Ein Weg bezeichnet eine Folge von besuchten Knoten $W = v_1 v_2 v_3 \dots v_n$ in einem Graphen G . Der Weg W besitzt dann den Startknoten v_1 und Endknoten v_n . Ein Graph kann auch mehrere Wege besitzen [Diestel 2017]. Der Graph G_3 in Abbildung 2.14 besitzt z.B. die Wege $W_1 = ABC$

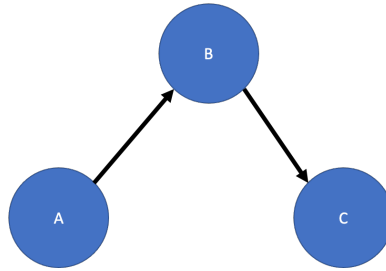


Abbildung 2.16: Ein Beispiel für einen Pfad aus dem Graphen G_1

und $W_2 = AC$ mit Startknoten A und Endknoten C . Ein Pfad stellt einen Weg dar, welcher jeden Knoten höchstens einmal besucht. Falls Start- und Endknoten gleich sind, bezeichnet man dies als Kreis. Ein Pfad ist somit kein Kreis [Briskorn u. a. 2020].

Kürzeste Pfade sind Pfade, welche eine minimale Anzahl an Kanten zwischen einem Start- und Endknoten besitzen und keinen Kreis beinhalten. Das Kürzeste-Wege-Problem besteht darin, einen solchen Pfad als Teilgraph zu finden [Needham u. a. 2019]. Um dieses Problem zu lösen, existieren unterschiedliche Algorithmen. Aufgrund der Nutzung einer Neo4j Graphen-Datenbank [Neo4j 2023], wird im Folgenden nur der Kürzeste-Wege-Algorithmus aus [Needham u. a. 2019] beschrieben. Der Grund dafür ist, dass dieser Algorithmus bereits in Neo4j angeboten wird.

Der in dieser Arbeit verwendete Algorithmus zur Bestimmung kürzester Pfade ist Yen's k -Shortest-Paths Algorithmus [Yen 1971]. Dieser findet den kürzesten Pfad und die bis zu $k - 1$ nächst-kürzeren Pfade in einem Graphen.

2.8.2 Graph-basierte Modellierung von Lernpfaden

In dieser Arbeit werden Graph-basierte Ansätze zur Bestimmung der Result-Performance und Progress-Performance verwendet, um die Performance des Nutzers zu bestimmen. Um die $Perf^P$ zu ermitteln, werden Graph-basierte Lernpfade genutzt. In diesem Abschnitt wird erklärt, wie diese Pfade in Graphen modelliert werden und auf welche Weise sich mithilfe von solchen Pfaden die $Perf^P$ berechnen lässt. Zunächst müssen die verschiedenen Arten von Pfaden definiert werden, um darauf aufbauend die Berechnung der $Perf^P$ anhand sogenannter idealer Pfade zu ermitteln.

Nutzungs- und Lernpfade

Ein Nutzungspfad beschreibt eine geordnete Sequenz von Interaktionen des Nutzers mit einem System. Diese Nutzungspfade werden mithilfe von Beobachtungsdaten erstellt und in Form

von gerichteten Graphen $G(V, E)$ modelliert, wobei die Knoten die Nutzerinteraktionen und den Nutzer selbst beinhalten [Streicher, Schönbein u. a. 2021]. In der oberen Hälfte von Abbildung 2.17 ist ein Nutzungspfad mit Interaktionen des Nutzers mit einem SG abgebildet. Lernpfade sind eine spezielle Form von Nutzungspfaden, welche sich auf den E-Learning-Bereich beziehen. Üblicherweise sind diese Lernpfade vordefinierte Sequenzen von Interaktionen mit einer E-Learning-Plattform. Lernpfade erweisen sich als essenzieller Bestandteil von adaptiven E-Learning-Systemen, da sie Informationen darüber liefern, wie die Lerneinheit strukturiert ist. Zusätzlich geben sie an, ob sich der Nutzer auf dem richtigen Weg befindet, um die Lerneinheit zu absolvieren. Weiterhin können Lernpfade zu Berechnung der Progress-Performance verwendet werden [Streicher, Schönbein u. a. 2021].

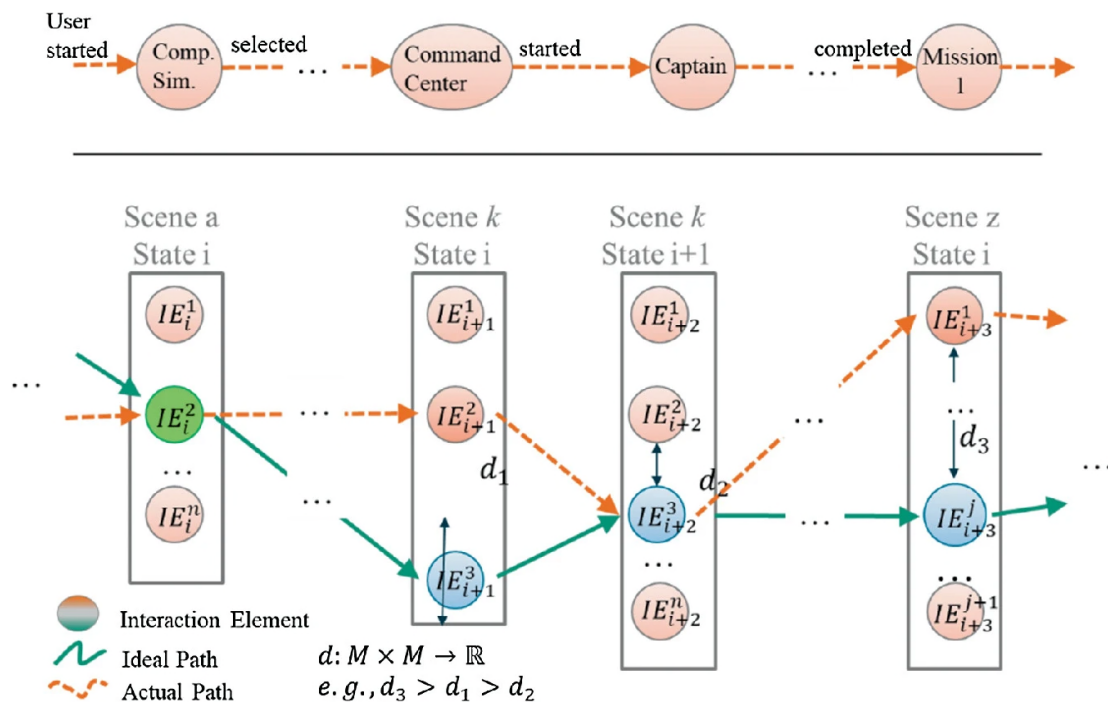


Abbildung 2.17: Beispiel für einen Nutzungspfad (oben) und das Ideal Path Model Konzept (unten) [Streicher, Leidig u. a. 2018]

Ideal Paths Models (IPMs)

Ein Lösungsansatz, wie sich Graph-basierte Modelle von Lernpfaden für adaptives Verhalten realisieren lassen, ist denkbar durch die Nutzung standardisierter Beobachtungsdaten. Diese Pfade werden anschließend dazu verwendet, Ideal Paths Models (IPMs) für adaptive Assistenzsysteme zu nutzen. Das Referenzmodell IPMs dient der Steuerung von Adaptivität

[Streicher, Schönbein u. a. 2021]. IPMs beschreiben alle notwendigen Schritte, um das Lernziel auf direktem Weg zu erreichen. Das kann beispielsweise ein kürzester Weg oder der Weg eines Experten, also eines erfahrenen Nutzers, sein. Einen solchen Pfad nennt man idealen Pfad (IP) [Streicher, Leidig u. a. 2018]. Ein IPM besteht laut Streicher, Leidig u. a. 2018 aus mehreren Komponenten, die in der unteren Hälfte von Abbildung 2.17 zu erkennen sind:

- Szene: Die Szene ist eine Menge von Zuständen, welche jeweils durch eine Interaktion und den Vorgängerzustand definiert sind (Interaktion als Zustandsübergangsfunktion).
- Interaktionen: Interaktionen werden aus den Beobachtungsdaten gewonnen und sind Aktionen, die vom Nutzer ausgeführt werden. Ein Beispiel für eine Interaktion auf einer E-Learning-Plattform könnte wie folgt lauten: Nutzer X schließt den Kurs „Deutschland Sehenswürdigkeiten“ erfolgreich ab.
- Ideale Pfade: Hierbei handelt es sich um die Durchläufe durch die Szenen und Interaktionen des Systems.
- Lernpfade: Der tatsächliche Pfad, den der Nutzer gewählt hat und welcher die ausgeführten Aktionen als Sequenz beschreibt.

Laut Baur 2022 kann man IPs auch als Expertenpfade betrachten. Experten sind Nutzer, die bei der Interaktion mit dem E-Learning-System stets die höchste Punktzahl erzielen und möglichst schnell das anspruchsvollste Lernmaterial erreichen. Die erfassten Daten können in Form von xAPI-Statements aufgezeichnet und anschließend als Knoten und Kanten für mögliche Pfade in Graphen abgebildet werden [Streicher, Schönbein u. a. 2021]. Aufgrund der bereits vorhandenen Graph- und Social-Network-Algorithmen bieten sich Graphendatenbanken an, um die Graphen zu sichern [Needham u. a. 2019]. Abbildung 2.18 zeigt ein Beispiel mit einer möglichen Sequenz an Aktivitäten $P = \{A, B, C, D, E, F\}$ und drei Nutzern mit drei unterschiedlichen Pfaden. Mithilfe von musterbasierter Suche für Graphen auf den Pfaden der Nutzer kann eine gemeinsame Menge von Knoten $C \subseteq P$ gefunden werden. Das Ergebnis ist ein gemeinsamer Pfad $C = \{A, C, F\}$. In diesem Kontext erweist sich der gemeinsame Pfad als idealer Pfad, da alle Nutzer die gleichen Aktivitäten in ihren Sequenzen besucht haben. Im Rahmen dieses Modells, bedeutet das, die Knoten des Pfades C sind relevant, um den Lernfortschritt steigern zu können. Damit bilden diese Knoten Lernziele für andere Nutzer. Jedoch besteht ein persönliches Lernziel des Nutzers in einer subjektiven Anschauung und dient daher nur als Basis zur Bestimmung von IPs [Streicher, Schönbein u. a. 2021].

Die Progress-Performance lässt sich berechnen, indem der aktuelle Lernpfad mit einem IP verglichen, gewichtet und kombiniert wird. Dies bedeutet, dass die Ähnlichkeit des aktuellen

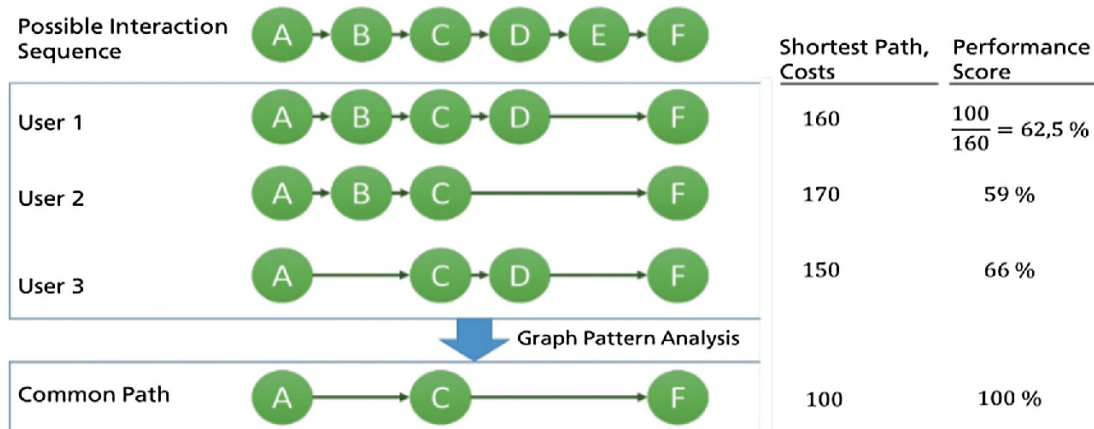


Abbildung 2.18: Konzept, um einen gemeinsamen Pfad zu identifizieren und eine Performance zu berechnen [Streicher, Schönbein u. a. 2021].

Nutzungspfads mit dem idealen Pfad gemessen wird. Je höher die Übereinstimmung des aktuellen Nutzungspfads mit den IPs, desto höher ist die $Perf^P$ des Nutzers und somit auch dessen Lerneffizienz [Baur 2022]. In Delaunay 2022 wird das Referenzmodell IMPs zusätzlich dazu verwendet, um adaptive Lernpfade zu ermöglichen, indem die Lerninhalte auf dem IP empfohlen werden. Um die $Perf^P$ für einen aktuellen Nutzungspfad zu berechnen, wurde in Baur 2022 folgende Formel definiert:

$$Perf^P(P_U, P_I) = \frac{Progress}{Time} = \alpha \frac{|V(P_U) \cap V(P_I)|}{V(P_U)} + (1 - \alpha) \frac{|E(P_U) \cap E(P_I)|}{E(P_U)} \quad (2.1)$$

Die Erklärung der einzelnen Variablen:

- P_U : Der Lernpfad eines Nutzers
- P_I : Ein idealer Pfad
- α : Ein Gewichtungsparemeter mit $\alpha \in [0, 1]$

Es gilt weiterhin $Perf^P \in [0, 1]$. Aufgrund mehrerer idealer Pfade gibt es nicht nur eine $Perf^P$ sondern mehrere, allerdings soll dem Nutzer nur ein Lernpfad empfohlen werden. Daher wird der IP verwendet, auf welchem der Nutzer die höchste Performance erreicht [Baur 2022]:

$$Perf^P(P_U) = \max_{P_I \in IP_s} Perf^P(P_U, P_I) \quad (2.2)$$

Abschließend wird zur $Perf^P$ noch der Graph Performance Service (GPS) vorgestellt, welcher im Rahmen von Baur 2022 implementiert und von Delaunay 2022 erweitert wurde.

2.8.3 Graph Performance Service (GPS)

Zur Berechnung der Progress-Performance eines Nutzers mithilfe von Lernpfaden wurde der Graph Performance Service (GPS) von Baur 2022 mit Python implementiert und von Delaunay 2022 erweitert. Der GPS führt anhand der Berechnung aus dem vorherigen Abschnitt Vergleiche zwischen dem aktuellen Lernpfad des Nutzers und den IPs durch. Mittels dieser Berechnung wird gemessen, wie effizient der Nutzer auf seinem aktuellen Lernpfad ist. Darauf aufbauend spricht der GPS eine Empfehlung für die nächste Lerneinheit aus [Delaunay 2022].

Der GPS nutzt die Graphdatenbank Neo4j, um den Graphen aus xAPI-Statements zu generieren und zu sichern. Dabei bilden die Akteure und Objekte die Knoten des Graphen und die Verben werden als Kanten hinzugefügt. Die xAPI Daten werden aus einem LRS des Fraunhofer IOSB geladen. Der Graph wird aus den Daten mittels sogenannten Cypher-Anfragen erstellt. Cypher ist eine deklarative Programmiersprache für die Datenabfrage in Neo4j [Neo4j 2023]. Die Cypher-Anfragen stellen zusätzlich eine große Auswahl an Graphalgorithmen, wie z.B. den Yen's k -Shortest-Paths Algorithmus, zur Verfügung [Baur 2022].

Als Microservice wird der GPS mittels REST-API angefragt. Eine Antwort vom Service erfolgt im JSON Format und kann wie in Auflistung 2.8.3 abgebildet aussehen. Nach Baur 2022 besteht die Antwort aus drei Komponenten: „success“, „performance“ und „recommendation“. „success“ beschreibt, ob die Anfrage erfolgreich bearbeitet wurde. „performance“ teilt die $Perf^P$ als Gleitkommazahl mit und „recommendation“ gibt die nächste empfohlene Lerneinheit an, falls vorhanden.

```

1 {
2   "success": true ,
3   "performance": 0.75 ,
4   "recommendation": {
5     "next_activity": {
6       "name": "HarryPotterBridge" ,
7       "id": "'https://www.iosb.fraunhofer.de/spotx/game/play/47'"
8     }
9   }
10 }
```

2.9 Result-Performance

In diesem Abschnitt werden die verwendeten und zusammenhängenden Modelle erläutert, welche für die Berechnung der Result-Performance ($Perf^R$) nötig sind. Zunächst wird jedoch der Begriff der Wissenskomponente (WK) (engl. knowledge component) definiert werden.

Anschließend wird das Knowledge Tracing (KT), eine in der Literatur weit verbreitete Methode, vorgestellt [De Sande 2013]. Zum Schluss erfolgt eine Erläuterung des in dieser Arbeit operationalisierten Modells Performance Factor Analysis (PFA) sowie dessen Vergleich mit KT. In diesem Kapitel werden ausschließlich die klassischen Modelle vorgestellt. Eine Gegenüberstellung der unterschiedlichen Varianten der Methoden KT und PFA erfolgt in Kapitel 3.2.

2.9.1 Wissenskomponenten

Um die Modelle zur Berechnung von Result-Performance näher zu beleuchten, ist es zunächst notwendig, die Begriffe Daten und Wissenskomponente (WK) zu definieren. Wenn in dieser Arbeit von Daten die Rede ist, sind damit die gesammelten Daten des Nutzers in einem E-Learning-System, wie z.B. Spot-X oder Moodle, gemeint. Wie diese Daten mittels Experience API Statements akquiriert werden können, wurde bereits in Abschnitt 2.6 erläutert. Die Daten werden in Situationen generiert, in deren Analyse einen Mehrwert bringt. Im Fall von Serious Games, wie Spot-X, werden die xAPI Statements nach Abschluss jeder Quizfrage generiert. Bei E-Learning Plattformen wie Ilias können die Statements nach Abschluss der Aufgabe generiert werden. Die Gemeinsamkeit all dieser E-Learning-Systeme ist, dass das Wissen des Nutzers abgefragt werden soll. Das Ziel dieser Systeme besteht darin, den Lernfortschritt des Nutzers in Form des besseren Abschneidens bei Aufgaben oder Problemen zu steigern.

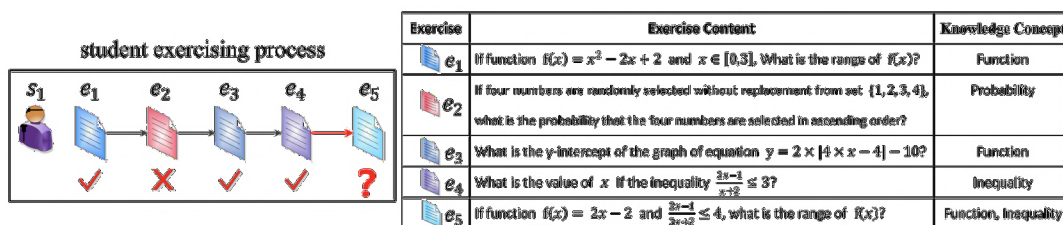


Abbildung 2.19: Die linke Box zeigt den Lernprozess eines Nutzers, der bereits vier Fragen beantwortet hat und als Nächstes die fünfte Frage beantworten wird. Die rechte Tabelle zeigt die einzelnen Aufgaben und die betroffenen Wissenskomponenten [Liu, Huang u. a. 2021].

In den wissenschaftlichen Artikeln über die Lernfortschrittsmodellierung für intelligente Tutoring Systeme wird der Begriff Wissenskomponente häufig verwendet. Corbett u. a. 1994 beschreiben erstmals Knowledge Tracing. In diesem Artikel wird eine Methode vorgestellt, mit welcher der Lernfortschritt eines Nutzers für intelligente Tutor Systeme messbar gemacht werden kann. Corbett u. a. 1994 beschreiben den Begriff WK als „component of a knowledge-rich skill“, also als Komponente einer wissenshaltigen Fähigkeit. Laut [Kenneth R Koedinger

u. a. 2012] definiert sich eine WK wie folgt: Um eine Aufgabe oder ein Problem lösen zu können, muss der Nutzer die betreffende/n Wissenskomponente/n verstanden haben. Eine WK ist also eine Generalisierung von alltäglichen Termen wie Konzepte, Prinzipien, Fähigkeiten oder Fakten. Eine WK kann somit alleine oder in Kombination mit anderen WKs dazu verwendet werden, erfolgreich ein (Teil-) Problem zu lösen. In Abbildung 2.19 erkennt man, wie sich die WKs für einen Lernprozess zusammenstellen. In Gong, Joseph E Beck u. a. 2010 wird der Term Wissenskomponente mit dem Begriff Fähigkeit (engl. Skill) gleichgesetzt. Dabei werden Probleme oder Aufgaben, die mehrere Wissenskomponenten einbeziehen, als Multi-Skill Fragen bezeichnet. Ähnlich hierzu beschreiben auch S. Pu u. a. 2021, dass WKs im Rahmen ihrer Arbeit ein Äquivalent zu Fähigkeiten darstellen. In der Veröffentlichung von [Nakagawa u. a. 2019] wird der Begriff WK durch den Begriff Wissenskonzept ersetzt und auch in der Abschlussarbeit von Delaunay 2022 so übernommen. Pavlik Jr u. a. 2009 verwendet den Term WK und definiert ihn analog zu Kenneth R Koedinger u. a. 2012 als eine Generalisierung für Fähigkeit, Konzept oder Fakt. In der vorliegenden Arbeit wird der Begriff Wissenskomponente angewendet, da die Definition von Kenneth R Koedinger u. a. 2012 einen umfangreicheren Term bietet und sowohl Konzepte als auch Fähigkeiten miteinbezieht. Ein weiterer Grund, diesen Begriff zu wählen, ist es, einen Zusammenhang zwischen dieser Arbeit und den Arbeiten von Pavlik Jr u. a. 2009 und Pustovojtovskij 2021 herzustellen.

2.9.2 Knowledge Tracing (KT)

Knowledge Tracing (KT) erweist sich als eines der grundlegenden Modelle, um die Performance des Nutzers zu bestimmen und in adaptiven Assistenzsystemen einzusetzen [Liu, Shen u. a. 2021]. Laut De Sande 2013 hat KT zwei unterschiedliche Anwendungsfälle:

- Das Bayesian Knowledge Tracing (BKT) Hidden Markov Modell (HMM) sagt die Wahrscheinlichkeit vorher, mit der ein Nutzer eine WK korrekt anwenden wird, wenn er eine Möglichkeit dazu hat. Üblicherweise wird dieses Modell zur Berechnung der Performance des Nutzers verwendet.
- Der Knowledge Tracing Algorithmus wird genutzt, um in Echtzeit zu bestimmen, ob ein Nutzer eine WK gelernt hat oder nicht. Die Wahrscheinlichkeit, dass die WK verstanden wurde, wird durch die Leistung des Nutzers in jeder Anwendung der WK angepasst.

Das Hidden Markov Modell ist eine bekannte Technik zur Verhaltensmustererkennung, welche in adaptiven Tutoring Systemen eingesetzt wird [Yadav u. a. 2020]. Zur Berechnung der Wahrscheinlichkeit für das Verständnis der WK werden die vergangenen erfolgreichen sowie

fehlgeschlagenen Anwendungen der WK des Nutzers verwendet. Diese berechnete Wahrscheinlichkeit wird in adaptiven Tutoring Systemen anschließend als Metrik für den Lernerfolg dieser WK genutzt, um herauszufinden, ob der Nutzer eine konkrete Wissenskomponente beherrscht [Falakmasir u. a. 2013].

Laut Falakmasir u. a. 2013 ist das HMM in der Anwendung ein zweistufiges Verfahren. In der ersten Etappe wird der vergangene Verlauf der Wissenskomponente ausgegeben. Aus diesem Verlauf lassen sich folgende fünf Parameter bestimmen [Gong, Joseph E Beck u. a. 2010]:

- Initiales Wissen: Die Wahrscheinlichkeit, dass der Nutzer bei der ersten Anwendung des adaptiven Systems die Wissenskomponente bereits beherrscht.
- Lernrate: Die Wahrscheinlichkeit, dass der Nutzer die Fähigkeit bei der Gelegenheit erwirbt, die WK zu nutzen.
- Schätzung (Guess Parameter): Beschreibt die Möglichkeit, dass der Nutzer eine korrekte Antwort generiert hat, obwohl er die WK nicht beherrscht.
- Ausrutscher (Slip Parameter): Sogar Nutzer, die eine WK beherrschen, können Fehler generieren. Dies kann aufgrund von Unachtsamkeit oder Konzentrationsverlust erfolgen.

Wie in Abbildung 2.20 zu sehen ist, wird der aktuelle Wissensstand der WK mithilfe der Parameter iterativ berechnet.

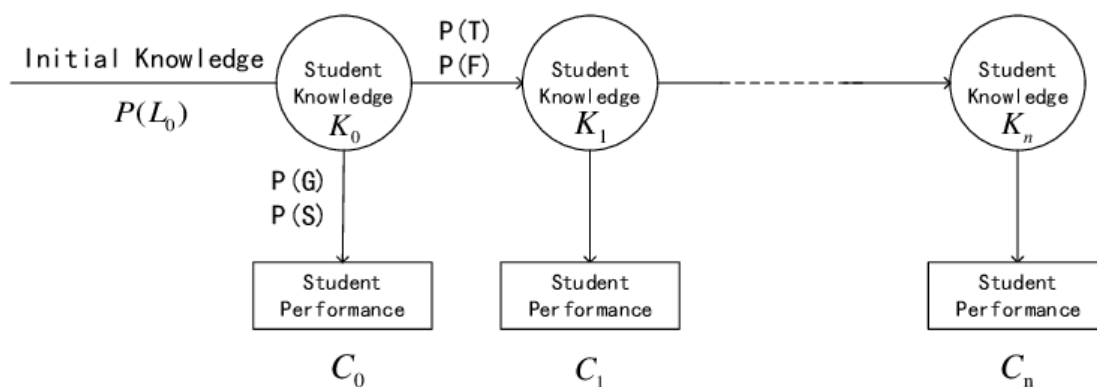


Abbildung 2.20: Knowledge Tracing Modell [Y. Pu u. a. 2018] mit $P(L_0)$ als initialer Wissensstand; $P(T)$, $P(F)$ als Lernrate über Erfolg und Misserfolg; $P(G)$, $P(S)$ als Guess und Slip Parameter; C_i mit $i \in \{1, 2, \dots, n\}$ für die Performance des Nutzers im Iterationsschritt i .

Probleme mit Knowledge Tracing

Wie bereits in Joseph E. Beck u. a. 2007 erwähnt, gibt es bei KT zwei Schwierigkeiten hinsichtlich der Bestimmung von Parametern: lokale Minima und diverse globale Maxima. Das erste Problem, welches durch sogenannte „Error Surface“ hervorgerufen wird, kann durch mehrere Neustarts behoben werden. Das Problem der diversen globalen Maxima ist auch als Problem der Mehrdeutigkeit bekannt [Gong, Joseph E Beck u. a. 2010]. Demnach passen für dieselbe Modellstruktur mit gleichen Daten mehrere Mengen an Parametern. Aufgrund der Gleichwertigkeit der unterschiedlichen Lösungsmengen werden zusätzliche Eigenschaften, wie die Genauigkeit sowie die Parameterplausibilität, hinzugezogen, um für das Modell eine geeignete und eindeutige Lösung zu finden. Für dieses Problem gibt es ebenfalls Lösungsansätze, wie z.B. der Expectation Maximization (EM) Algorithmus [Moon 1996] oder der Brute Force Ansatz [Baker u. a. 2008]. Die Ansätze, welche versuchen, die Probleme zu umgehen, sind jedoch schwer umsetzbar, da sich die Modellsuche als äußerst rechenintensiv erweist [Gong, Joseph E Beck u. a. 2010]. Im Gegensatz zu KT bestehen für die auf der logistischen Regression basierenden Verfahren LFA und PFA keine Schwierigkeiten beim Erreichen von globalen Maxima.

Das Bayesian Knowledge Tracing Framework besitzt einen weiteren großen Nachteil: Wenn eine Frage gestellt wird, die erfordert, dass diverse Wissenskomponenten gleichzeitig geprüft werden, fehlt dem Modell die Fähigkeit alle WKs gleichzeitig zu behandeln [Gong, Joseph E Beck u. a. 2010]. Ursache des Problems ist, dass die KT bei einer Abfrage ausschließlich den vergangenen Verlauf einer einzelnen WK überprüft. Für ein adaptives Tutoring System ist es jedoch wichtig, die zuvor genannte Eigenschaft zu besitzen, da viele Aufgabenstellungen meist nicht nur eine, sondern mehrere WKs zeitgleich abfragen. Für dieses Problem existieren ebenfalls Lösungen, wie z.B. die Frage bei der Berechnung zu wiederholen, bis alle Wissenskomponenten der Frage einmal geprüft wurden. Dies bedeutet, dass die Frage mit mehreren WKs in viele Fragen mit nur einer WK aufgeteilt wird. Dabei wird die Wahrscheinlichkeit auf Overfitting erhöht [Pardos u. a. 2010] und die Aufteilung erhöht den Rechenaufwand. Bei der LFA und PFA Modellierung entsteht jenes Problem nicht, da beide diese Eigenschaften bereits implizit besitzen.

2.9.3 Logistische Regressionsmodelle

Um im Folgenden auf die Modelle LFA sowie PFA einzugehen, ist es zunächst notwendig, die logistische Regression als Grundstein der PFA näher zu betrachten. Wie bereits in Pustovojtovskij 2021 beschrieben, ist die Regressionsanalyse eine konzeptuelle Methode, um funktionelle Beziehungen zwischen Variablen zu untersuchen [Chatterjee u. a. 2015]. Dabei

wird die Beziehung der Variablen in Form einer Gleichung oder eines Modells dargestellt. Die Beziehung besteht dabei zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen. Solche Gleichungen oder Modelle nennt man Regressionsmodelle. Es gibt viele unterschiedliche Regressionsmodelle, wie z.B. das multiple oder lineare Regressionsmodell. Die in dieser Arbeit behandelte PFA basiert auf dem klassischen logistischen Regressionsmodell. Aus diesem Grund wird nur auf dieses Modell weiter eingegangen und die anderen werden nicht erwähnt.

Das logistische Regressionsmodell, oder auch Logit-Modell, ist eine Erweiterung des linearen Regressionsmodells. Es wird verwendet, um die Wahrscheinlichkeit für das Eintreten eines bestimmten Ereignisses darzustellen. Dabei wird das Ereignis als kategoriales Kriterium vorhergesagt [Kleinbaum u. a. 2002]. Zum Beispiel liegt eine Patientenakte vor und man möchte prüfen, ob der Patient an Krankheit X erkrankt. Dies wäre ein kategoriales Kriterium, da der Patient in Bezug auf Erkrankung X , entweder krank oder gesund ist. Die Zuordnung des Ereignisses zu einer Klasse nennt man Klassifikation. Die logistische Regression wird in die zwei folgenden Regressionsmodelle unterteilt:

- Binär logistische Regression: Die abhängige Variable des Regressionsmodells verfügt über zwei Ausprägungen (dichotom).
- Multinomiale logistische Regression: Die abhängige Variable des Regressionsmodells besitzt mehr als zwei Kategorien.

Die vorliegende Arbeit behandelt ausschließlich die binäre logistische Regression, da geprüft wird, ob die PFA das Wissen des Nutzers als verstanden oder nicht verstanden klassifiziert. In Abbildung 2.21 erkennt man den Unterschied zwischen der logistischen und der linearen Regressionsfunktion. Die logistische Funktion wird auch als Sigmoidfunktion bezeichnet. Diese wird in Abschnitt 2.9.4 nochmal aufgegriffen.

2.9.4 Learning Factor Analysis (LFA)

Learning Factor Analysis (LFA) ist eine Educational Data-Mining Methode und das Fundament für das im nächsten Unterkapitel vorgestellte Performance Factor Analysis Modell [Pavlik Jr u. a. 2009]. Zunächst wird der Begriff Educational Data-Mining definiert. Durch die stetigen Fortschritte im Bereich Data-Mining wurden diese Erkenntnisse auf einen neuen Aufgabenbereich angewandt: Educational Data-Mining [C. Romero u. a. 2007]. Mit Lernumgebung sind Schulen, Universitäten, Hochschulen, Kindergärten, aber auch E-Learning-Systeme und web-basierte Bildungseinrichtungen gemeint. In Abbildung 2.22 wird der Zyklus von C. Romero

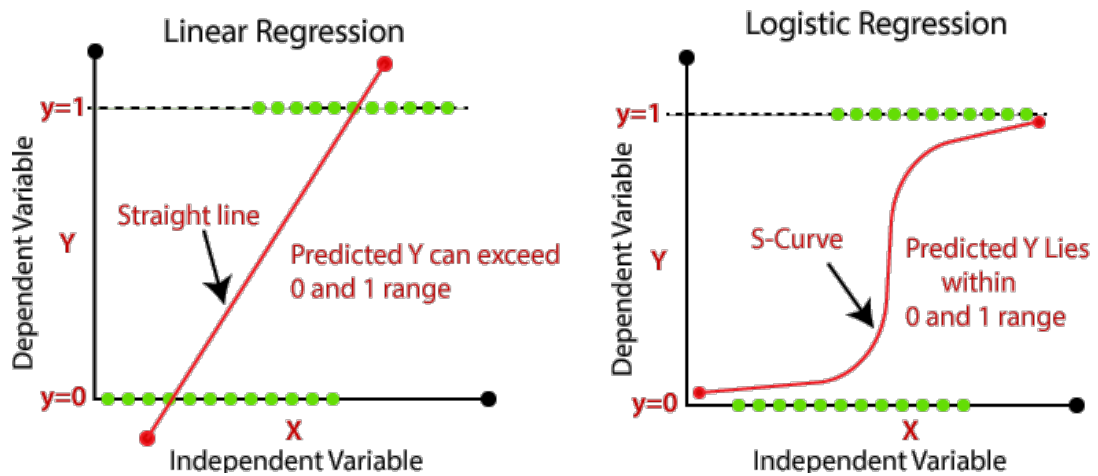


Abbildung 2.21: Verteilungsfunktionen der linearen und logistischen Regressionsmodelle [Regression 2023]

u. a. 2007 zur Anwendung von Data-Mining auf Lernumgebungen dargestellt. Durch die Interaktionen des Nutzers mit der Lernumgebung, oder durch die Nutzung des Systems, werden Daten akquiriert. Nach der Vorverarbeitung der vorliegenden Daten können Data-Mining Techniken, wie z.B. Clustering, Klassifikation, Ausreißererkenkung und viele mehr, auf den Datensatz angewendet werden. Die Ergebnisse führen anschließend zu neuen Erkenntnissen, welche dazu genutzt werden können, die Lernumgebung weiterhin zu fördern oder dem Nutzer Empfehlungen auszusprechen [C. Romero u. a. 2007].

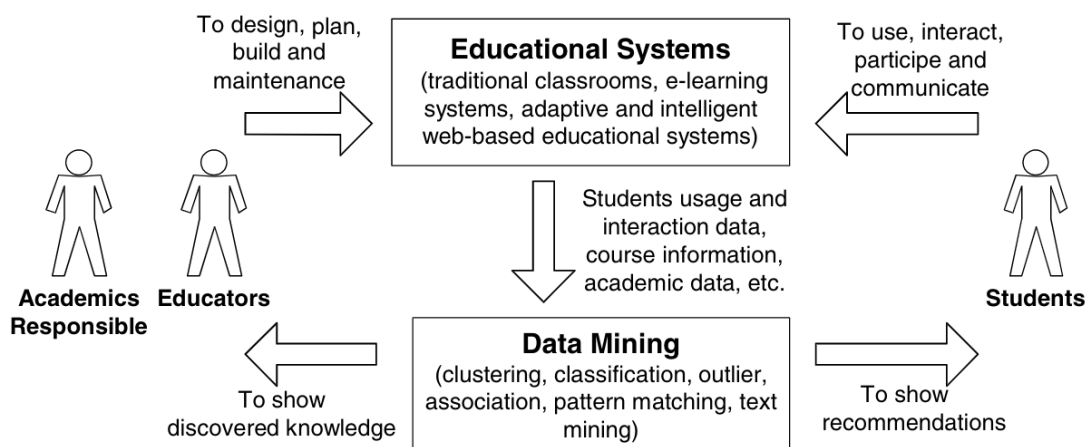


Abbildung 2.22: Zyklus zur Anwendung von Data-Mining auf Lernumgebungen [C. Romero u. a. 2007]

Learning Factor Analysis wird als Educational Data-Mining Methode eingesetzt, um kognitive Modelle und Daten zwischen Lernenden und Tutoren zu untersuchen. Kognitive Modelle sind entweder Beschreibungen oder berechnete Repräsentationen des menschlichen Denkens über bestimmte Konzepte oder Fähigkeiten [Lane 2012]. Nach [Cen, Kenneth R. Koedinger u. a. 2007] ist es möglich, mittels LFA die Schwierigkeit sowie die Lernrate von Wissenskomponenten zu bestimmen und eine Vorhersage für die Performance jeder einzelnen WK zu treffen. Folgt man [Pavlik Jr u. a. 2009] erfasst LFA drei wichtige Eigenschaften, welche den Lernprozess und die Performance des Nutzers beeinflussen:

- Fähigkeit der WK: Für jede WK wird in einem Parameter festgehalten, was die Fähigkeit des Nutzers ist.
- Leichtigkeit der einzelnen WK: Gibt den Schwierigkeitsgrad der WK an.
- Lernrate der einzelnen WK: Für jede WK wird mithilfe eines Parameters die Lernrate aufgenommen.

Berechnung der Learning Factor Analysis

In diesem Abschnitt erfolgt eine Beschreibung der klassischen LFA Berechnung (s. Formel 2.3) sowie eine Erklärung der einzelnen Parameter:

$$m(i, j \in KCs, n) = \alpha_i + \sum_{j \in KCs} (\beta_j + \gamma_j n_{i,j}) \quad (2.3)$$

- Logit-Wert m , der die akkumulierte Result-Performance für einen Nutzer i repräsentiert: $m \in \mathbb{R}^+$, dabei steht 0 für das fehlende Wissen des Nutzers bezüglich der WK. m kann für eine oder mehrere WK j berechnet werden.
- Nutzer i : ist eine eindeutige Kennzeichnung für einen Nutzer.
- Wissenskomponente j : Eine oder mehrere WKs, die ein Fach oder Thema repräsentieren, welches erlernt werden soll.
- Vorteil von häufigen vorherigen Abfragen n : Eine Funktion n für vorherige Beobachtungen eines Nutzers i mit WK j . $\gamma_{i,j}$ ist eine Gewichtsfunktion von n für jede WK.
- Fähigkeit oder auch Lernobjekt α : Für jeden Nutzer i die Fähigkeit, die eine oder mehrere WK enthalten kann.
- Schwierigkeitsgrad der WK $\beta \in [0,1]$: Gibt die Leichtigkeit der Wissenskomponente an.

Laut Becker u. a. 2016 beschreibt der Logit-Wert den natürlichen Logarithmus einer Wahrscheinlichkeit p durch eine Gegenwahrscheinlichkeit $1 - p$. Nachdem m berechnet wurde, kann mithilfe der Sigmoidfunktion, welche in Formel 2.4 dargestellt ist, die Wahrscheinlichkeit der jeweiligen Vorhersage der Analyse berechnet werden. Die Sigmoidfunktion, oder kurz S-Funktion, ist eine logistische Funktion mit einem S-förmigen Graphen (s. Abbildung 2.21). Die Funktion beschreibt eine stetig eindimensionale Wahrscheinlichkeitsverteilung und bietet eine Darstellung von Sättigungsprozessen. Dabei beschreibt e die Eulersche Zahl. Mithilfe dieser Funktion kann m in eine Vorhersage beobachteter Wahrscheinlichkeiten konvertiert werden [Han u. a. 1995].

$$p(m) = \frac{1}{1 + e^{-m}} \quad (2.4)$$

Vorteile und Einschränkungen der LFA

Wie bereits im Absatz 2.9.2 beschrieben, hat das KT Modell klare Einschränkungen, wenn es darum geht, den Lernfortschritt für Fähigkeiten mit diversen Wissenskomponente zu berechnen. Im Vergleich dazu wurde das LFA Modell eingehend als Lösung für das Problem von diversen WK innerhalb einer Fähigkeit untersucht [Leszczenski u. a. 2007]. Die Lösung für dieses Problem erscheint sehr nützlich, da viele adaptive Tutoring Systeme Antworten erwarten, in denen mehrere WK involviert sind. Eine große Herausforderung des LFA besteht in der eingeschränkten Fähigkeit des Modells, den Nutzern einzelne WK zuzuordnen. Das liegt daran, dass das Modell korrekte und inkorrekte Aussagen eines bestimmten Nutzers ignoriert [Pavlik Jr u. a. 2009]. Das Modell ist demnach nicht dazu in der Lage, eine $Perf^R$ für einen Nutzer zu berechnen und daher für die Anwendung im Rahmen dieser Arbeit ungeeignet. LFA ordnet lediglich jedem Nutzer die gleiche Performance zu. Jedoch bietet LFA ein gutes Fundament für die Performance Factor Analysis. Im nächsten Abschnitt wird das Learning Factor Analysis Modell angepasst, um das Modell adaptiv für jeden einzelnen Nutzer verwenden zu können.

2.9.5 Performance Factor Analysis

Die Performance Factor Analysis (PFA) ist nach [Pavlik Jr u. a. 2009] eine auf der Learning Factor Analysis basierende Methode, um die Result-Performance eines einzelnen Nutzers zu berechnen. Eine Stärke der LFA ist, dass die Berechnung des Lernerfolgs mehrere WKs betrachten kann. Aus diesem Grund bildet die Learning Factor Analysis die Basis für die Performance Factor Analysis. Ein großer Nachteil der LFA liegt darin, dass sich die Datenanalyse nicht auf einen Nutzer fokussiert, sondern alle vorhandenen Daten zur Berechnung verwendet und daraus eine durchschnittliche Performance für alle Nutzer berechnet. Daher wird das PFA

Modell um einen zusätzlichen Parameter erweitert. Dieser neue Parameter repräsentiert den Nutzer und ermöglicht es somit, dessen individuelle Performance zu berechnen. Pavlik Jr u. a. 2009 erklären, dass ein ausschlaggebender Unterschied zwischen den beiden Modellen die Empfindlichkeit der PFA gegenüber der Performance des Nutzers ist. Die Performance ist aus zwei Gründen besonders wichtig für das Lernverhalten des Nutzers. Erstens sind korrekte Antworten ein starkes Anzeichen dafür, dass das Wissen bereits sehr hoch ist. Dementsprechend ist diese Information alleine schon hilfreich, um den Wissensstand des Nutzers zu bestimmen. Zweitens führen korrekte Antworten zu einem besseren Lernfortschritt als falsche Antworten. Das kann unterschiedliche Gründe besitzen. Der Prozess einer richtigen Antwort wird zum einen besser verarbeitet und zum anderen ist die Überprüfung einer falschen Antwort üblicherweise weniger aufschlussreich. Es hat viele Vorteile, korrekte Antworten als Informationsquelle miteinzubeziehen, aber es mag auch sinnvoll erscheinen, fehlerhafte Antworten zu betrachten [Pavlik Jr u. a. 2009]. Die Sensibilisierung auf falsche Antworten kann einen weiteren Anhaltspunkt für das Lernverhalten bieten. Dabei repräsentiert der Misserfolg die Inverse des Erfolges. Gemeinsam können Erfolg und Misserfolg nicht nur die Quantität der Antworten, sondern auch eine Relation zwischen den beiden liefern [Pavlik Jr u. a. 2009]. Dieses Konzept ist in Gleichung 2.5 zu erkennen, da Erfolg und Misserfolg nun beide in die Formel integriert sind. Ein weiterer Unterschied lässt in der Gleichung 2.5 zu erkennen: Der Parameter α aus der LFA (s. Formel 2.3) ist weggefallen, da dieser Parameter in einer Situation, in der Adaptivität gelten muss, als schwer zu bestimmen gilt [Corbett u. a. 1994].

Pavlik Jr u. a. 2009 beschreiben PFA als eine vielversprechende Alternative zu dem bekannten Modell KT. Im Gegensatz zu KT kann PFA bei der Berechnung mehrere WKs gleichzeitig erfassen und somit kommen nun auch Fragen mit mehreren Komponenten zur Analyse in Betracht.

Berechnung der Performance Factor Analysis

Dem folgenden Abschnitt ist die klassische PFA Berechnung sowie die Erklärung der einzelnen Parameter nach Gong, Joseph E Beck u. a. 2010 zu entnehmen:

$$m(i, j \in KCs, s, f) = \sum_{j \in KCs} (\beta_j + \gamma_j s_{i,j} + \rho_j f_{i,j}) \quad (2.5)$$

Die Erklärung der einzelnen Variablen:

- m, i, j und β entsprechen den Parameter bei LFA.
- $s_{i,j}$: Vorherige Erfolge der WK j für Nutzer i ; $s_{i,j} \in \{0,1\}$ oder $s_{i,j} \in [0,1]$

- $f_{i,j}$: Vorherige Misserfolge der Wissenskomponente j für Nutzer i ; $f_{i,j} \in \{0,1\}$ oder $f_{i,j} \in [0,1]$
- γ, ρ : Skalierungsfaktoren für Erfolg $0 \leq \gamma \leq 1$ und Misserfolg $-1 \leq \rho \leq 0$ der WK (Gewichtsfunktionen)

$s_{i,j}$ und $f_{i,j}$ können verschiedene Werte annehmen, je nachdem, ob man das Beenden eines Lerninhalts binär als Erfolg bzw. Misserfolg betrachtet, oder, ob das Ergebnis stetig im Intervall $[0, 1]$ ist. Nachdem m berechnet wurde, kann genauso wie im Learning Factor Analysis Modell (s. Abschnitt 2.9.4) mithilfe der Sigmoidfunktion der Logit-Wert in die Wahrscheinlichkeitsverteilung der $Perf^R$ konvertiert werden [Gong, Joseph E Beck u. a. 2010].

Nachteile von PFA

Wie bereits aufgezeigt, werden bei der Performance Factor Analysis Erfolge und Misserfolge jedes einzelnen Lernobjektes bzw. jeder einzelnen Wissenskomponente festgehalten. Bei der Iteration über die WK wird die Reihenfolge der jeweiligen Antworten jedoch nicht berücksichtigt [Gong, Joseph E Beck u. a. 2010]. Zur Veranschaulichung ein Beispiel: Ein Nutzer soll eine Frage zum Satz des Pythagoras beantworten. Zuvor hat derselbe Nutzer vier Fragen zum selben Lernobjekt beantwortet. Von diesen vier Fragen wurden zwei richtig und zwei falsch beantwortet. Wie im Problem bereits beschrieben, ist die Reihenfolge der Antworten nicht einzusehen. Nun gibt es genau sechs Möglichkeiten, welche zu dieser Konstellation an Antworten führen. In der Performance-Berechnung sollte es einen Unterschied machen, ob ein Nutzer zuerst zweimal falsch und dann zweimal richtig, oder zuerst zweimal richtig und anschließend zweimal falsch antwortet. Beide Fälle bieten abweichende Erkenntnisse über den vorherigen Performance-Score des Nutzers [Pustovojtovskij 2021].

Für dieses Problem bietet [Gong, Joseph E Beck u. a. 2010] jedoch eine Antwort: Die Grundidee ist es, „Data Aging“ in die Berechnung von PFA einfließen zu lassen. Die modifizierte Berechnung basiert auf der Annahme, dass Antworten, welche weiter in der Vergangenheit liegen, weniger Einfluss auf den Performance-Score haben sollten. Im selben wissenschaftlichen Schreiben wird der Verfallfaktor („decay factor“) δ ($0 < \delta \leq 1$) vorgestellt. Mithilfe dieses Faktors werden Antworten, die weiter in der Vergangenheit liegen, mit absteigender Bedeutung versehen.

$$\begin{aligned} success_count_t &= \sum_{1 \leq k \leq t-1} P_k * \delta^{t-1-k} \\ failure_count_t &= \sum_{1 \leq k \leq t-1} |P_k - 1| * \delta^{t-1-k} \end{aligned} \quad (2.6)$$

Diese Anpassungen ersetzen $s_{i,j}$ und $f_{i,j}$ aus der klassischen PFA Formel 2.5. Dabei bezeichnet t den t -ten Lerninhalt, welchen er beantwortet wird. P_k repräsentiert den Erfolg des k -ten Lerninhalts.

Nachdem die grundlegenden Modelle und Methoden vorgestellt und analysiert wurden, zielt Kapitel 3 darauf ab, verschiedene Arbeiten hinsichtlich verwandter Problematiken zu präsentieren und diese gegenüber der vorliegenden Arbeit abzugrenzen. Insbesondere werden Varianten sowie verbesserte Konzepte der hier dargelegten Modelle Bayesian Knowledge Tracing und Performance Factor Analysis erklärt.

3 Stand der Forschung und Technik

In diesem Kapitel wird der aktuelle Stand der Forschung und Technik im Hinblick auf die Untersuchung von Performance-Messungen für adaptive Assistenzsysteme behandelt. Zunächst werden bereits vorhandene Publikationen und Forschungsarbeiten zum Thema adaptive Assistenzsysteme im Bereich von E-Learning Plattformen und Serious Games vorgestellt, um anschließend zu klären, weshalb diese Lösungsansätze die Problemstellung dieser Arbeit nicht vollständig lösen. Im Anschluss folgt ein Überblick über die unterschiedlichen Methoden der Performance-Messung und deren Bezug zur vorliegenden Arbeit.

3.1 Adaptive Assistenzsysteme und E-Learning

In Kapitel 2.2.1 wurden adaptive Assistenzsysteme bereits motiviert und erläutert. Im Folgenden werden wissenschaftliche Publikation zu dem Thema adaptive Assistenzsysteme im E-Learning-Bereich vorgestellt und von dem adaptiven Assistent dieser Arbeit abgegrenzt.

Einer der ersten web-basierten intelligenten Tutoring Systeme war, der in [Brusilovsky u. a. 1996], vorgestellte ELM Adaptive Remote Tutor (ELM-ART). ELM-ART ist ein System, das eine Kombination an zwei Paradigmen bietet: adaptive Assistenz und adaptive Hypermedia Technologie. Dieses System bietet dem Nutzenden ein interaktives und intelligentes web-basiertes Textbuch [Weber u. a. 2016]. Der Nutzende kann zu den Inhalten des Textbuches Fragen stellen und sich Beispiele anzeigen lassen. Das Ziel dieses Systems ist es, dem Nutzenden die Konzepte des Textbuches zu lehren, mit Möglichkeiten zum Nachfragen im Gegensatz zu einem starren Buch [Weber u. a. 2016]. Im Vergleich zu ELM-ART soll der in dieser Arbeit konzipierte adaptive Assistent keine inhaltlichen Fragen zu den Konzepten der E-Learning-Plattform beantworten, sondern den Nutzenden navigieren. Das Ziel dieser Arbeit ist es, dem Nutzenden den Lerninhalt darzustellen, der für ihn angemessen ist. Dabei heißt angemessen, dass sich der Nutzende in einem Zustand zwischen Über- und Unterforderung befindet.

Ein weiteres bekanntes Beispiel für ein adaptives Assistenzsystem ist „Assessment and Learning in Knowledge Spaces“ (ALEKS), das eines der meist eingesetzten Systeme im Bereich E-Learning in den USA ist [Fang u. a. 2019]. ALEKS ist eine kostenpflichtige E-Learning-Plattform mit Lerninhalten in den Bereichen, Mathematik, Chemie und Betriebswissenschaft

[Falmagne u. a. 2013]. Wie in Matayoshi u. a. 2021 genauer erläutert wird, kombiniert ALEKS die Ansätze eines rekurrenten neuronalen Netzes und der sogenannten „Knowledge Space Theory“. Ein rekurrentes neuronales Netz ist ein neuronales Netz, das sequenzielle Daten oder Zeitreihendaten verwendet, um zu lernen [IBM 2023]. Das Modell wird in Abschnitt 3.2 noch einmal genauer erläutert. Die „Knowledge Space Theory“ besagt, dass bestimmte Wissenskomponenten (Kapitel 2.9.1) nur gemeistert werden können, wenn vorher eine Untermenge anderer Komponenten gemeistert wurde [Falmagne u. a. 2013]. Aufgrund dieser Struktur ist es den Nutzenden von ALEKS nur möglich eine Lerneinheit zu üben, falls sie die Voraussetzungen für die benötigten WKs erfüllt wurden. In Abbildung 3.1 ist ein Beispiel für eine Aufgabe aus dem System ALEKS zu sehen.

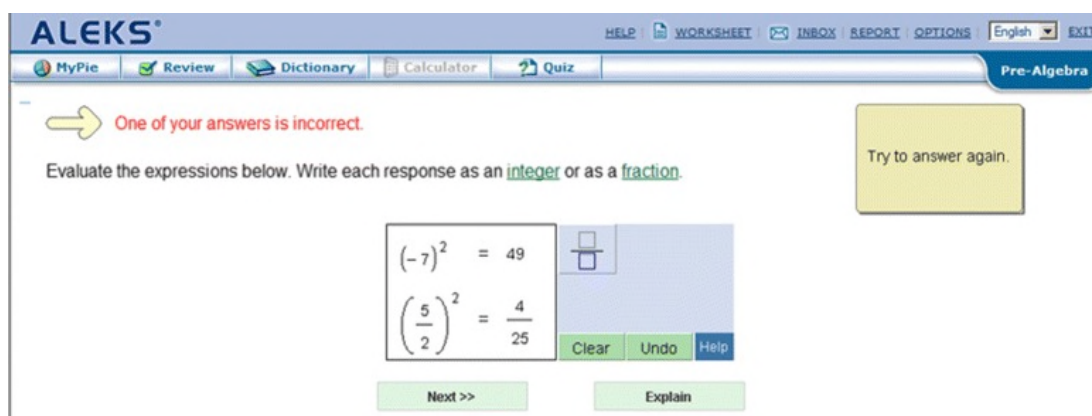


Abbildung 3.1: Aufgabe aus der Lernplattform ALEKS mit den Möglichkeiten, eine andere Aufgabe zu wählen, einen Hinweis zu erhalten oder sich die Aufgabe erklären zu lassen [Nye u. a. 2018].

In dieser Arbeit liegt der Fokus jedoch darauf, wie in Dörner u. a. 2016 beschrieben, die Performance des Nutzenden auf zwei Komponenten aufzuteilen. Die erste Komponente, Result-Performance, wird anhand einer oder mehrerer zusammenhängender Wissenskomponenten des Nutzenden berechnet und betrachtet den Erfolg bzw. Misserfolg dieser WK. Mithilfe der zweiten Komponente, Progress-Performance, soll ein sogenannter Nutzungspfad erstellt werden, mit dem man den Lernfortschritt des Nutzenden messen kann. Mit der Kombination dieser beiden Werte soll dem Nutzenden nicht nur nach jedem Kurs Adaptivität geboten werden, sondern nach Abschluss jeder einzelnen Lerneinheit. Um Adaptivität in E-Learning-System zu gewährleisten, muss die Performance des Nutzenden gemessen werden. Möglichkeiten, die $Perf^P$ zu berechnen, wurde bereit in Baur 2022 und Delaunay 2022 diskutiert und in Kapitel 2.8 vorgestellt. Berechnungen für die $Perf^R$ wurde bereits in Kapitel 2.9 präsentiert. Im Folgenden werden verschiedene Ansätze und Varianten der $Perf^R$ Berechnung dargelegt.

3.2 Lernfortschrittsmodellierung bei adaptiven Assistenzsystemen

Im folgenden Abschnitt werden die wissenschaftlichen Artikel der Result-Performance aus Kapitel 2.9.2 und 2.9.5 verglichen und zu dem in dieser Arbeit entstandenen Modell abgetrennt. Am Ende dieses Abschnitts wird in Abbildung 3.7 ein zeitlicher Verlauf der hier erwähnten Modelle gezeigt. In der darauffolgenden Tabelle 3.2 werden die für diese Arbeit relevantesten Bestandteile der Modelle verglichen und in Relation zu GPFA gesetzt.

Knowledge Tracing

Knowledge Tracing ist eine essenzielle Methode, um den Wissensstand jedes Nutzers einzeln zu ermitteln und damit die Performance für zukünftige Aufgaben vorherzusagen. Dabei stellt das Bayesian Knowledge Tracing (BKT) eines der bekanntesten Methoden [Liu, Huang u. a. 2021] dar. Bei diesem Modell wird angenommen, dass jeder Nutzende über eine Menge Wissenskomponenten verfügt, welche sich aus einer Anzahl binärer Variablen zusammensetzen. Jede Variable repräsentiert dabei, ob ein Nutzer ein konkretes Teilproblem bzw. Konzept gemeistert hat oder nicht. Im Allgemeinen verwendet BKT ein Hidden Markov Modell (HMM), um den Wissensstand jedes Nutzenden zu aktualisieren und anschließend die Performance des Nutzers auszugeben. In Abbildung 3.2 ist dieses Modell zu sehen. Auf Basis dieses Modells wurden weitere Annahmen über die Zusammensetzung der Aufgaben und den Ablauf des Lernprozesses gemacht. Beispielsweise wurde weiterhin angenommen, dass ein Nutzer eine korrekte Antwort auf eine Frage geben kann, obwohl er das Wissen für diese WK nicht besitzt (Guess Parameter) oder eine falsche Antwort gibt, obwohl er die WK verstanden hat (Slip Parameter). Eine weitere Annahme war, dass die gestellten Aufgaben unterschiedliche Schwierigkeitsstufen besitzen. Diese Annahmen wurden in dem Modell integriert, um die Performance möglichst präzise berechnen zu können [Gong, Joseph E Beck u. a. 2010].

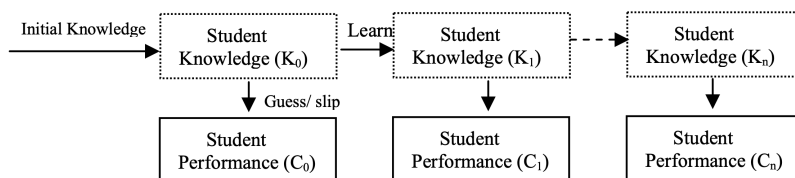


Abbildung 3.2: Das Knowledge Tracing Modell [Gong, Joseph E Beck u. a. 2010]

Beim Ermitteln der genannten Parameter lassen sich zwei Nachteile hinsichtlich des Knowledge Tracing Modell feststellen: lokale und mehrere globale Maxima. In Bezug auf Fehlerlo-

berflächen stellen lokale Maxima ein übliches Problem dar, für das jedoch bereits mehrere Lösungsansätze bestehen, wie beispielsweise das mehrfache Neustarten [Gong, Joseph E Beck u. a. 2010]. Der zweite Nachteil besteht im sogenannten „Identifiability“-Problem. Das bedeutet, dass dasselbe Modell, mit denselben Daten, mehrere unterschiedliche Mengen an Parametern ausgibt, welche gleich gut zu den Daten passen. Es gibt keine statistischen Methoden, um eine Präferenz in den Parametermengen zu erkennen. Dies bedeutet wiederum, dass eine Anpassung des Modells zusätzlich eine Anpassung der Parameterermittlung zur Folge hat [Gong, Joseph E Beck u. a. 2010]. Im Gegensatz zu KT bestehen für die auf der logistischen Regression basierenden Verfahren Learning Factor Analysis und Performance Factor Analysis keine Schwierigkeiten beim Erreichen von globalen Maxima [Gong, Joseph E Beck u. a. 2010].

Eine weitere Schwachstelle im KT Modell zeigt sich, wenn die Fragen mehrere WK gleichzeitig erfassen. Ein Beispiel für solch eine Frage wäre Aufgabe fünf in Abbildung 2.19, bei der sowohl Funktionen als auch Ungleichungen abgefragt werden. Das Problem hierbei besteht darin, dass KT lediglich die Historie einer Wissenskomponente betrachtet [Gong, Joseph E Beck u. a. 2010]. Für dieses Problem zeigen Gong, Joseph E Beck u. a. 2010 verschiedene Lösungsansätze auf, wie beispielsweise das Auftrennen von Fragen mit mehreren WK in mehrere gleiche Fragen mit unterschiedlichen WKs. Auf diese Ansätze wird in Abschnitt 2.9.2 konkreter eingegangen. Für PFA gibt es keine Schwierigkeiten für Fragen mit mehreren WK, da in diesem Modell mehrere Komponenten auch auf natürliche Art und Weise betrachtet werden können.

Lernen stellt einen komplexen Prozess dar und der Lernerfolg kann nicht in unabhängige Wissenskomponenten unterteilt werden [Liu, Huang u. a. 2021]. Daher wurde Methoden und Verbesserungen in Betracht gezogen, bei denen die WK in Beziehungen zueinander stehen. Mit den bemerkenswerten Fortschritten der Deep Learning Techniken in vielen unterschiedlichen Bereichen, wie z.B. der Spracherkennung, der Bildgenerierung, dem Natural Language Processing (NLP) und vielen mehr, gab es bereits Versuche Deep Learning Modelle auf die Lernfortschrittsmodellierung im Bereich des E-Learning zu transferieren [Liu, Huang u. a. 2021]. In der Arbeit von Piech u. a. 2015 zu Deep Knowledge Tracing (DKT) wurden rekurrente neuronale Netze (RNN) und Long Short-Term Memory Networks (LSTM) verwendet, um den Lernprozess des Nutzens zu modellieren und damit eine Performance vorherzusagen. Der Aufbau eines solchen RNN für KT lässt sich der Abbildung 3.3 entnehmen.

Ein Problem des Deep Knowledge Tracing Ansatzes ist die hohe Modellkomplexität, welche für Schwierigkeiten bei der Skalierung sorgt. Zusätzlich dazu gewährt DKT aufgrund der hohen Komplexität einen sehr geringen Einblick in die gelernten Konzepte des Modells. Dadurch wird aus dem Modell nicht ersichtlich, welche Wissenskomponenten vom Nutzenden verstanden wurden oder welche ihm noch unbekannt sind [S. Pu u. a. 2021].

Sowohl Bayesian Knowledge Tracing als auch die Deep Learning Ansätze, wie z.B. DKT,

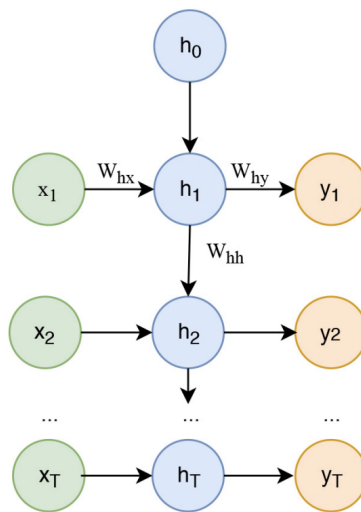
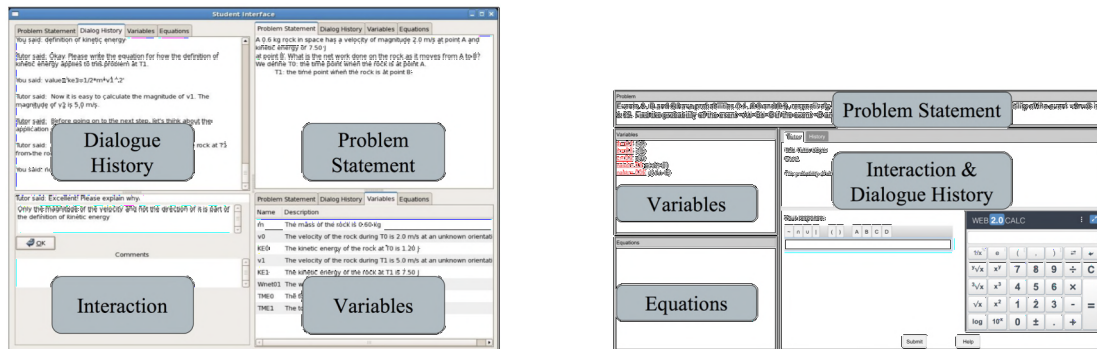


Abbildung 3.3: Das Deep Knowledge Tracing Modell als RNN, wobei x_t die Eingabe darstellt und y_t die Wahrscheinlichkeit, die Frage aus der Eingabe korrekt zu beantworten [Sapountzi u. a. 2019].

wurden in zwei E-Learning Systemen angewendet (Abbildung 3.4). Die erste Anwendung ist ein Natural Language Tutoring System, namens Cordillera, welches das Ziel verfolgt, dem Nutzenden, Physik beizubringen. Dabei besteht Cordillera aus insgesamt sieben WKs. Beim zweiten System handelt es sich um das intelligente Tutoring System Pyrenees, welches Wahrscheinlichkeitstheorie unterrichtet. Dieses besteht aus zehn WK zum Thema Wahrscheinlichkeiten, die unter anderem das Additionstheorem und das Bayes Theorem abdecken [Mao 2018].

Um weitere Einblicke sowie ein tiefergehendes Verständnis für die Beziehungen zwischen Aufgaben und Wissenskomponenten zu erhalten, wurde von Zhang u. a. 2017 ein Dynamic Key-Value Memory Networks für Knowledge Tracing (DKVMN KT) vorgestellt. Bei DKVMN KT werden die Beziehungen zwischen den zugrundeliegenden WKs ausgewertet und direkt als Können jeder einzelnen WK ausgegeben [Zhang u. a. 2017]. Um diese Beziehungen zwischen den Komponenten zu speichern, wurden sogenannte „Memory“-Matrizen eingeführt. Der Nachteil an diesem Modell äußert sich wiederum in der hohen Komplexität des Modells und der damit einhergehenden Rechenleistung.

Um die natürlichen Gegebenheiten der Beziehungen zwischen den Komponenten zu nutzen, wurde von Nakagawa u. a. 2019 das Graph Knowledge Tracing (GKT) Modell vorgestellt. Dabei werden die unterschiedlichen Komponenten als Knoten repräsentiert und die Beziehungen zwischen den Komponenten als Kanten. Eine solche Umformulierung des Knowledge Tracing Ansatzes ermöglicht, anhand der Graphstruktur, die unterschiedlichen Komponenten auch



Ein Natural Language Tutoring System, namens Cordillera, das darauf spezifiziert ist, Physik zu lehren.

Ein intelligentes Tutoring System, namens Pyrenees, welches das Thema Wahrscheinlichkeit unterrichtet.

Abbildung 3.4: Zwei Tutoring Systeme, die in Mao 2018 genutzt wurden, um BKT mit anderen Deep Learning Ansätzen zu vergleichen.

in eine zeitliche Relation zu setzen [Nakagawa u. a. 2019]. Dieses Modell nutzt die gewonnenen Erkenntnisse aus dem wissenschaftlichen Schreiben von Battaglia u. a. 2018 über Graph Neural Networks und die Bedeutung von Beziehungen hinsichtlich des Verständnisses für den Menschen. Dieses Konzept wurde in der Abschlussarbeit von Delaunay 2022 auf das Quizlernspiel Spot-X konzeptionell angewendet (Abbildung 3.5).

Die vorgestellten Modelle zeigen gegenüber den klassischen Verfahren, wie Bayesian Knowledge Tracing und Performance Factor Analysis einen überlegenen Fortschritt und haben damit einen großen Erfolg erzielt [S. Pu u. a. 2021]. In Abbildung 3.7 werden die vorherigen Entwicklungen und Varianten von BKT in einem Zeitabschnitt dargestellt. In dieser Arbeit wird ein Modell untersucht, das auf PFA basiert. Der Grund hierfür besteht darin, dass PFA gegenüber BKT einige Vorteile besitzt. Einer davon ist, dass PFA im Gegensatz zu KT keine Probleme mit Multi-Skill Fragen hat. Dies bedeutet, dass PFA auch den Wissensstand für Fragen untersuchen kann, die mehrere WKs umfassen. Um auch die fortgeschritteneren Methoden vorstellen zu können, werden im folgenden Abschnitt die PFA und die darauf basierende Learning Factor Analysis vorgestellt.

Performance Factor Analysis

In dieser Arbeit liegt der Fokus auf den Learning Analytics, besonders auf der Performance Factor Analysis (PFA). Um nachvollziehen zu können, wie das Performance Factor Analysis (PFA) Modell aus dem Artikel von Pavlik Jr u. a. 2009 entstanden ist und weshalb es in wissen-

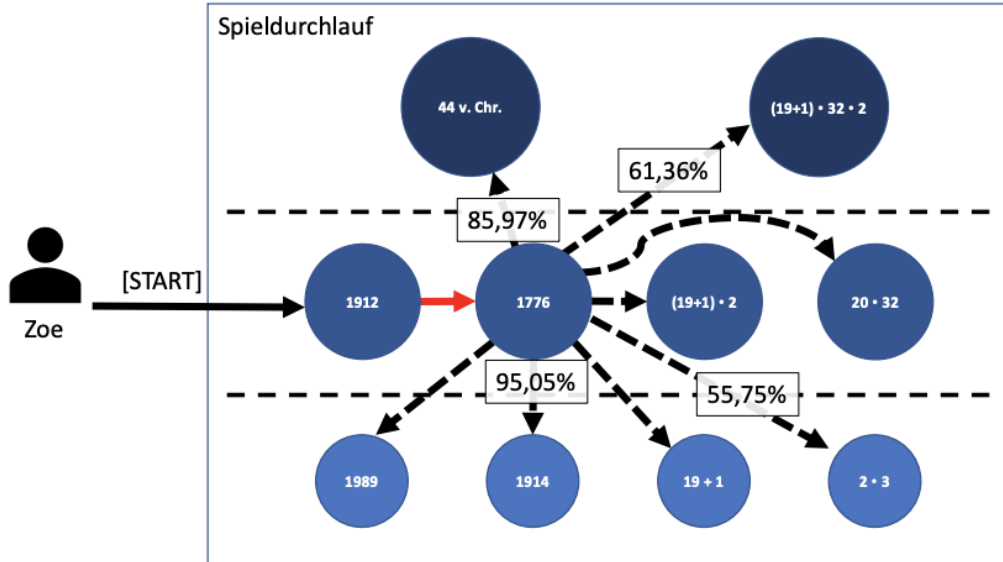
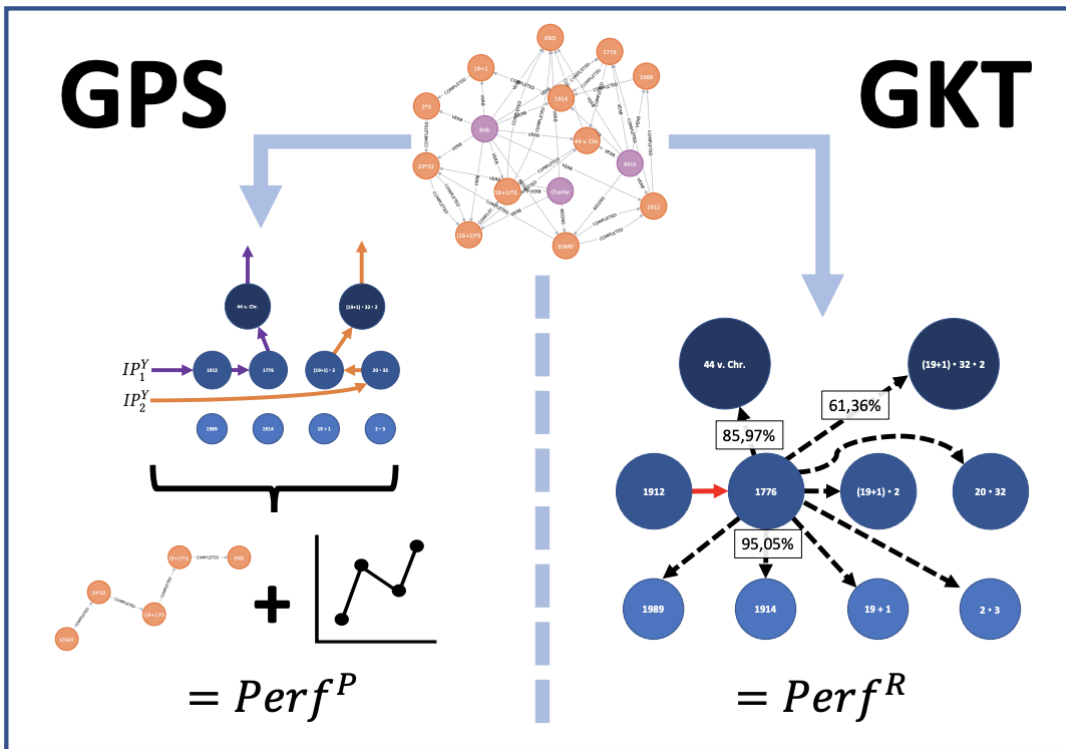


Abbildung 3.5: Oben: Anwendung von GKT, um die Performance eines Nutzers zu berechnen. Unten: Eine konkrete Performance-Vorhersage für den Nutzenden Zoe im Quizlernspiel Spot-X [Delaunay 2022].

schaftlichen Arbeiten, wie beispielsweise von Gong, Joseph E Beck u. a. 2010 sowie Pavlik Jr u. a. 2009, genauer untersucht wurde, erläutert Kapitel 2.5 das Modell im Detail.

Wie bereits im vorherigen Abschnitt erwähnt, ist es mit KT möglich, den aktuellen Wissensstand eines konkreten Nutzers einer bestimmten WK zu berechnen. Falls die gestellte Aufgabe jedoch mehrere WKs gleichzeitig abfragt, kann der Lernfortschritt mit KT nicht mehr angezeigt oder angepasst werden. Da die meisten fortgeschrittenen adaptiven Lernspiele mehr als nur eine WK gleichzeitig prüfen, scheint das KT für moderne SG nicht optimal zu sein [Pavlik Jr u. a. 2009].

Aus dieser Problematik ist die Idee für die alternative PFA entstanden. PFA basiert auf dem logistischen Regressionsmodell Learning Factor Analysis (LFA). LFA ist ein Educational Data-Mining Ansatz, der für unterschiedliche Zwecke eingesetzt wird. Beispielsweise wird LFA als Teil eines Suchalgorithmus verwendet, um Lerninhalte in Wissenskomponenten zu unterteilen. Dabei sollen die WKs die Daten bestmöglich repräsentieren und die Modellkomplexität möglichst einfach gehalten werden. Das Problem der Multi-Skill Fragen, das bei KT auftritt, wird bei LFA durch die Erfassung mehrerer WK gleichzeitig gelöst [Pavlik Jr u. a. 2009].

Das Problem bei LFA ist jedoch, dass die gespeicherten Daten keinem konkreten Nutzer zugewiesen werden können. Wie in Pavlik Jr u. a. 2009 bereits ausgearbeitet, ist das LFA Modell für adaptive Tutoring Systeme nicht geeignet, da die Analyse der Daten nicht an einen bestimmten Nutzer angepasst wird, sondern nur für den Durchschnitt aller Nutzer gilt.

Da sich LFA für die Analyse von Datensätzen anbietet und die Möglichkeit besteht, mehrere WKs einzubeziehen, wird LFA als Basis für das in Pavlik Jr u. a. 2009 neu entwickelte Modell verwendet. In dieser Arbeit LFA zu Performance Factor Analysis (PFA) umkonfiguriert. Das PFA Modell soll die Daten, wie in LFA, analysieren und die Erkenntnisse daraus an einen konkreten Nutzer binden [Pavlik Jr u. a. 2009]. In Abschnitt 2.5 wird das gesamte PFA Modell genauer erklärt.

In Pavlik Jr u. a. 2009 und Gong, Joseph E Beck u. a. 2010 werden die beiden Modelle BKT und PFA in mehreren Projekten miteinander verglichen. Beide wissenschaftlichen Artikel kommen zu dem Fazit, dass in diesen konkreten Umgebungen PFA besser abschneidet als BKT und damit besser für adaptive Tutoring Systeme geeignet ist [Gong, Joseph E Beck u. a. 2010][Pavlik Jr u. a. 2009]. In der Abschlussarbeit von Pustovojtovskij 2021 wurde das PFA Modell untersucht und für ein Serious Games modelliert und implementiert (Abbildung 3.6).

Ähnlich wie es beim KT Modell Fortschritte der Deep Learning Techniken gab, wurden auch Deep Learning Ansätze auf die Performance Factor Analysis angewendet. In der Arbeit von S. Pu u. a. 2021 wird das Deep Performance Factor Analysis (DPFA) Modell vorgestellt, welches ein Feed-Forward Neuronal Network verwendet, um den aktuellen Wissensstand zu berechnen. Wie in Tabelle 3.1 zu sehen ist, übertrifft DPFA sowohl PFA als auch DKT und liefert



Abbildung 3.6: Ein Beispiel für die Anwendung von PFA zur Berechnung des Lernfortschritts des Nutzenden: Eine Aufgabe aus dem Serious Games Streamlined Lost Earth 2307. Aufgrund der Performance Berechnung erhält der Nutzende einen Hinweis vom adaptiven Assistenten LISA [Pustovojtovskij 2021].

vergleichbare Ergebnisse wie DKVMN KT auf weit verbreiteten öffentlichen Datensätzen.

Datasets	Statistics					Average Test AUC			
	Attempts	Stu	Items	Skills	BKT	PFA	DKT	DKVMN	DPFA
Syn-5	20k	4,000	50	5	N/A	N/A	0.8239	0.8267	0.8348
NIPS	15.8M	119K	27K	1078	0.6890	0.7249	0.7771	0.7973	0.7965
ASSIST	392.8K	1,709	4,117	102	0.7091	0.6624	0.7317	0.7958	0.7964
STAT	135.3K	316	987	279	0.7298	0.7435	0.7928	0.8046	0.8061

Tabelle 3.1: Diese Tabelle vergleicht die Performance der Modelle DPFA mit den anderen klassischen und Deep Learning Verfahren. AUC steht für Fläche unter der ROC-Kurve, wobei ROC-Kurve die Leistung eines Klassifizierungsmodells für alle Klassifizierungsschwellenwerte zeigt [S. Pu u. a. 2021].

Ähnlich wie bei den anderen Deep Learning Methoden hat auch DPFA durch die hohe Modellkomplexität Schwierigkeiten mit der Skalierbarkeit und Anwendung auf kleineren Datensätze [S. Pu u. a. 2021]. Eine mögliche Lösung für dieses Problem stellen Graph-basierte Ansätze, wie z.B. Graph Knowledge Tracing, bereit. Zum Beispiel bietet der Ansatz von Nakagawa u. a. 2019 sowohl eine bessere Vorhersage zur Performance, als auch eine bessere Interpretierbarkeit der Vorhersagen im Vergleich zu den anderen Methoden.

In der vorliegenden Arbeit wird eine mögliche weitere Umsetzung für das Performance Factor Analysis Modell vorgestellt. Dieses Modell soll den Graph-basierten Ansatz aus dem GKT auf PFA anwenden. Im Rahmen dieser Arbeit wird das Modell Graph Performance Factor Analysis (GPFA) genannt. Dieses Modell kann generisch für unterschiedliche Assistenzsysteme verwendet werden, um einen sogenannten „Performance Score“ zu berechnen. Weiterhin wird in der Implementierung dieser Arbeit ein intelligenter Tutor erweitert, der anhand des berechneten Wertes mit den Nutzenden über das SG Interface interagieren kann. Im folgenden Kapitel werden die entwickelten Konzepte für die Modellierung sowie die Implementierung vorgestellt.

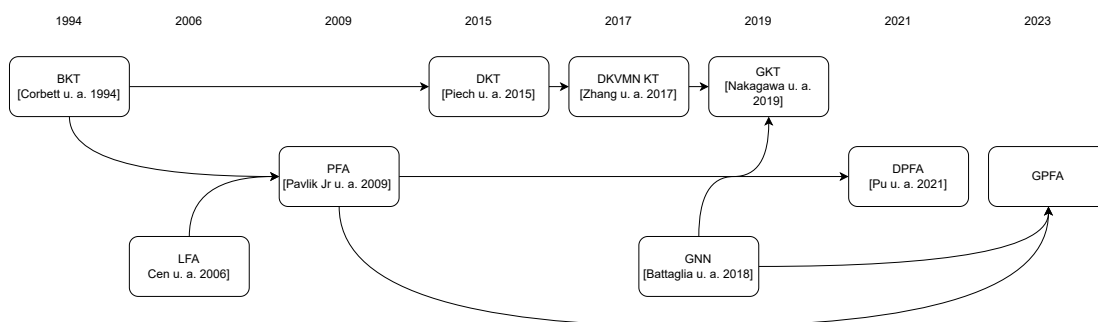


Abbildung 3.7: Diese Abbildung zeigt den zeitlichen Verlauf und die Entwicklung der entstandenen Modelle. Ein Pfeil von Modell A zu Modell B zeigt, dass Modell B basierend auf Modell A gebaut oder Modell B von Modell A motiviert wurde.

Modelle zur Verbesserung der Lernfortschrittsmodellierung						
Modelle	Educational Data-Mining EDM	Learning Analytics LA	Graph-Based	Artificial Neuronal Network ANN	Multi-Skill Question	Trancparency
BKT [Corbett u. a. 1994]		X				X
DKT [Piech u. a. 2015]		X		X		
DKVMN KT [Zhang u. a. 2017]		X		X		X
GKT [Nakagawa u. a. 2019]		X	X	X		
LFA [Cen, K. Koedinger u. a. 2006]	X				X	X
PFA [Pavlik Jr u. a. 2009]		X			X	X
DPFA [S. Pu u. a. 2021]		X		X	X	
GPFA		X	X		X	X

Tabelle 3.2: Diese Tabelle dient als Überblick für die Haupteigenschaften der Modelle zur Verbesserung der Lernfortschrittsmodellierung. Im Vergleich zu den anderen Modellen werden die Eigenschaften des Modells, das in der vorliegenden Arbeit konzipiert wurde, dargestellt.

4 Konzept für Performance-Messungen für adaptive Assistenzsysteme

In diesem Kapitel wird die Operationalisierung für Performance-Messungen für adaptive Assistenzsysteme vorbereitet. Die erarbeiteten Konzepte dieser Thesis lassen sich in zwei wesentliche Erkenntnisse untergliedern. Die erste Erkenntnis stellt der sogenannte „Graph Performance Service“ (GPS) dar, welcher bereits von Baur 2022 und Delaunay 2022 entwickelt wurde. Dabei handelt es sich um ein Performance-Modell, das dafür zuständig ist, mithilfe von vorliegenden Nutzerdaten einen „Performance-Score“ zu berechnen. Baur 2022 und Delaunay 2022 haben die Performance mittels der in Kapitel 2.8 erklärten Progress-Performance gemessen. Dieser Service soll im Rahmen dieser Thesis um eine kombinierte Performance, bestehend aus Progress-Performance und Result-Performance, erweitert werden. Die kombinierte Performance wurde gemeinsam mit Alexander Streicher konzipiert und basiert auf der Definition von Performance nach Wiemeyer u. a. 2016. Der zweite Aufschluss dieser Arbeit ist das generische Performance-Modell, welche für Assistenzsysteme allgemein eingesetzt werden kann, unter der Voraussetzung, dass diese Systeme xAPI-basierende Beobachtungsdaten verwenden. Durch Interaktionen mit dem Nutzer entstehen hierfür in einem Learning Management System (LMS) (z.B. Moodle, ILIAS) oder einem SG (z.B. Spot-X) Daten in xAPI-Format. Das Modell hat Zugriff auf diese Daten, welche in einem Learning Record Store (LRS) hinterlegt sind und berechnet daraus eine kombinierte Performance. Der berechnete Score wird anschließend zurück an das Assistenzsystem übermittelt und verwendet, um adaptive Anpassungen zu ermöglichen.

In Abbildung 4.1 stellt den Entwurf dar, welcher auf Basis dieser Arbeit entstanden ist. Der Nutzer bedient ein LMS oder ein SG, beispielsweise Moodle oder Spot-X. Aus Nutzerinteraktionen, wie z.B. Kurse bearbeiten oder Aufgaben spielen, werden innerhalb der verwendeten Plattform xAPI Statements (s. Abschnitt 2.6) generiert, die im Learning Record Store gespeichert werden. Anhand der vorliegenden xAPI Statements wird ein JSON Format mit Informationen, wie z.B. $Perf^P$, $Perf^R$ und kombinierter Performance des Nutzers, erstellt. Das JSON Objekt wird vom adaptiven Assistenten über eine HTTP Anfrage angefragt und anhand des „Performance Scores“ entscheidet sich der Assistent für eine adaptive Anpassung des Systems. Dies kann beispielsweise in Form einer Anpassung des Schwierigkeitsgrads oder eines angezeigten

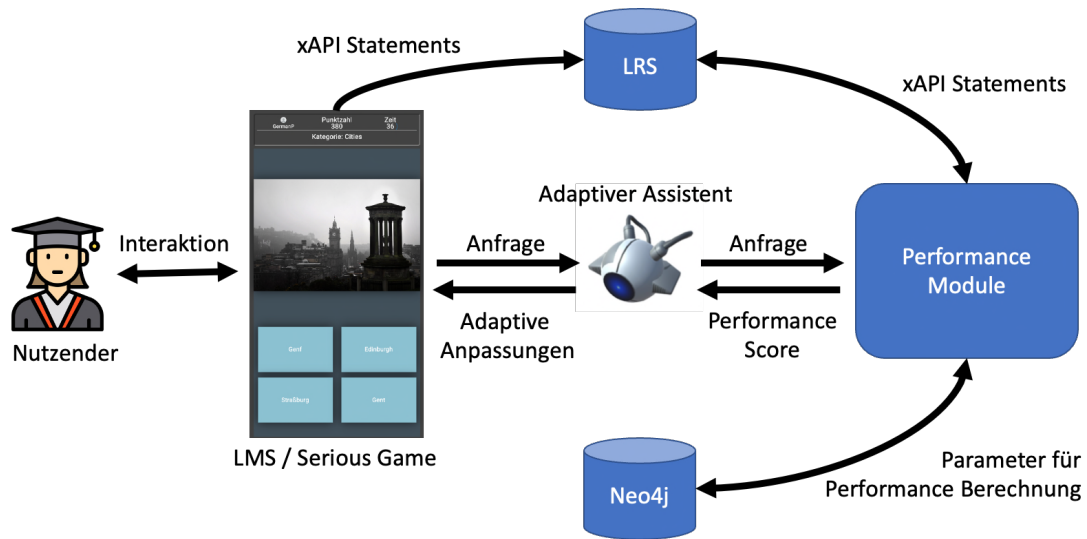


Abbildung 4.1: Architekturdiagramm der Kommunikation des Nutzers mit dem System und die Nutzung des Performance-Modells

Hinweises sein.

4.1 Erweiterung des GPS um eine kombinierte Performance

Im Rahmen dieser Arbeit wird der GPS um eine kombinierte Performance erweitert. Die Basis hierfür bildet die Arbeit von Baur 2022 und Delaunay 2022, welche um weitere Funktionalitäten erweitert wird. Im ersten Schritt erfolgt die Realisierung der kombinierten Performance im GPS. Die Grundidee dieser Berechnung stammt aus Wiemeyer u. a. 2016, welche die Performance in zwei Komponenten unterteilen: Progress und Result. Die Formel für die kombinierte Performance, welche in dieser Thesis untersucht und operationalisiert wird, lautet wie folgt:

$$p_u(e) = \alpha_1 Perf_u^P(e) + \alpha_2 Perf_u^R(e) \quad (4.1)$$

Dabei stellt $p_u(e)$ die kombinierte Performance für einen Nutzer u sowie den Zustandsübergang e mit $p_u \in [0, 1]$ dar. Zudem ist $Perf_u^P$ die Progress-Performance und $Perf_u^R$ die Result-Performance für einen Nutzer u . Die Parameter $\alpha_1, \alpha_2 \in [0, 1]$ sind Skalierungsfaktoren für die beiden Performance-Messungen. Da in der Literatur keine weiteren Anmerkungen hinsichtlich der Relationen beider Performance-Werte ersichtlich sind, wurde sich auf $\alpha_1, \alpha_2 = 0.5$ geeinigt.

Die Performance-Messung $Perf^P$ wurde bereits von Baur 2022 und Delaunay 2022, basierend auf den Ideal Paths Models (IPMs), modelliert und implementiert. Jedoch mussten

einige Anpassungen getroffen werden, welche in Abschnitt 5.3 behandelt werden. Der Fokus dieser Thesis liegt auf der Modellierung und Implementierung von $Perf^R$. Aufgrund der Modellierung von $Perf^P$ existiert bereits ein Graph, der den Nutzer, die Aktivitäten sowie die Aktivitätsübergänge darstellt. Die Grundidee war es, den bereits existierenden Graphen zu verwenden, um eine $Perf^R$ basierend auf PFA zu berechnen. Zusätzlich verwendet dieses Modell die anderen Vorteile der Graphstruktur, wie z.B. eine bessere Nutzung der Beziehungsgegebenheiten zwischen den Komponenten, sowie die Interpretierbarkeit der Vorhersagen (s. Abschnitt 3.2). Das in dieser Arbeit entstandene Modell der PFA für Graphen wird im Folgenden Graph Performance Factor Analysis (GPFA) genannt. Der nächste Abschnitt widmet sich der konzeptionellen Darstellung des GPFA.

4.2 Graph Performance Factor Analysis (GPFA)

Sei G ein gerichteter Graph mit der Knotenmenge V und der Kantenmenge E , sodass $G = (V, E)$ gilt. Weiterhin gelte $V = \{U, W\} \cup \{v_{start}, v_{end}\}$, wobei $U = \{u_1, u_2, \dots, u_n\}$ die Menge der Nutzer für alle $n \in \mathbb{N}$ und $W = \{W_1, W_2, \dots, W_m\}$ die Menge der WKs für alle $m \in \mathbb{N}$ sei. Für jedes $W_i \in W$ gilt, $W_i = \{w_1, w_2, \dots, w_k\}$ aus $k \in \mathbb{N}$ Aktivitäten w_j besteht. Im Folgenden versteht man $w_{m,k}$ als die Aktivität k der Wissenskomponente m . Ein Knoten $u \in U$ wird in dieser Arbeit als Nutzerknoten und ein Knoten $w \in W$ als Aktivitätsknoten bezeichnet. Zusätzlich gelte, dass jeder Nutzer einen Nutzerpfad G_u hat, der eine Menge an Kanten E_u besitzt, formal bedeutet das: $E = E_1 \cup E_2 \cup \dots \cup E_n$ mit $E_u = \{e_1, e_2, \dots, e_k\}$. Im Folgenden versteht man e_{u,w_k} als die ausgehende Kante aus der Aktivität w_k von Nutzer u . Eine Kante markiert immer einen Zustandsübergang. Zum Beispiel kennzeichnet die Kante $e = (u, v_{start})$, dass der Nutzer u ein System, wie z.B. das Quizlernspiel Spot-X, gestartet hat. Zusätzlich zu den Zustandsübergängen gibt es Aktivitätsübergänge. Das sind Kanten (w_k, w_m) , welche von einem Aktivitätsknoten $w_k \in W_k$ zu einem anderen Aktivitätsknoten $w_m \in W_m$ gehen, für $W_n, W_m \in W$. In diesem Aktivitätsübergang wird gespeichert, ob es sich bei der ausgehenden Aktivität um einen Erfolg s oder Misserfolg f bei einem Nutzer u handelt. Eine Übersicht der Zuweisung von Beobachtungsdaten im xAPI-Format zu Grapheigenschaften lässt sich der folgenden Tabelle entnehmen:

xAPI Parameter	Grapheigenschaften
Actor	Nutzerknoten u
Verb	Kante als Zustandsübergang
Object	Aktivitätsknoten mit WK als zusätzliche Eigenschaft
Result	Erfolg oder Misserfolg als Eigenschaft der Kante

Bevor der Entwurf des Performance-Modells vorgestellt wird, erfolgt die Darstellung einer Modellierung von GPFA. Wie bereits in Pustovojtovskij 2021 erläutert wurde, bedarf es für die Berechnung von PFA bzw. GPFA einiger Strukturen, z.B. WKs sowie Lernobjekte. Um diese Strukturen zu modellieren, ist ein charakterisierendes Ziel des adaptiven Assistenten (s. Abschnitt 2.2.2) erforderlich. In dieser Arbeit handelt es sich bei jenem Ziel um die Wissenslehre. Für die Modelle zur Berechnung der $Perf^R$ muss das charakterisierende Ziel noch weiter in Fähigkeiten („Skills“) unterteilt werden, welche auch Lernobjekte genannt werden. Diese Lernobjekte erhält man, indem man sie mithilfe von didaktischen Mitteln ausarbeitet. Meistens ist das Aufgabe eines Lehrers in der Schule, eines Entwicklers bei einem Spiel oder eines Dozenten bei einer Vorlesung und nicht Bestandteil oder Fokus dieser Arbeit. Um nun das Modell für die GPFA zu vervollständigen, werden den Lernobjekten WKs zugewiesen. Dabei kann ein Lernobjekt eine oder mehrere WKs besitzen. Ähnlich wie für die Lernobjekte, werden die WK über didaktische Mittel erarbeitet. In Abbildung 4.2 ist ein Graph mit den oben genannten Eigenschaften, sowie einer Aufteilung der WKs und Lernobjekte zu sehen. In orange ist der Nutzerknoten für einen Nutzer „Bob“ abgebildet. Bob schließt drei Aktivitätsknoten $T1$, $T2$ und $T3$ ab und beendet anschließend die Lernobjekte. Die Kante beinhaltet den Erfolg bzw. Misserfolg der Aktivität des ausgehenden Aktivitätsknotens, sowie die ausgeführte Aktivität. Beispielsweise umfasst die Kante $(T1, T2)$ den Misserfolg mit einem $Score = 0$ der Aktivität $T1$. Die unterschiedlichen WKs sind in grün und grau gekennzeichnet. $T1$ und $T3$ bilden dabei die WK „Geschichte“ und $T2$ die WK „Algebra“. Ein Aktivitätsknoten kann dabei mehrere WKs aufweisen.

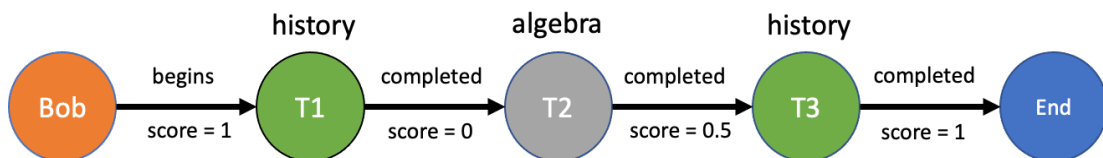


Abbildung 4.2: Ein Beispiel für einen konstruierten Graphen

Nachdem die Struktur des Graphen beschrieben wurde, folgt nun die GPFA Formel, welche aus PFA modifiziert wurde:

$$\begin{aligned}
 G_u &= (V_u, E_u) \\
 v, w &\in V_u, e = (v, w) \in E_u \\
 m_u(e) &= \sum_{j \in W_v \cap V_u} (\beta_j + \gamma_j s_j(e) + \rho_j f_j(e))
 \end{aligned} \tag{4.2}$$

Die Erklärung der einzelnen Variablen ist identisch mit der Beschreibung von Formel 2.5

im Abschnitt zu PFA (s. Abschnitt 2.9.5). Die $Perf^R$ wird hier für einen Nutzer u und den Aktivitätsübergang $e = (v, w)$ berechnet. Dabei werden alle bisherigen Aktivitäten j aus der Knotenmenge V_u mit der Wissenskomponente W_u betrachtet. Anschließend wird der Erfolg bzw. Misserfolg anhand der ausgehenden Kante von j bestimmt. Um dieses Modell generisch für adaptive Assistenzsysteme einsetzen zu können, muss die E-Learning-Plattform einige Voraussetzungen erfüllen. Zum einen sollte die Plattform unterschiedliche Klassen an Lernobjekten besitzen, damit die einzelnen Lernobjekte den verschiedenen WKs zugeordnet werden können. Zum anderen sollte das System unterschiedliche Schwierigkeitsstufen unterscheiden können. Beispielsweise bieten die meisten Lernplattformen die Möglichkeit zwischen den Stufen schwer, mittel und leicht zu wählen. Diese Stufen sind für die Auswahl von β für eine WK in Formel 4.2 relevant.

Die xAPI Statements werden alle aus dem LRS für den jeweiligen Nutzer abgefragt und sortiert. Eine Sortierung erfolgt nach dem Zeitstempel aus dem xAPI-Format, um das in Abschnitt 2.9.5 erläuterte Problem mithilfe von „Data Aging“ zu beheben. Anschließend wird mittels GPFA die $Perf^R$ berechnet. Wie im wissenschaftlichen Schreiben von Gong, J. Beck u. a. 2011 wurden unterschiedliche Ansätze des PFA Modells evaluiert. Als schwer zu interpretieren, erwies sich hierbei das Ergebnis, dass das „All Skill“ PFA Modell gegenüber dem klassischen Modell besser abschnitt. Bei der „All Skill“ PFA werden nicht nur die WK des aktuell betrachteten Lernobjekts in der Berechnung für m verwendet, sondern alle vorhandenen WKs [Gong, J. Beck u. a. 2011]. Dieses Ergebnis erfordert jedoch noch weitere Überprüfungen und Versuche, um eine eindeutige Schlussfolgerung aus diesem Ergebnis auf das GPFA Modell übertragen zu können. Zuletzt werden die kombinierte Performance und weitere Parameter, wie z.B. $Perf^P$, $Perf^R$ und Empfehlungen, in dem JSON Format an den adaptiven Assistenten zurückgesendet. Ein Beispiel für eine Antwort des GPS Moduls zeigt folgende Auflistung:

```

1 {
2   "success": true ,
3   "performance": 0.75 ,
4   "result_performance": 1 ,
5   "progress_performance": 0.5 ,
6   "recommendation": {
7     "difficulty": "hard" ,
8     "next_activity": {
9       "name": "HarryPotterBridge" ,
10      "id": "'https://www.iosb.fraunhofer.de/spotx/game/play/47'"
11    }
12  }
13 }
```

Spot-X / Quizlernspiel Spot-X / Preview

QUIZ

Quizlernspiel Spot-X

Quiz Settings Questions Results Question bank More ▾

Back

Question 2
Not yet answered
Marked out of 1.00
Flag question
Edit question

Löse folgende Gleichung durch Äquivalenzumformung:

$$\frac{1}{2}x + 2 = 12$$

a. 4
 b. 26
 c. 2
 d. 6
 e. 14
 f. 10
 g. 20
 h. 5

Previous page Next page

Abbildung 4.3: Moodle Lernobjekt mit den Wissenskomponenten Mathematik und Gleichungen

4.3 Generisches Performance-Modell für Assistenzsysteme

Im Rahmen der vorliegenden Arbeit wird der GPS als generisches Performance-Modell verwendet, um allgemein für Assistenzsysteme eingesetzt zu werden. In diesem Abschnitt erfolgt die Vorstellung einer Architektur, welche es Learning Management Systemen (LMS) oder Serious Games ermöglicht, das Performance-Modell über einen adaptiven Assistenten zu nutzen.

- Die Nutzerdaten werden mittels xAPI-Statements in einem LRS gesichert.
- Die Lernobjekte des LMSs oder SGs sind einem oder mehreren WKs zugeordnet.
- Das Absolvieren eines Lernobjekts kann der Kategorie Erfolg oder Misserfolg zugeordnet werden und werden durch eine erreichte Punktzahl repräsentiert.

Um die Funktionsweise des Performance-Modells besser vorzustellen, wird das Konzept anhand von ILIAS und Moodle das Konzept erläutert. Anschließend wird in Kapitel 5 mittels eines weiteren Beispiels, dem Quizlernspiel Spot-X, die Operationalisierung des Modells vorgestellt. Wie bereits in Abbildung 4.1 zu sehen ist, müssen für die Performance-Messungen

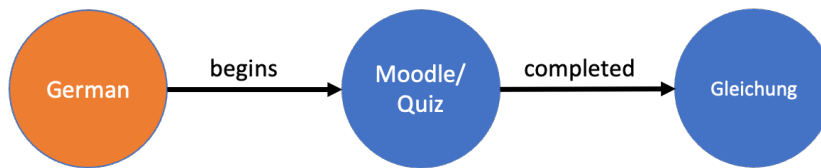


Abbildung 4.4: Ein Graph erstellt aus erfassten Interaktionen des Nutzers German aus dem LMS Moodle.

erst einmal Nutzerdaten in Form von xAPI-Statements zur Verfügung gestellt werden. Hierfür müssen für das SG oder LMS Lerneinheiten erstellt werden. Diese erhält man, indem man mithilfe von didaktischen Mitteln solche Fähigkeiten ausarbeitet (s. Abbildung 4.3). Wie bereits in Kapitel 2.6 erläutert, können dann xAPI-Statements aus den Interaktionen mit einem Nutzer erzeugt werden. Ein Beispiel für eine angelegte Frage im LMS Moodle ist in Abbildung 4.3 zu sehen. Dabei hat der Nutzer die Möglichkeit, eine Antwortoption auszusuchen und anschließend seine Antwort zu bestätigen.

Um den Lerninhalt aus Moodle so zu gestalten, dass dieser für das Performance-Modell benutzt werden kann, müssen die erfassten Statements eine konkrete Struktur aufweisen. Wenn der Nutzer einen Kurs startet, muss dieser mit dem Verb „begins“ gekennzeichnet werden. Die dazugehörige Aktivität ist das Objekt des Statements, welches in diesem Fall der Kurs ist. Sobald der Nutzer Aufgaben oder ähnliche Lerninhalte absolviert, muss das Verb „completed“ verwendet werden, um das Beenden zu kennzeichnen. Zusätzlich soll das xAPI-Statement einen Kontext liefern, wie beispielsweise das Ergebnis sowie die benötigte Zeit. Aus diesen Statements kann der GPS einen Graphen für die Berechnung der kombinierten Performance erstellen. Abbildung 4.4 stellt einen Graphen dar, welcher aus erfassten Interaktionen anhand der Frage aus Abbildung 4.3 errichtet wird.

Die Statements müssen in der Kontexterweiterung Aussagen über die Schwierigkeit und die WKs der Aufgabe beinhalten. Das bedeutet, dass der Tutor bzw. Entwickler beim Erstellen des Kurses dafür zuständig ist, jeder Aufgabe eine Schwierigkeitskategorie sowie benötigte WKs zuzuordnen. Für das Beispiel aus der vorherigen Abbildung wäre eine WK „Gleichung“ denkbar. Analog zu Moodle muss auch ILIAS diese Anforderungen erfüllen, um den GPS nutzen zu können. Abbildung 4.5 demonstriert, wie ein Graph für das LMS ILIAS mit einer identischen Aufgabe aussehen würde.

Zusammenfassend kann man sagen, dass eine konkrete Struktur für xAPI-Statements von einem LMS bzw. SG gefordert sind, um die Funktionen des GPS nutzen zu können. Es sind sowohl feste Verben als auch Kontextparameter nötig, um einen Graphen aus Interaktionsdaten zwischen Nutzer und E-Learning-Plattform aufzubauen (s. Abbildung 4.6). Wenn die-

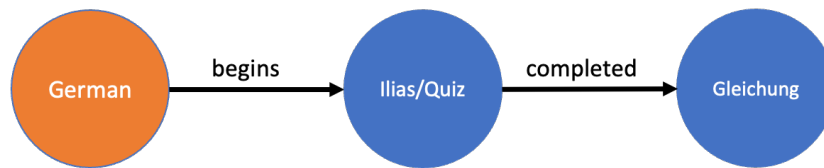


Abbildung 4.5: Ein Graph erstellt aus erfassten Interaktionen des Nutzers German aus dem LMS ILIAS.

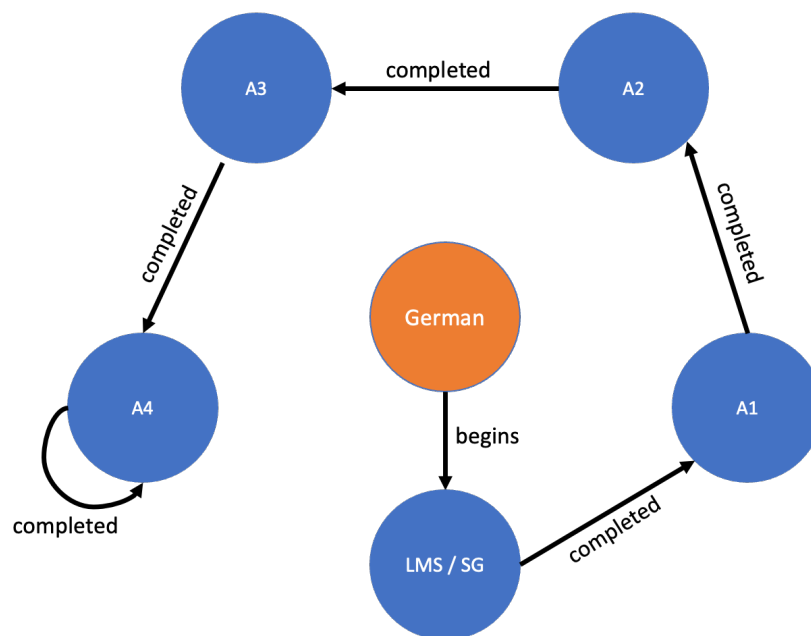


Abbildung 4.6: Ein Graph erstellt aus erfassten Interaktionen des Nutzers German aus einem LMS bzw. SG.

se Anforderungen erfüllt sind und ein solcher Graph erstellt werden kann, lässt sich das Performance-Modell generisch von anderen Systemen verwenden.

Nachdem die Nutzerinteraktionen in einem LRS gespeichert sind, kann das Performance-Modell mit einem Zugriff auf die LRS Instanz die nötigen Statements abrufen. Durch das Beenden einer Aufgabe oder einer Frage wird eine HTTP Anfrage mit der Methode GET an das Performance-Modell geschickt. Dabei wird ein eindeutiger Identifikator des aktuellen Nutzers mitgesendet, um ihre xAPI-Statements abzurufen. Wie genau die Berechnung für die kombinierte Performance erfolgt, wurde bereits in Abschnitt 4.1 beschrieben. Anschließend muss der errechnete Performance-Score aus dem JSON Format ausgelesen werden. Basierend auf dem Performance-Score und der Aufgabe, in der sich der Nutzer befindet, bestimmt der Assistent anhand von Grenzwerten eine Anpassung des Schwierigkeitsgrades und zeigt anschließend die

nächste Aufgabe. Wie in Abbildungen 2.8 zu sehen ist, wird hier eine Frage beim Quizlernspiel Spot-X gestellt, jedoch sind die Einstellungen der Schwierigkeitsstufe unterschiedlich.

Um aus Fragen xAPI-Statements zu erzeugen, benötigen manche LMS Erweiterungen. Diese ermöglichen es anschließend, die Interaktionen des Systems mit dem Nutzenden zu verfolgen. Im Rahmen dieser Arbeit wurde für ILIAS das Plugin von Stefan Schneider 2018 installiert. Auf der Seite von *Ilias Documentation 2023* befindet sich eine ausführliche Dokumentation zur Installation des Plugins. Zum Zeitpunkt der Ausarbeitung dieser Thesis wurde jedoch kein Plugin gefunden, welches die Manipulation der erzeugten xAPI-Statements zulässt. Zusätzlich zu ILIAS wurde das xAPI Plugin für Moodle aus Pesce 2022 installiert. Die Dokumentation und Anleitung ist auf *Moodle Documentation 2023* zu finden. Sowohl Moodle als auch ILIAS bieten nur eine Erfassung von xAPI-Statements mittels der genannten Plugins an. Das verhindert das Erstellen eigener Statements für diese Plattformen und somit eine Anpassung der Statements nach den hier vorgestellten Anforderungen. Daher konnte im Rahmen dieser Arbeit das Konzept eines generischen Performance-Modells für adaptive Assistenzsysteme im Allgemeinen nur konzeptionell umgesetzt werden. Das ermöglicht jedoch eine Erweiterung der Plattformen für zukünftige Arbeiten.

Abschnitt 2.2.1 konnte bereits verdeutlichen, dass die adaptive Anpassung an den Nutzer ein iteratives Verfahren darstellt. Dies bedeutet, dass nach der Präsentation des neuen Inhalts der Nutzer die Aufgabe wieder beantwortet, was wiederum für die neue Performance-Messung berücksichtigt werden muss. Im folgenden Kapitel wird die Operationalisierung des Performance-Modells mit einer kombinierten Performance-Berechnung am Beispiel des Quizlernspiels Spot-X vorgestellt.

5 Operationalisierung der Performance-Messungen

In diesem Kapitel erfolgt die Darstellung der im Rahmen der vorliegenden Arbeit entstandenen Operationalisierung. Hierzu werden die erarbeiteten Konzepte aus Kapitel 4 umgesetzt. Zur Operationalisierung werden zunächst einige Hypothesen, darunter auch die Nullhypothese festgelegt, welche durch diese Operationalisierung belegt werden sollen. Anschließend werden die erfassten xAPI-Statements untersucht und potenzielle Einflussfaktoren zur Performance-Berechnung ausgewählt. Zu Beginn wird der Graph Performance Service (GPS) um eine Result-Performance ($Perf^R$) erweitert (s. Abschnitt 4.1). Dieses Performance-Modell ist der Proof Of Concept für diese Arbeit und wird in Abschnitt 5.2 erläutert. Anschließend wird der intelligente, adaptive Tutor im SG Spot-X um die Funktionalitäten des Performance-Modells ergänzt. Während der Nutzerstudie, die in Kapitel 6 behandelt wird, haben die Performance-Messungen der $Perf^P$ nicht den erwarteten Werten entsprochen. Nach genauerer Untersuchung der Berechnung sind zwei weitere Implementierungsdetails aus einer vorherigen Thesis aufgefallen, welche dazu geführt haben, dass die theoretische Berechnung der $Perf^P$ nicht mit jener in der Anwendung übereingestimmt hat. Aufgrund dieser Untersuchung werden in Abschnitt 5.6 die Anpassungen und Erweiterungen des GPS nach den neu gewonnenen Erkenntnissen vorgestellt.

5.1 Festlegung der Hypothesen

Zunächst müssen Vermutungen und Annahmen über die kombinierte Performance-Berechnung in Form von Hypothesen getroffen werden. Dazu werden in diesem Abschnitt die Hypothesen bzw. Nullhypothesen bestimmt, die durch operative Performance-Messungen in Form einer Nutzerstudie und geprüft werden sollen. Um die Hypothese zu formulieren, werden eine Annahme und eine These benötigt. Wie bereits in Abschnitt 2.2.1 erläutert, hat Adaptivität eines SG Einfluss auf den Lernerfolg eines Nutzers. Aufbauend auf dieser Annahme wurde eine kombinierte Performance berechnet. Dieser „Performance Score“ wurde mittels $Perf^P$ und $Perf^R$ (s. Abschnitt 4.1) bestimmt. Die These dazu lautet, dass der Gebrauch dieses be-

rechneten „Performance Scores“ die Adaptivität an den Nutzer verbessert. Aus diesen beiden Formulierungen lässt sich die Hypothese H_0 und H_1 aus Tabelle 5.1 ableiten. Weiterhin wurden Hypothesen H_2 und H_3 aufgestellt, welche durch operative Performance-Messungen in Form einer Nutzerstudie geprüft werden sollen.

Abkürzung	Hypothese
H_0	Keine Veränderung der kombinierten Performance durch die Nutzung der generierten Lernempfehlung.
H_1	Die kombinierte Performance steigt durch die Verwendung der generierten Lernempfehlung, die durch den GPS bereitgestellt wird.
H_2	Die Selbsteinschätzung des Nutzers korreliert mit dem „Performance Score“.
H_3	Die Berücksichtigung der erreichten Punktzahl und der benötigten Zeit führt zu sinnvollen Performance Scores. Sinnvolle Scores meinen hierbei einen Wert, der die Stereotypen (Anfänger, Intermediate, Experte) des Nutzers widerspiegelt.

5.2 Implementierung des Performance-Modells für Graphen

Für die Performance-Berechnungen werden xAPI-Statements benötigt (s. Kapitel 4.2). Die erfassten Interaktionsdaten des Quizlernspiels Spot-X mit dem Nutzer folgen dem xAPI Standard und sind in einem LRS zugänglich. Im Rahmen dieser Arbeit werden die Performance-Berechnungen mit Daten realer Anwender in Form einer Nutzerstudie durchgeführt. Ähnlich wie in der Arbeit von Delaunay 2022, in welcher die unterschiedlichen Attribute aus den Beobachtungsdaten für die $Perf^P$ operationalisiert wurden, erfolgt in diesem Abschnitt die Operationalisierung der Daten für $Perf^R$. Die Attribute werden in den xAPI-Statements mitgeliefert und zur Anwendung des GPS genutzt. Folgende Tabelle zeigt notwendige Attribute für die Berechnung der kombinierten Performance aus einer abgeschlossenen Spot-X-Aufgabe:

Attribute	Wertebereiche
Punktzahl	[0, 100]
Erfolg	{ <i>True, False</i> }
Schwierigkeitskategorie	{ <i>easy, normal, hard</i> }
Benötigte Zeit	[0, 75]
Kategorien	{ <i>Mathematik, Geschichte, Sehenswürdigkeiten, TV&Filme, ...</i> }

Der Wert für das Attribut Punktzahl ist 0, wenn die Aufgabe falsch (*Erfolg = False*) und 100, wenn die Aufgabe vollständig richtig (*Erfolg = True*) beantwortet wurde. Die benötigte Zeit wird in Sekunden angegeben, wobei die maximal zugestandene Zeit für die Aufgabe nach 75 Sekunden ausläuft und die Frage automatisch als nicht bzw. falsch beantwortet gilt. Die Kategorien definieren eine Submenge von Wissenskomponenten, welche in der abgeschlossenen Aufgabe abgefragt wurden.

Neben den abgeschlossenen Aufgaben werden weitere xAPI-Statements zu Interaktionen mit dem Nutzer in Spot-X an den LRS gesendet. Darunter fallen z.B. das Ein- und Ausloggen aus dem System oder das Betrachten von Ergebnissen eigener Spieldurchläufe. Diese Interaktionen könnten ebenfalls Informationen über die Performance des Nutzers liefern und damit bei einer adaptiven Anpassung nützlich sein, werden aber im weiteren Verlauf der Arbeit nicht berücksichtigt. Diese Interaktionen in die Berechnungen von $Perf^P$ oder $Perf^R$ zu integrieren geht über den Rahmen der Aufgabenstellung dieser Arbeit hinaus und kann in zukünftige Arbeiten miteinfließen. Im Rahmen dieser Arbeit wird der Graph für Spot-X auf die Aktivitäten eines Spielstarts, einer Beendigung des Spiels und den Abschluss von Aufgaben eingegrenzt. Um die irrelevanten Interaktionen herauszufiltern, wurde von Delaunay 2022 bereits der GPS erweitert, sodass xAPI-Statements, wie z.B. Ein- und Ausloggen, nicht berücksichtigt werden.

5.3 Implementierung der Result- und Progress-Performance

Der folgende Abschnitt widmet sich der Operationalisierung graph-basierter Performance-Messungen durch die Anwendung des GPS auf Spot-X. Im Rahmen dieser Arbeit wird die Implementierung zur Berechnung der $Perf^P$ angepasst und um die $Perf^R$ erweitert. Anschließend erfolgt eine Kombination dieser beiden Performance-Messungen kombiniert (siehe Abschnitt 4.1). Dabei werden verschiedene Einflussfaktoren operationalisiert, welche in den Interaktionsdaten aus Spot-X vorhanden sind.

Abbildung 5.1 soll das allgemeine Vorgehen des hier angewendeten GPS verdeutlichen. Hierzu werden verschiedene Indikatoren aus den Benutzerdaten zur Bestimmung idealer Pfade und der WKs für die GPFA mitberücksichtigt. Im Vorfeld dieser Arbeit wurden die Auswirkungen der Interaktionen des Nutzers ausschließlich anhand der entstehenden Lernpfade sowie der $Perf^P$ gemessen [Delaunay 2022]. Diese Auswirkungen werden nun um die Einflüsse der WKs und der $Perf^R$ erweitert. Somit werden die Ergebnisse der Aktivitäten der einzelnen WKs für einen Nutzer berechnet, um daraus eine Performance-Vorhersage für die nächste Aufgabe zu ermitteln. Diese Vorhersage wird anschließend für eine Anpassung der Schwierigkeitskategorie verwendet.

Die Aktivitätsknoten, d.h. die einzelnen Aufgaben aus Spot-X, besitzen das Attribut Katego-


```

3 MATCH(t) - [e: COMPLETED] -> (m: Activity)
4 WHERE e.userid = $username
5 RETURN DISTINCT {
6     knowledge_component: t.categories ,
7     taskId: t.id ,
8     difficulty: t.difficulty ,
9     success: e.success ,
10    score: e.score
11 } as task

```

Die $Perf^R$ wird anhand der GPFA ermittelt, welche in dieser Arbeit auf der „All Skill“ PFA basiert [Gong, J. Beck u. a. 2011]. Wie bereits in Abschnitt 4.2 erläutert, zeigt die „All Skill“ PFA Variante bessere Ergebnisse bei der Performance-Vorhersage in [Gong, J. Beck u. a. 2011] an und wird daher auch in der Operationalisierung verwendet. Eine Alternative wäre weiterhin die klassische GPFA, welche nur die WKs \$KCs betrachtet, die in der vorherigen Aufgabe abgefragt wurden. Für die Implementierung der Cypher Abfrage würde das einen weiteren Filter in Auflistung 5.3 Zeile vier bedeuten:

```
WHERE e.userid = $username AND ($KCs CONTAINS t.categories)
```

Wie bereits in Kapitel 2.9.5 beschrieben, erhält die Reihenfolge der einzelnen Antworten bei der PFA keine Berücksichtigung [Gong, Joseph E Beck u. a. 2010]. Um diesen Nachteil auszugleichen, wurde ein „decay factor“ eingeführt, welcher jedoch eine Sortierung der Aufgaben erfordert. Hierfür wurden die zurückgelieferten Objekte aus Auflistung 5.3 absteigend sortiert.

Der GPS wurde so konzipiert, dass eine Aufgabe entweder ein Erfolg oder Misserfolg ist. Dies bedeutet für die Interpretation der Attribute, dass man aus einem Misserfolg $success = False$ & $score = 0$ und einem Erfolg $success = True$ & $0 < score \leq 1$ schließt. Um einen passenden „decay factor“ zu bestimmen, wurde die Information des wissenschaftlichen Schreibens von Gong, Joseph E Beck u. a. 2010 verwendet. Dort wird ein „decay factor“ von $\delta = 0.9$ ausgewählt, da $\delta = 1$ die klassische Performance Factor Analysis bedeuten würde und ein niedrigerer Faktor ($\delta < 0.9$) ältere Antworten als deutlich weniger wichtig einstufen würde. Da die berechnete $Perf^R$ einen Wert m im Intervall $m \in [0, \infty)$ ausgibt, aber ein Wert $0 \leq m \leq 1$ erwartet wird, wird ein normalisierter Performance-Score berechnet.

Nachdem der soeben vorgestellte $Perf^R$ Score und der in Delaunay 2022 vorgestellte $Perf^P$ Score berechnet wurden, erfolgt die Berechnung der kombinierten Performance $Perf^{combined}$ mit der Formel 4.1. Anhand dieser kombinierten Performance wird dann eine Empfehlung für die nächste Aufgabe ausgesprochen. Hierfür wurden die folgenden Schwellenwerte betrachtet:

- $0 \leq Perf^{combined} < 0.3$: Empfehlung Schwierigkeitsgrad „easy“

- $0.3 \leq Perf^{combined} < 0.6$: Empfehlung Schwierigkeitsgrad „normal“
- $0.6 \leq Perf^{combined} \leq 1$: Empfehlung Schwierigkeitsgrad „hard“

Nach der Berechnung liefert der GPS eine Antwort in einem JSON Format an das SG Spot-X zurück. Eine mögliche Antwort ist in Auflistung 4.2 abgebildet. Eine Schwierigkeit, die im Prozess der Operationalisierung entstanden ist, stellen die Empfehlungen dar. Sowohl die Lernpfade als auch die GPFA Berechnung liefern eine Empfehlung. Die Empfehlung durch die IPMs erfolgt in Form einer nächsten Aufgabe und jene durch die GPFA in Form einer Anpassung des Schwierigkeitsgrads. Welche dieser Empfehlungen nun wahrgenommen werden, wird vom adaptiven Assistenten entschieden. Das Performance-Modell ist nur dafür zuständig, die Empfehlungen zu berechnen und auszugeben. Für Spot-X wurde dieser die Entscheidung getroffen, dass die priorisierte Empfehlung aus der kombinierten Performance stammt. Das bedeutet, dass die Empfehlung nur eine Schwierigkeitsanpassung zur Folge hat. Falls die angepasste Schwierigkeitsstufe mit der Schwierigkeitsstufe der empfohlenen Aufgabe der Lernpfade übereinstimmt, wird diese als nächste Aufgabe selektiert. Sonst wird eine andere Aufgabe aus derselben Schwierigkeitskategorie ausgewählt.

5.4 Erweiterung des adaptiven Assistenten für Spot-X

Damit das Performance-Modul bzw. der GPS für das Quizlernspiel Spot-X verwendet werden kann, muss der adaptive Assistent E-Learning Artificial Intelligence (ELAI) auf den Microservice zugreifen. Außerdem müssen für die Evaluation der Nutzerstudie, die in Kapitel 6 beschrieben wird, sowohl die Selbsteinschätzung des Nutzers nach jeder Aufgabe von Spot-X als auch die $Perf^P$, die $Perf^R$ und die kombinierte Performance festgehalten werden. Zur Selbsteinschätzung wurde bereits ein Dialog in Spot-X entwickelt, der in dieser Arbeit verwendet werden konnte. Abbildung 5.2 zeigt ein Dialogfenster zur Selbsteinschätzung nach einer Aufgabe im Quizlernspiel Spot-X.

Das Dialogfenster verwendet die Likert-Skala als eine graduelle Antwortskala. Eine Likert-Skalierung eignet sich grundsätzlich für jede Befragung, bei der es sich um die persönliche Meinung der Befragten handelt [Joshi u. a. 2015]. Die Anzahl an Kategorien wurde ungerade gewählt, weil es dem Nutzer die Möglichkeit auf eine neutrale Position gibt und somit zu keiner Tendenz zwingt, so wie es bei der geraden Anzahl an Kategorien der Fall wäre. In dieser Arbeit wurden die in Guy u. a. 1977 und Garland 1991 empfohlenen fünf Kategorien ausgewählt. Im Dialogfenster von Spot-X wird der Nutzer zuerst gefragt, wie zufrieden er mit seiner eigenen Leistung in dieser Aufgabe ist und erhält die Möglichkeit, seine Meinung in einer 5-stufigen Likert-Skala abzugeben. Dabei bedeutet ein Stern, dass der Nutzer überhaupt

Your help is needed: to be able to improve the AI functionality in the long term please share your assessment here. This assessment is only used in the context of the evaluation and has no effect on your current session

How satisfied are you with your performance in the last task? (One: "I did poorly", five: "I did great")

☆☆☆☆☆

How difficult was the last task for you? (One: Very easy, five: very difficult)

☆☆☆☆☆

Did you feel that the difficulty of the task was appropriate for you? (Too difficult, appropriate, too easy)

☹️ 😐 😊

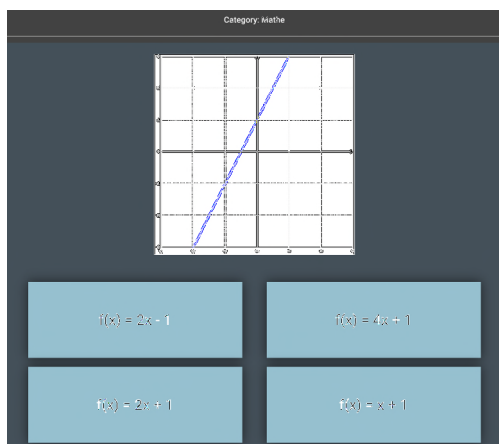
Send

Abbildung 5.2: Ein Dialog zur Selbsteinschätzung nach einer Aufgabe im Quizlernspiel Spot-X.

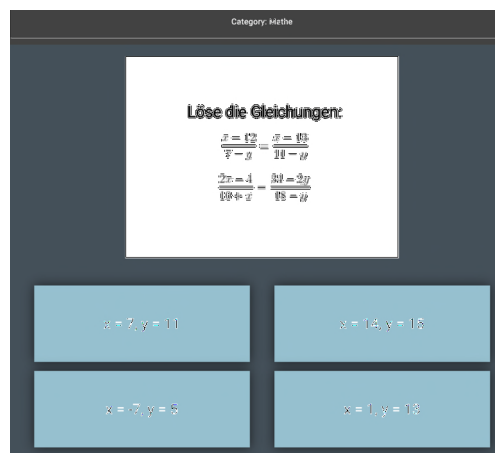
nicht zufrieden ist und fünf Sterne eine hohe Zufriedenheit. Anschließend wird er in der gleichen Skalierungsmöglichkeit gefragt, wie schwer die Aufgabe für ihn war. In diesem Fall wird ein Stern als eine sehr leichte Aufgabe und fünf Sterne als eine sehr schwere Aufgabe interpretiert. Die letzte Frage erkundet mit einer 3-stufigen Likert-Skalierung, ob die Schwierigkeit der Aufgabe für den Nutzer angemessen war. Zusätzlich musste der Assistent um eine Performance-Messung ohne Anpassungen erweitert werden. Die Performance des Nutzers wird jetzt sowohl in einem Spieldurchlauf ohne adaptive Anpassungen berechnet als auch in einem Spieldurchlauf mit adaptiver Anpassung. Diese Funktion wird in der Auswertung zum Vergleich der Performance-Messungen mit und ohne Adaptivität verwendet.

5.5 Vorstellung der Anwendungsszenarien

In diesem Abschnitt werden die Anwendungsszenarien aus dem Quizlernspiel Spot-X vorgestellt und die einzelnen Adaptivitätsmechanismen aufgezeigt. Zur Durchführung wird das SG mit einem Pseudonym gespielt, aus welchem Interaktionsdaten in xAPI-Format entstehen. Für das Beispielszenario werden bereits bestehende Spot-X Aufgaben durch einige weitere Aufgaben ergänzt und anschließend verwendet. Diese erfassten Daten dienen einer Operationalisierung der Performance-Messung für das adaptive Assistenzsystem Spot-X. Das Aufgabenset besteht aus insgesamt zehn Aufgaben unterschiedlicher Kategorien. Die Kategorien wurden für das Quizlernspiel bereits festgelegt und werden für die Performance-Messung verwendet. Die Kategorien umfassen Mathe, Sehenswürdigkeiten, TV & Filme und viele mehr. Dabei gibt es für jede Kategorie jeweils Aufgaben mit der Schwierigkeitskategorie leicht, normal und schwer. Die Aufgabe einer Kategorie zuzuordnen, liegt in der Verantwortung des Tutors oder des Entwicklers. Abbildung 5.3 zeigt zwei Beispielaufgaben aus dem Aufgabenset in Spot-X.



Aufgabe aus der Kategorie Mathe mit dem Schwierigkeitsgrad normal.



Aufgabe aus der Kategorie Mathe mit dem Schwierigkeitsgrad schwer.

Abbildung 5.3: Zwei Aufgaben aus dem SG Spot-X mit der selben Kategorie, aber unterschiedlichen Schwierigkeiten.

Für die $Perf^P$ ist es notwendig, ein Lernziel festzulegen, um die idealen Pfade bestimmen zu können. In diesem Fall besteht das Lernziel in der richtigen Beantwortung einer Aufgabe der Schwierigkeitskategorie schwer. Für die $Perf^R$ ist es notwendig, Wissenskomponenten festzulegen, weshalb die Kategorien als WKs behandelt werden (s. Kapitel 4.2).

Im ersten Anwendungsszenario spielt Alice eine Runde, bestehend aus zehn Aufgaben, des Quizlernspiels Spot-X ohne adaptive Anpassung durch den Assistenten ELAI. Alice startet das Spiel und ihr wird die erste Frage gestellt, wie in Abbildung 5.4 zu sehen ist.

Alice ordnet allen Begriffen das korrekte Bundesland ohne Fehlversuche zu und erhält dafür die volle Punktzahl, d.h. 100 Punkte. Da dieser Spieldurchlauf keinen adaptiven Anpassungen ausgesetzt ist, bekommt Alice, unabhängig von ihrer gegebenen Antwort sowie ihrer benötigten Zeit, eine weitere Frage gestellt. Diese Frage wird zufällig aus den Schwierigkeitskategorien ausgewählt. Auf die Frage in Abbildung 5.5 antwortet Alice falsch und der Schwierigkeitsgrad des Quizlernspiels wird, trotz falsch beantworteter Frage, nicht angepasst. Nach Bearbeitung der restlichen Aufgaben wird Alice auf eine Seite mit ihrer Spielstatistik weitergeleitet. Dort kann Alice die Ergebnisse ihrer Antworten sowie ihre Punkte der jeweiligen Aufgaben einsehen.

Nun wird das Szenario vorgestellt, in dem Alice von ELAI begleitet wird und Anpassungen am Schwierigkeitsgrad vorgenommen werden. Da Alice das Quizlernspiel zum ersten Mal spielt und noch keine Informationen über die Performance von Alice im LRS oder der Graphdatenbank existieren, beantwortet sie die erste Frage ohne Assistenzsystem. Dieses Problem

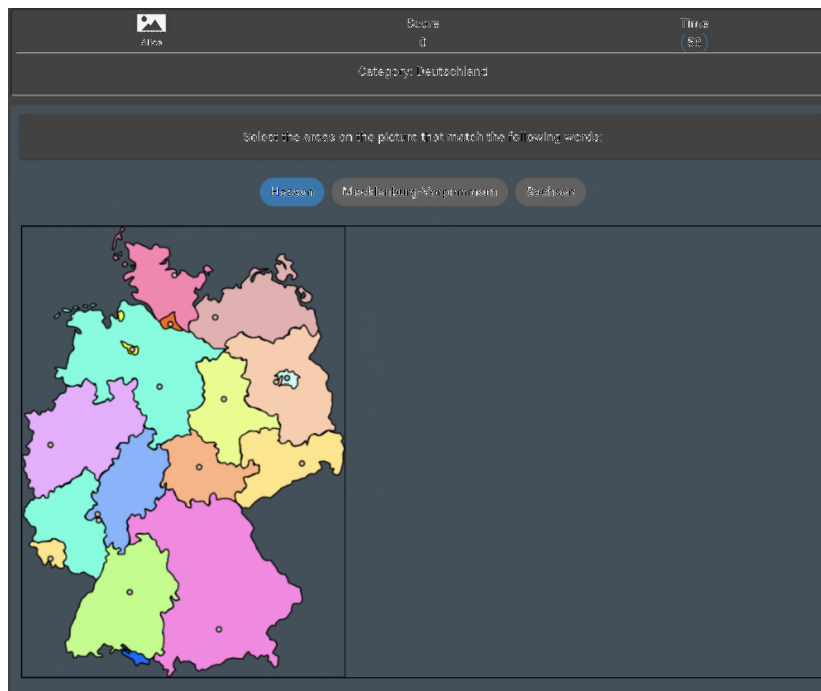


Abbildung 5.4: Alice wird eine Frage im ersten Spieldurchlauf gestellt.

ist bekannt als Kaltstartproblem. Wie bereits im Szenario ohne Assistenzsystem beantwortet Alice die erste Frage korrekt und die Interaktion wird in Form von xAPI-Statements im LRS gespeichert. Eine Selbsteinschätzung zur soeben absolvierten Aufgabe abgibt (s. Abbildung 5.2), wird eine Anfrage von ELAI an den GPS gesendet. Im GPS wird anschließend aus den Interaktionsdaten im LRS ein Graph für Alice in einer Neo4j Datenbank angelegt. Abbildung 5.6 zeigt diesen Graphen, nachdem Alice zehn Aufgaben durchlaufen hat. Hiernach werden aus diesem Graphen die $Perf^P$ mittels idealer Lern- und Nutzungspfaden, welche aus der vorgegebenen Zielfunktion bestimmt werden, und die $Perf^R$ mittels GPFA gewonnen. Zuletzt wird im GPS die kombinierte Performance aus $Perf^P$ und $Perf^R$ berechnet und eine Antwort im JSON Format (s. Auflistung 4.2) an ELAI weitergeleitet.

Aufgrund der Interaktionsdaten von Alice erhält ELAI die Empfehlung, zur Schwierigkeitskategorie schwer zu wechseln. Die Schwierigkeitskategorie der Empfehlung für die nächste Aufgabe, welche auf dem idealen Pfad liegt, stimmt mit der Empfehlung durch die kombinierte Performance überein. Daher wählt der adaptive Assistent als nächste Aufgabe die Empfehlung durch den idealen Pfad aus dem GPS. Alice wird nun die nächste Aufgabe angezeigt. So wie es auch im Adaptivitätszyklus dargestellt ist, wiederholt sich dieses Vorgehen iterativ und es kommen immer mehr Interaktionsdaten von Alice dazu. Nach zehn Aufgaben wird Alice

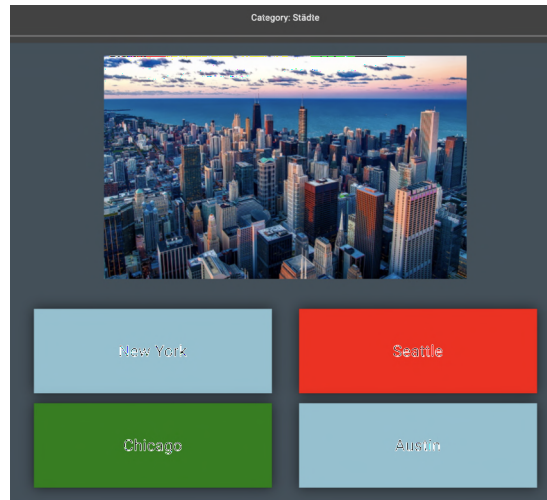


Abbildung 5.5: Alice beantwortet die zweite Frage im ersten Spieldurchlauf falsch.

wieder zu ihrer Spielstatistik weitergeleitet.

5.6 Korrektur des GPS nach der Nutzerstudie

In diesem Abschnitt der Arbeit werden die Korrekturen während der Untersuchung der Performance-Messungen vorgestellt. Nachdem die Implementierung des Performance-Modells sowie des adaptiven Assistenten für Spot-X abgeschlossen waren, hat nach einer Vorstudie die Hauptstudie begonnen. Während dieser wurden die Performance-Messungen in regelmäßigen Abständen auf ihre Plausibilität hin überprüft. Dabei sind in einer dieser Überprüfungen einige Unklarheiten zu den Werten der gemessenen $Perf^P$ aufgetreten. Sobald es einen Nutzer gab, der alle Fragen des Aufgabesets falsch beantwortete, war die erwartete $Perf^P$ eine andere, als die gemessene $Perf^P$. Als erste Maßnahme wurden für die von Delaunay 2022 geschriebenen Tests überprüft und erweitert. Eine Skizze für einen dieser Tests ist in Abbildung 5.7 einsehbar.

Um einen besseren Überblick über die aktuelle Messung von $Perf^P$ zu erhalten, wurden zwei weitere Test geschrieben. Der erste Test simulierte einen Nutzer, der alle Fragen falsch beantwortete. Der zweite Test wurde analog für einen Nutzer erstellt, welcher alles richtig beantwortet. Im Fall der Nutzerstudie ist folgende Erkenntnis erlangt worden:

Fall	erwarteter Wert	gemessener Wert
alles falsch	0.5	0.25
alles richtig	1.0	Wert \neq 1.0

Nachdem die Tests für die oben genannten Fälle geschrieben wurden, konnte festgestellt

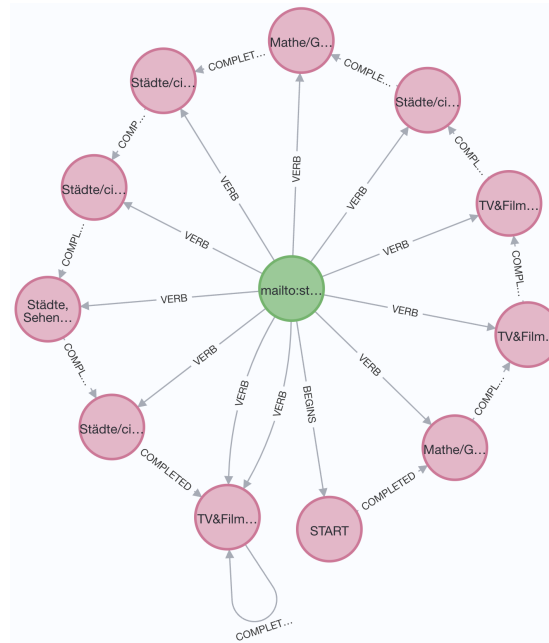


Abbildung 5.6: Ein erzeugter Graph in Neo4j aus den Interaktionsdaten von Alice.

werden, dass die Kantentupel zum Teil keine korrekte Reihenfolge besaßen. Demnach wurde bei einem Aktivitätsübergang von Aktivität A_1 zu Aktivität A_2 eine Kante (A_2, A_1) , anstelle einer Kante (A_1, A_2) erstellt. Diese Kanten wurden dann für den idealen Pfad verwendet, während die Kanten im Lernpfad des Nutzers korrekt sortiert waren. Dies hatte zur Folge, dass der gemessene Wert bei einem Spieldurchlauf mit ausschließlich richtigen Antworten ungleich 1 war. Daraufhin wurde die $Perf^P$ Berechnung mithilfe von Tests erneut überprüft und für valide erachtet.

Anschließend erfolgte eine weitere Überprüfung mit Interaktionsdaten von Teilnehmern der Nutzerstudie, wobei sich die Berechnung der $Perf^P$ weiterhin als fehlerhaft einstufen ließ. Die erwarteten Werte stimmten nicht mit den gemessenen Werten überein, obwohl die Tests für die theoretische Berechnung korrekt waren. Aufgrund der synthetisch erstellten Daten für die Tests der $Perf^P$, wurde im nächsten Schritt betrachtet, welche Interaktionsdaten aus dem Graphen in Neo4j gewonnen werden. Dabei ist aufgefallen, dass die Cypher-Anfragen für die Kanten des Lernpfades falsch abgefragt wurden. In der Berechnung für $Perf^P$ werden alle absolvierten Aktivitäten als Knoten des Lernpfades betrachtet, unabhängig davon, ob die Aktivität erfolgreich beantwortet wurde oder nicht. Die Kanten hingegen sollen nur gelten, wenn die Aktivität des ausgehenden Knotens ein Erfolg war. Die Kanten für die Interaktionen aus Spot-X wurden unabhängig von Erfolg oder Misserfolg zum Lernpfad hinzugefügt. Dies konnte

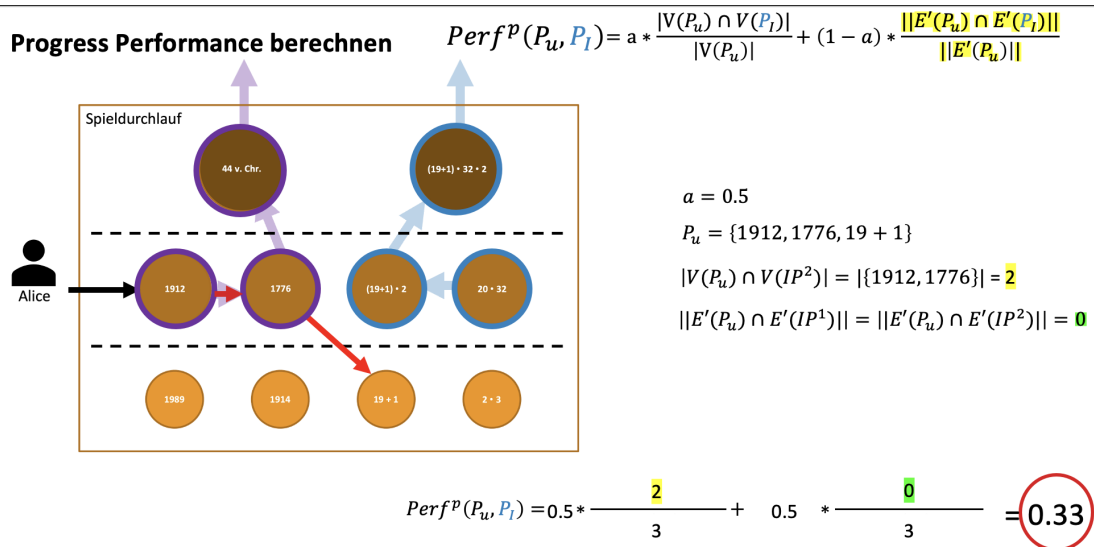


Abbildung 5.7: Ein Beispielszenario für die Berechnung der $Perf^P$ aus der Arbeit von Delaunay 2022.

mit einer zusätzlichen Cypher-Anfrage korrigiert werden und hatte zur Folge, dass sowohl die Tests als auch die Berechnung auf den Spot-X Interaktionsdaten korrekt funktionierten.

Ein weiterer Fehler ist im Zuge der Lernzielanpassung aufgefallen. Das Lernziel ist standardmäßig auf die Schwierigkeitskategorie schwer gesetzt. Sofern jedoch keine der insgesamt zehn Aufgaben den Schwierigkeitsgrad schwer besitzt, ist es nicht möglich, ein Ziel zu finden und folglich einen idealen Pfad zu berechnen. Ohne ideale Pfade kann auch keine $Perf^P$ berechnet werden. Um dieses Problem zu lösen, wurde im Rahmen dieser Arbeit eine kaskadierende Überprüfung entwickelt. In diesem Fall wird immer die nächst leichtere Schwierigkeitskategorie für ein Lernziel selektiert. Das bedeutet, dass beim Fehlen einer Aufgabe mit der Schwierigkeitskategorie schwer, das Lernziel normal gewählt wird. Analog für die Schwierigkeitsstufen normal und leicht. Abbildung 5.8 veranschaulicht die gemessene $Perf^P$ vor und nach der Korrektur für den Fall, dass ein Nutzer alle zehn Aufgaben falsch beantwortet. Der Fall für die falsche Messung ist eingetreten, falls die Zielfunktion nicht passend gesetzt war.

Diese Schwierigkeiten und Probleme sind im Laufe der Untersuchung aufgetaucht und wurden im Rahmen dieser Arbeit korrigiert. Im Folgenden widmet sich Kapitel 6 einer Erläuterung sowie Diskussion der Ergebnisse und Folgen dieser Korrekturen.

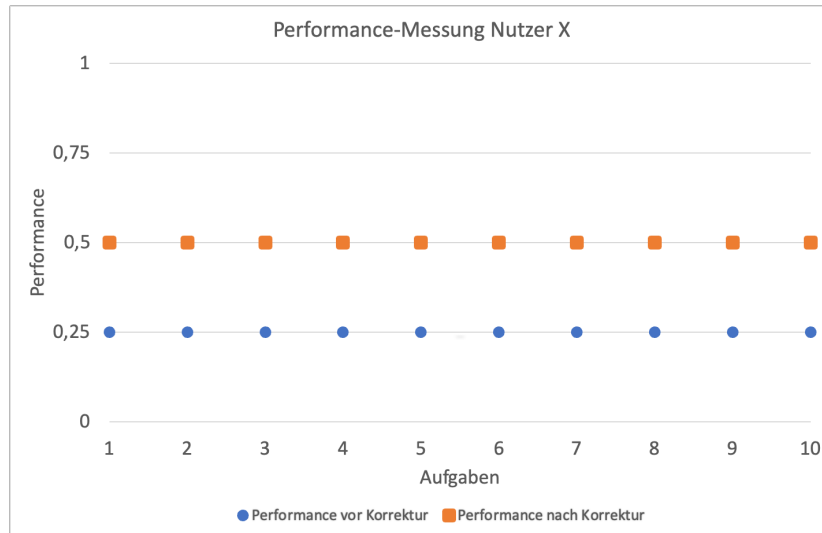


Abbildung 5.8: Zwei Messungen der $Perf^P$ für den Fall, dass der Nutzer alles falsch beantwortet. Die blauen Kreise zeigen die Performance vor und die orangenen Quadrate die Performance nach der Korrektur.

5.7 Diskussion

Das folgende Unterkapitel intendiert eine Diskussion darüber, inwieweit der Entwurf des Kapitels 4 im Kontext der Operationalisierung dieser Arbeit realisiert werden konnte. Ähnlich zu den vorherigen Kapiteln wird zunächst die Erweiterung des GPS um die kombinierte Performance und anschließend das Performance-Modell besprochen.

Der GPS kann eine Anfrage eines Assistenzsystems entgegennehmen und mithilfe der weitergeleiteten Informationen eine kombinierte Performance mit zusätzlichen Empfehlungen berechnen. Dafür müssen xAPI Statements aus dem LRS geladen und in der Graphdatenbank Neo4j gesichert werden. Anhand der Darstellung der Interaktionsdaten kann das Modul die $Perf^R$ mittels Graph Performance Factor Analysis berechnen und diesen mit der $Perf^P$ zu einer kombinierten Performance zusammensetzen. Diese wird anschließend als JSON Objekt zurückgesendet. Somit wurden alle Funktionalitäten aus Abschnitt 4.2 implementiert.

Die Implementierung der GPFA Berechnung beruht nicht vollständig auf dem Stand der Forschung und Technik Methodik und kann daher verbessert werden. Damit das Modul sowohl für das Serious Games als auch die Nutzerstudie funktioniert, wurde die Implementierung einfach gehalten. Die fehlenden Parameter und Faktoren (z.B. Wissenskomponente) sollten, sobald das Quizlernspiel Spot-X um einige Aufgaben und Kategorien erweitert wurde, in zukünftigen Arbeiten ergänzt und untersucht werden. Die Erkenntnisse von Gong, Joseph E

Beck u. a. 2010 sind, dass das klassische PFA schlechtere Ergebnisse liefert als das „All-Skill“-PFA, welches alle WKs und Fähigkeiten verwendet. Aufgrund dieses Ansatzes wurde nicht nach WKs gefiltert, sondern alle Fähigkeiten für die Analyse verwendet.

Nach der Implementierung dieser Arbeit ist das Performance-Modell in der Lage, eine Performance für die Anfrage eines SGs oder einer E-Learning-Plattform zu berechnen. Hierfür muss die Plattform jedoch noch einige Vorbedingungen erfüllen. Zum Beispiel ist der GPS nur in der Lage, Graphen in der Neo4j Datenbank aus xAPI-Statements zu erstellen. Weiterhin ist es notwendig, dass die Statements ein bestimmtes Format besitzen. Dies bedeutet, dass das Abschließen einer Aufgabe oder einer Lerneinheit mit dem xAPI Verb „completed“ gekennzeichnet werden muss. Für weitere Arbeiten bietet das die Möglichkeit, den GPS weiter zu generalisieren und für andere Plattformen leichter zugänglich zu gestalten.

Das Performance-Modell ist durch die bereits vorhandene Anbindung an den LRS und die Neo4j Datenbank leicht durch weitere Einflussfaktoren auf die Performance, wie z.B. Eye-Tracking oder Hilfestellungen in Form von Hinweisen, erweiterbar. Die Empfehlung der Schwierigkeitskategorie wird durch den GPS gewährleistet. Zwar existiert für jede Wissenskomponente eine Auswahl an leichten, normalen und schweren Aufgaben, es kann jedoch durch die Anpassung des SG oder der E-Learning-Plattform weitere Erweiterungen der Empfehlungen geben. Dies kann beispielsweise in Form verschiedener Schwierigkeitskategorien innerhalb einer Aufgabe oder durch eine unterschiedliche Anzahl an Antwortmöglichkeiten realisiert werden. Zusätzlich wurde die Aufzeichnung der Interaktionen in einem Graphen erweitert, was von großem Interesse für andere, darauf aufbauende Themen oder die Weiterführung dieser Arbeit sein kann.

6 Nutzerstudie: Evaluation der Performance-Messung

In diesem Kapitel wird die Durchführung der Operationalisierung abgeschlossen. Hierzu erfolgt eine nähere Betrachtung der im Rahmen dieser Arbeit entstandenen Nutzerstudie. Zunächst wird der Aufbau der Studie dargelegt, um anschließend deren Folgen und Ergebnisse, vor und nach den Korrekturen vorzustellen (s. Abschnitt 5.6). Abschließend wird eine Interpretation sowie ein Vergleich der Ergebnisse mit den im Unterkapitel 5.1 definierten Hypothesen intendiert.

6.1 Experimentelles Verfahren

Für den Aufbau der Nutzerstudie wird das SG Spot-X verwendet. Die Studie wurde im Zeitraum von März bis Mai 2023 durchgeführt und erfolgte in Präsenz. Abbildung 6.1 veranschaulicht den Ablauf der Nutzerstudie. Nach der Begrüßung des Nutzers wird diesem die Thematik der Arbeit sowie das weitere Vorgehen dargelegt. Daraufhin wird dem Nutzer eine Datenschutzerklärung in Form einer LimeSurvey Umfrage ausgehändigt (s. Abbildung 6.2), deren Unterzeichnung für die weitere Teilnahme erforderlich ist. Im Anschluss daran beginnen die Spieldurchläufe.

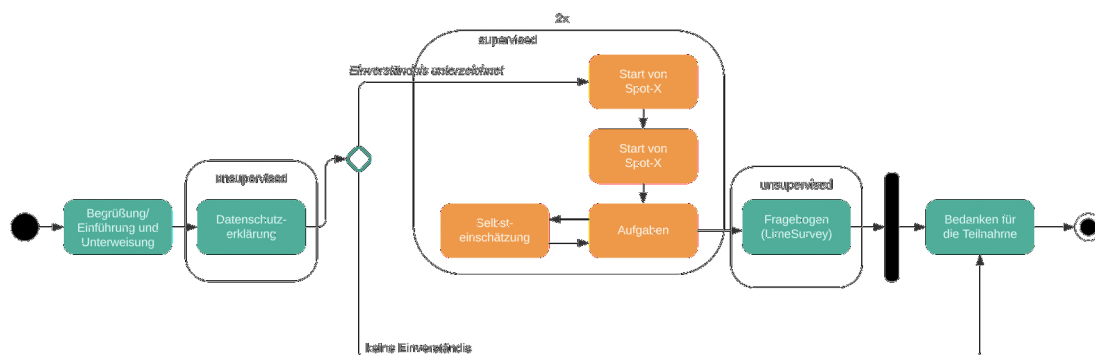


Abbildung 6.1: Ein Aktivitätsdiagramm zum Ablauf der Nutzerstudie

Spot-X Umfrage zur Nutzbarkeit

In dieser Umfrage werden Sie darum gebeten, mehrere Spieldurchläufe von Spot-X zu spielen. Anschließend werden Sie darum gebeten, einige Fragen zu beantworten. Ich freue mich über ihre Teilnahme und wünsche Ihnen viel Spaß!

In dieser Umfrage sind 8 Fragen enthalten.

Dies ist eine anonyme Umfrage.

In den Umfrageantworten werden keine persönlichen Informationen über Sie gespeichert, es sei denn, in einer Frage wird explizit danach gefragt.

Wenn Sie für diese Umfrage einen Zugangscode benutzt haben, so können Sie sicher sein, dass der Zugangsschlüssel nicht zusammen mit den Daten abgespeichert wurde. Er wird in einer getrennten Tabelle aufbewahrt und nur aktualisiert, um zu speichern, ob Sie diese Umfrage abgeschlossen haben oder nicht. Es gibt keinen Weg, die ZugangsCodes mit den Umfrageergebnissen zusammenzuführen.

Vielen Dank, dass Sie sich die Zeit nehmen, an unserer Studie teilzunehmen.

Einverständnis zur Teilnahme an der Studie „Nützlichkeit eines adaptiven Assistenzsystems für Spot-X“

1. Ich bin über das Forschungsvorhaben ausreichend informiert. Art, Umfang und Bedeutung der Studie, das Studienziel, die Studienlänge und weitere studienbedingte Erfordernisse wurden genau und verständlich dargelegt.
2. Ich hatte ausreichend Zeit, mich für oder gegen eine Studienteilnahme zu entscheiden und bin mir bewusst, dass die Teilnahme an der Studie freiwillig erfolgt.
3. Auf die Erfassung personenbezogener Daten wird verzichtet. Die Datenerfassung (z.B. Umfrageantworten) erfolgt in anonymisierter Form.
4. Es wird gewährleistet, dass alle gesammelten Daten nicht an Dritte weitergegeben werden. Bei der Veröffentlichung in einer wissenschaftlichen Zeitschrift wird aus den Daten nicht hervorgehen, wer an dieser Untersuchung teilgenommen hat. Die Erklärung zum Datenschutz können Sie hier einsehen.

Der Fragebogen startet, sobald Sie sich mit vorstehender Vorgehensweise einverstanden erklären und der Studienteilnahme zustimmen.

Um die Umfrage zu öffnen, akzeptieren Sie bitte unsere Datenschutzerklärung.

Weiter

Abbildung 6.2: Die Datenschutzerklärung, die der Nutzer unterzeichnen muss, bevor dieser am Rest der Studie teilnehmen darf.

Jeder Nutzer spielt genau zwei Durchläufe des Quizlernspiels Spot-X. Abbildung 6.3 zeigt den Versuchsaufbau und einen Nutzer während eines Spieldurchlaufs. Dabei besteht ein Spieldurchlauf aus zehn Fragen. Um die Hypothesen unter möglichst korrekten Bedingungen zu prüfen, wurden die Studienteilnehmer in zwei Gruppen unterteilt. Gruppe eins hat den ersten Spieldurchlauf ohne adaptiven Assistenten und somit ohne Anpassungen gespielt. Daraufhin erfolgt ein zweiter Durchlauf mit adaptiver Anpassung an den Nutzer. Teilnehmer der zweiten Gruppe spielen zunächst mit und den zweiten Durchlauf ohne Assistenz. Mit dieser Unterteilung der Gruppen können die Ergebnisse der Studie unabhängig von der Reihenfolge der Spieldurchläufe betrachtet werden. Beim Spieldurchlauf mit Assistenzsystem muss der Nutzer nach jeder Aufgabe einen Dialog zur Selbsteinschätzung (s. Abbildung 5.2) ausfüllen.

Nach den beiden Spieldurchläufen füllt der Nutzer einen Fragebogen aus (s. Anhang 7), in dem er zuerst allgemeine Fragen bezogen auf sein Alter, seinen Bildungsabschluss sowie seine Erfahrungen mit Videospiele bzw. Quizlernspielen beantwortet. Die restlichen Fragen orientieren sich an der „System Usability Scale“ (SUS) [Brooke u. a. 1996] und dem „NASA Task Load Index“ (NASA-TLX). Der SUS ist ein Fragebogen, mithilfe dessen sich die Nutzbarkeit eines Systems bewerten lässt. Der NASA-TLX stellt ein Werkzeug dar, welches die Bewertung/Beurteilung der wahrgenommenen Arbeitsbelastung ermöglicht. Diese beiden Konzepte werden in Abschnitt 6.2 ausgewertet und besprochen. Abschließend erhält der Nutzer im



Abbildung 6.3: Ein Foto von einem Nutzer während eines Spot-X Spieldurchlaufs

Rahmen der Umfrage die Möglichkeit, Feedback hinsichtlich des Spiels Spot-X sowie den Ablauf der Studie mitzuteilen. Im Rahmen dieser Thesis wurde entschieden, die Umfragen auf LimeSurvey ungeführt zu gestalten, um den Probanden die Freiheit für uneingeschränktes Feedback zu geben. Die Durchführung der Spieldurchläufe erfolgten jedoch, aufgrund von möglicherweise aufkommenden Unklarheiten und Fragen bei der Ausführung der Vorstudie, beaufsichtigt.

6.2 Ergebnisse

Im folgenden Abschnitt werden die Ergebnisse der Nutzerstudie vorgestellt. Dabei werden die Ergebnisse in zwei unterschiedliche Kategorien unterteilt: vor und nach der Korrektur der $Perf^P$ Berechnung. Daraufhin erfolgt eine Überprüfung und Interpretation der in Kapitel 5.1 aufgestellten Hypothesen. Abschließend werden mögliche Fehler und Verbesserungen der Studie diskutiert.

Das Anwendungsszenario wurde mit insgesamt zwölf Probanden vor und vier Probanden nach der Korrektur der Implementierung durchgeführt. In Abbildung 6.4a wird die Performance

des Nutzers mit und ohne dynamischer Anpassung vor der Korrektur verglichen. Analog für die vier Probanden nach der Korrektur in Abbildung 6.4b. Unter dynamischer Anpassung versteht man in dieser Arbeit die beschriebene Unterstützung des adaptiven Assistenten. Die x-Achse zeigt die kombinierte Performance im Intervall $[0,1]$. Auf der y-Achse werden die zehn Aufgaben gekennzeichnet. Für einen Nutzer wird jeweils mit einem blauen Kreise die Performance ohne adaptiven Assistenten dargestellt, während die roten Kreuze die Performance mit adaptiver Assistenz markieren. Das rote Kreuz zeigt die Performance mit adaptiver Assistenz an. Der Nutzer hat insgesamt zehn Aufgaben gespielt und die Werte in Aufgabe i entsprechen dem „Performance Score“, welcher bis zur i -ten Aufgabe akkumuliert wurde.

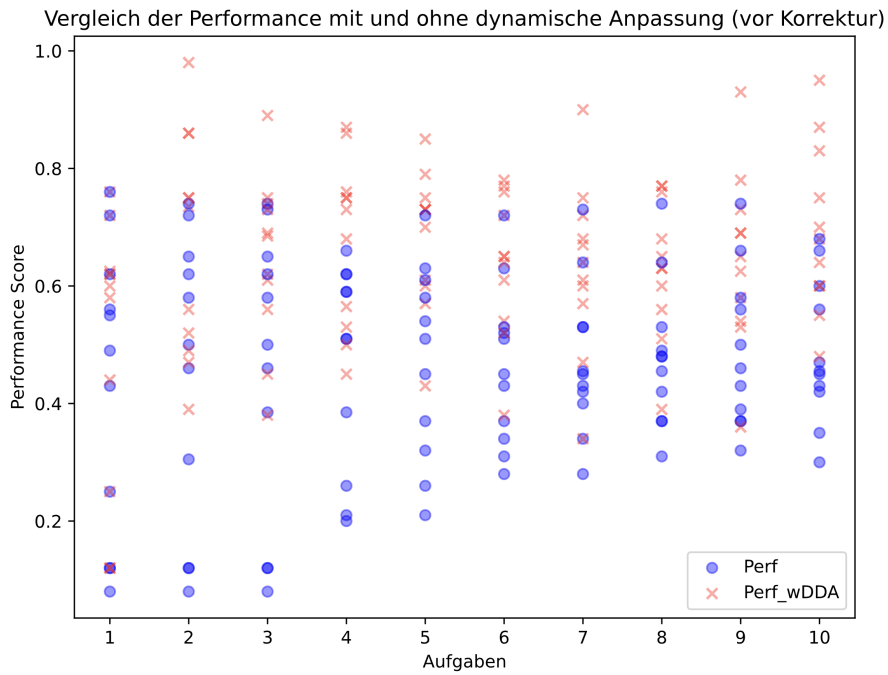
Im Folgenden werden die Hypothesen H_1 und H_0 , die in Abschnitt 5.1 besprochen wurden, analysiert. Hierfür nochmal die beiden Hypothesen:

H_0 : Keine Veränderung der kombinierten Performance durch die Nutzung der generierten Lernempfehlung.

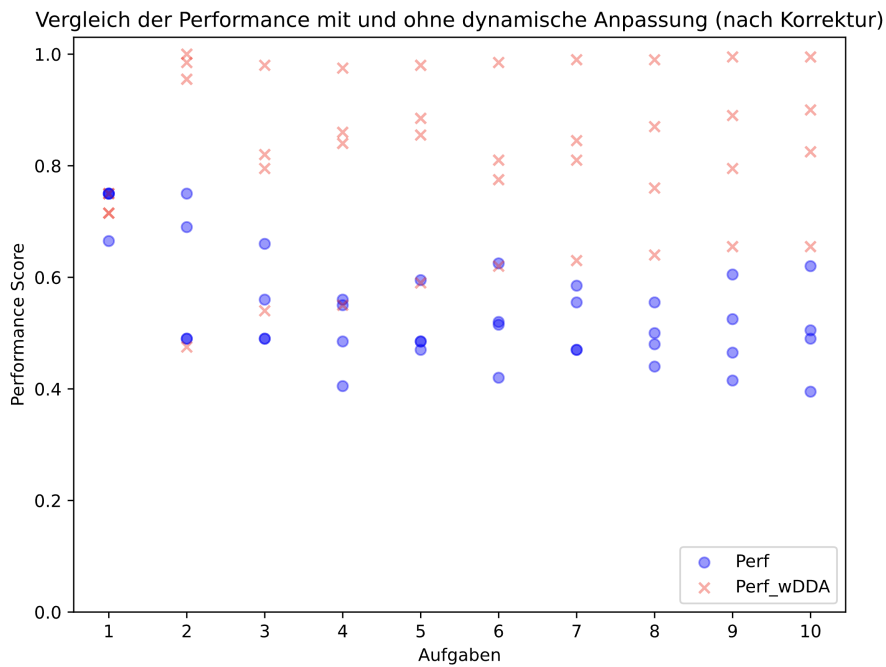
H_1 : Die kombinierte Performance steigt durch die Verwendung der generierten Lernempfehlung, welche durch den GPS bereitgestellt wird.

Da es sich hier um kleine Stichproben handelt, wäre es sinnvoller, den nicht-parametrischen Mann-Whitney-U-Test zu verwenden, der eine robuste Methode ist, um Unterschiede in der Verteilung der Daten zwischen zwei Gruppen zu testen. Daraufhin wurde anhand der vorliegenden Daten ein Mann-Whitney-U-Test für den Mittelwert zweier unabhängiger Proben durchgeführt [Mann-Whitney-U-Test 2023]. Dabei wurden die Daten zu zwei unterschiedlichen Zeitpunkten untersucht: vor und nach der Korrektur. Der Test zeigt, dass der „Performance-Score“ mit $p = 0.99$ bei der Verwendung mit adaptiver Assistenz höher ist als die Performance ohne. Dieses Ergebnis ist aufgrund der Größe der Stichprobe jedoch unter Vorbehalt zu betrachten und benötigt weitere Daten, um eine robuste Aussage treffen zu können. Daraufhin wurde eine Power-Analyse mit dem Signifikanzniveau $\alpha = 0.05$, der Effektgröße $\alpha = 0.5$ und der Power $p = 0.5$, um herauszufinden, welche Stichprobengröße erreicht werden muss, um H_0 ablehnen zu können. Diese Analyse ergab, dass die Stichprobengröße $n = 6.4$ sein muss, um die statistische Signifikanz von H_1 zu zeigen. Das bedeutet, dass die Signifikanz mit 7 Nutzern gezeigt werden kann.

Dennoch unterscheiden sich die Mittelwerte der beiden Kategorien, die Mittelwerte für die Performance vor der Korrektur sind $(Performance, Performance_{DDA}) = (4.5325, 6.3925)$ und nach der Korrektur $(Performance, Performance_{DDA}) = (5.43125, 8.1125)$. Wie es in den Abbildungen 6.4 zu erkennen ist, ist der Mittelwert der „Performance Score“ mit adaptiven Anpassungen nach der Korrektur deutlich gestiegen. Das Gleiche gilt auch für die Performance



(a) Vergleich der Durchläufe des Quizlernspiels Spot-X mit und ohne adaptives Assistenzsystem vor Korrektur der Implementierung mit $n = 12$ Nutzern.



(b) Vergleich der Durchläufe des Quizlernspiels Spot-X mit und ohne adaptives Assistenzsystem nach Korrektur der Implementierung mit $n = 4$ Nutzern.

Abbildung 6.4: Zwei Vergleiche der Performance in Abhängigkeit vom Assistenzsystem vor und nach der Korrektur.

ohne dynamische Anpassung jedoch, um einen geringeren Faktor. Die Ergebnisse nach der Korrektur weisen auf eine Verbesserung der Performance mit dynamischer Anpassung hin, welche jedoch noch weitere Untersuchungen bedarf. Allgemeiner spiegelt sich auch das Feedback der Probanden in den Diagrammen wider. Einige Nutzer haben rückgemeldet, dass die Fragen des Quizlernspiels zu einfach waren oder sich manche Fragen wiederholt haben. Dies lässt sich an der Dichte der Performance-Berechnungen im Bereich $y \approx 0.4$ erkennen.

Anschließend wurde folgende Hypothese untersucht:

H_2 : Die Selbsteinschätzung des Nutzers korreliert mit dem „Performance Score“.

Hierfür wurde im Rahmen dieser Studie die eingebaute Selbsteinschätzung (s. Abbildung 5.2) verwendet. Dabei wird die Korrelation mittels Bravais Pearson Korrelation untersucht [Benning 2020]. Für die Berechnung der Pearson Korrelation wird die Formel 6.1 verwendet.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (6.1)$$

Zu diesem Zweck wurde nach der Korrektur der Implementierung eine Korrelation Heatmap erstellt, welche in Abbildung 6.5 zu sehen ist. Hier wird die akkumulierte Selbsteinschätzung des Nutzers bis zur i -ten Aufgabe mit den Performance-Messungen $Perf^P$, $Perf^R$ sowie der kombinierten Performance verglichen. Auf den Diagonaleinträgen wird die Messung immer mit sich selbst verglichen. Daraus folgt, dass die Diagonaleinträge immer 1 sind, d.h. die Messungen auf der Diagonalen korrelieren positiv miteinander.

Wie zu erwarten war, korrelieren die $Perf^P$ und die $Perf^R$ mit der kombinierten Performance. Trotz $\alpha_1 = \alpha_2 = 0.5$ für die kombinierte Performance, korreliert die kombinierte Performance stärker mit der $Perf^P$. Aktuell wird als Lernziel der Graph-basierten Lernpfade für die $Perf^P$ Berechnung das Erreichen des Schwierigkeitsgrades schwer betrachtet. Das hat zur Folge, dass eine korrekte Antwort auf eine Aufgabe mit der Schwierigkeitskategorie schwer eine Performance berechnet und anschließend keine Änderungen mehr an dem Wert vornimmt, weil das Ziel bereits erreicht wurde. Das verursacht keine weiteren Änderungen im Wert $Perf^P$ und führt regelmäßig zu dem Performance-Wert 1. Deshalb wird die kombinierte Performance auch generell auf ≥ 0.5 gezogen und somit liegt generell leichter näher an $Perf^P$.

Weiterhin korreliert die $Perf^R$ stark mit der Selbsteinschätzung des Nutzers. Das weist darauf hin, dass es einen linearen Zusammenhang zwischen der Selbsteinschätzung und der $Perf^R$ gibt. Andererseits scheint $Perf^P$ weder mit der Selbsteinschätzung noch der $Perf^R$ zu korrelieren. Diese beiden Faktoren führen dazu, dass die kombinierte Performance nur leicht mit der Selbsteinschätzung korreliert.

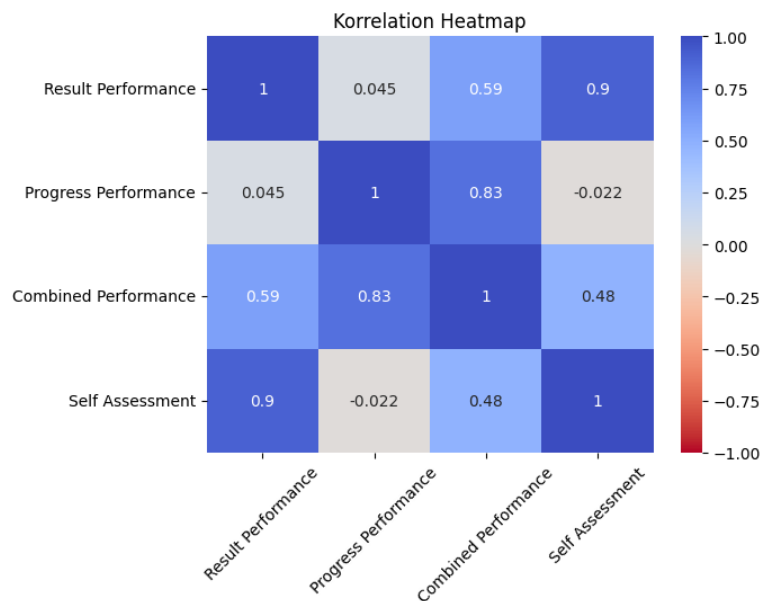


Abbildung 6.5: Pearson Korrelationen zwischen der Selbsteinschätzung des Nutzers und den verschiedenen Performance-Messungen dieser Arbeit mit $n = 4$ Nutzern nach der Korrektur. Das Diagramm wird durch eine kontinuierliche Skalierung im Intervall $[-1, 1]$ dargestellt. Blau bzw. 1 bedeutet indiziert eine positive Korrelation. Rot bzw. -1 hingegen weist auf eine negative Korrelation hin. Grau bzw. 0 bedeutet, dass keine Korrelation vorliegt.

Daraus ergibt sich, dass nur die $Perf^R$ Komponente der kombinierten Performance mit der Selbsteinschätzung des Nutzers korreliert. Dass die $Perf^P$ nicht mit der Selbsteinschätzung korreliert, bedeutet, dass es keinen Zusammenhang zwischen der Selbsteinschätzung und dem berechneten „Performance Score“ gibt. Aktuell wird als Lernziel der Graph-basierten Lernpfade für die $Perf^P$ Berechnung das Erreichen des Schwierigkeitsgrades schwer betrachtet. Dies hat zur Folge, dass eine korrekte Antwort auf eine Aufgabe mit der Schwierigkeitskategorie schwer eine Performance berechnet und anschließend keine Änderungen mehr an dem Wert vornimmt, da das Ziel bereits erreicht wurde. Dies verursacht keine weiteren Änderungen im Wert $Perf^P$ und führt zu keiner Korrelation zur Selbsteinschätzung.

Es folgt eine Übersichtstabelle, welche die formulierten Hypothesen und deren Ergebnisse zusammenfasst:

Hypothesen	erfüllt	Test bzw. Ergebnisse
H_1 : Die kombinierte Performance steigt durch die Verwendung der generierten Lernempfehlung, die durch den GPS bereitgestellt wird.	✓	Es wurden die Mittelwerte der Durchläufe mit und ohne adaptiven Assistenten verglichen. Anhand dessen konnte gezeigt werden, dass sich der „Performance-Score“ des Nutzers durch die adaptive Anpassung um 26.81 % mit einer Standardabweichung von 0.17 für eine Stichprobe $n = 4$ verbessert hat.
H_2 : Die Selbsteinschätzung des Nutzers korreliert mit dem „Performance Score“.	✓	Um die Korrelationen zwischen den Komponenten zu untersuchen wurde die Bravais Pearson Korrelation verwendet. Diese hat eine starke positive Korrelation (0.9) der $Perf^R$ mit der Selbsteinschätzung ergeben. Die $Perf^P$ korreliert in der angewendeten Variante nicht mit der Selbsteinschätzung, sodass die kombinierte Performance nur schwach mit der Selbsteinschätzung korreliert (0.48).
H_3 : Die Berücksichtigung der erreichten Punktzahl und der benötigten Zeit führt zu sinnvollen Performance Scores. Sinnvolle Scores bezeichnen hierbei einen Wert, der den Stereotypen (Anfänger, Intermediate, Experte) des Nutzers widerspiegelt.		Diese Hypothese wurde aufgrund von mangelnder Zeit nicht untersucht. Nach der Korrektur konnten vier Probanden untersucht werden. Diese Größe der Stichprobe ist nicht ausreichend, um die Nutzer in Klassen zu unterteilen.

6.3 Diskussion

In diesem Absatz werden mögliche Verbesserungen und Erweiterungen der Evaluation vorgestellt. Im Rahmen der Evaluation wurde deutlich, dass die $Perf^P$ weiterhin untersucht werden kann. Nach der Implementierung von Delaunay 2022, der den GPS mit der $Perf^P$ Berechnung entwickelt hat, funktionierte die Berechnung nur auf synthetischen Daten für das Quizlernspiel Spot-X. Nach den Korrekturen dieser Arbeit war die Berechnung auf Daten echter Nutzer

in Spot-X erfolgreich. Dennoch entsprechen die Ergebnisse nicht den Erwartungen, dass die $Perf^P$ mit der Selbsteinschätzung korreliert. Dies liegt aktuell an der Wahl des Lernziels für die Berechnung der idealen Pfade. Ein denkbarer Verbesserungsvorschlag besteht darin, einen kontinuierlichen Prozentsatz an korrekten schweren Antworten zu betrachten. Das bedeutet, dass der Nutzer für einen guten „Performance-Score“ einen Schnitt an korrekten Antworten halten muss. Eine weitere Möglichkeit zur Optimierung liegt in der Entkopplung der Performance-Berechnung bis zu einem gewissen Grad von der Schwierigkeitskategorie und der Untersuchung der Pfadlänge des Nutzers. Damit könnte man nicht nur den Erfolg der Aktivitäten auf einer Schwierigkeitskategorie belohnen, sondern auch ermitteln, wie häufig ein Nutzer in diese Kategorie eingestuft wurde und wie viele Aktivitäten er in dieser Kategorie absolviert hat. Es wäre interessant zu sehen, wie sich die Korrelationen durch eine Erweiterung der $Perf^P$ Berechnung ändert. Darüber hinaus stellt sich die Frage, wie sich die kombinierte Performance, durch eine Anpassung der Parameter α_1, α_2 ändert und damit ggf. die Korrelation zur Selbsteinschätzung.

Weiterhin wäre interessant zu beobachten, wie sich die Performance-Messungen bei einer größeren Auswahl an Aufgaben verhalten würden. Die Probanden haben die Rückmeldungen gegeben, dass sich die Aufgaben wiederholen oder auf eine eingeschränkte Bandbreite an Kategorien bezogen haben. Für ein breiteres Spektrum an Schwierigkeitskategorien wäre eine Unterteilung in Nutzerklassen, wie z.B. Experten, Intermediates und Anfänger denkbar.

7 Fazit und Ausblick

Das übergeordnete Ziel dieser Arbeit ist es, die Adaptivität der Assistenzsysteme für den E-Learning-Bereich zu verbessern. Konkret sollte eine umfassende Performance für einen Nutzer eines Serious Games oder einer E-Learning-Plattform bestimmt werden, um diese für die Lernfortschrittsoptimierung zu nutzen. Im Folgenden werden die Frage aus der Problemstellung aufgegriffen und durch die vorliegende Arbeit beantwortet:

- Wie kann die kombinierte Performance mittels PFA für Graphen und Graph-basierten Lernpfaden modelliert werden?

Hierfür wurde eine kombinierte Performance definiert, welche zwischen den beiden Komponenten Progress-Performance ($Perf^P$) sowie Result-Performance ($Perf^R$) differenziert. Für die Bestimmung der $Perf^R$ standen die beiden in der Literatur verbreiteten Modelle KT und PFA zur Auswahl. Eine umfangreiche theoretische Analyse zeigte auf, dass beide Modelle die gestellten Anforderungen, wie die gleichzeitige Performance-Berechnung mehrerer WKs oder eine vordefinierte Graphstruktur, nicht vollumfänglich erfüllen konnten. Aufgrund dessen wurde das auf PFA basierende GPFA Modell entwickelt. Dieser Entschluss wurde aufgrund der Vorteile von PFA gegenüber KT und den Eigenschaften Graph-basierter Modelle getroffen. Insbesondere kann PFA im Gegensatz zu KT eine Performance-Berechnung für komplexe Fragen mit einer Vielzahl von Wissenskomponenten durchführen. Hauptsächlich liegen die Vorteile Graph-basierter Modelle in der Nutzung natürlicher Gegebenheiten, der Beziehungen zwischen den Komponenten sowie der verständlichen Darstellung dieser Beziehungen. Bei der Implementierung der $Perf^R$ wurde das GPFA Modell um weitere Features ergänzt. Diese umfassen zum einen den Decay Factor, der zusätzlich Datenalterung berücksichtigt und zum anderen die „All Skill“-PFA, welche eine Berücksichtigung sämtlicher Wissenskomponenten des Quizlernspiels ermöglicht. Für die $Perf^P$ wurde das bereits erarbeitete Konzept der Graph-basierten Lernpfade von Delaunay 2022 vorgestellt und verwendet.

- Wie sieht so eine Modellierung für Assistenzsysteme allgemein aus?

Um die Nutzbarkeit für adaptive Assistenten zu vereinfachen, wurde formal eine allgemeine Schnittstelle definiert, welche die Kommunikation zwischen Graph Performance Service (GPS) und anderen E-Learning Systemen festlegt. Mit dem Ziel eines einheitlichen Standards bei der Modellierung wurden xAPI-Statements vorgestellt und verwendet.

- Wie sieht so eine Anwendung für E-Learning-Plattformen und Serious Games aus?

Darüber hinaus wurde eine Modellierung der kombinierten Performance für E-Learning-Plattformen und Serious Games realisiert, die den GPS, welcher die $Perf^P$ Berechnung bereits inkludierte, um die $Perf^R$ Berechnung zu erweitern. Das Modell wurde zur Evaluation im GPS umgesetzt. Es folgte eine Umsetzung notwendiger Funktionen für die Berechnung der kombinierten Performance, welche die Ausgabe einer Empfehlung der Schwierigkeitskategorie sowie die Überprüfung des „Performance Scores“ umfassen. Weiterhin mussten einige Korrekturen an der Modellierung der $Perf^P$ getroffen werden, da z.B. die Implementierung der synthetischen Daten nicht mit der theoretischen Umsetzung der Graph-basierten Lernpfade übereinstimmte.

- Wie sieht so eine Anwendung für ein konkretes Anwendungsbeispiel aus?

Zusätzlich wurden die Performance-Messungen für das Quizlernspiel Spot-X operationalisiert und eine praktische Implementierung der kombinierten Performance für ein konkretes Anwendungsbeispiel realisiert, um im Anschluss eine Evaluation des Ansatzes in Form einer Nutzerstudie durchzuführen. Abschließend erfolgte die Durchführung einer Nutzerstudie, bei der jeder Nutzer zwei Durchläufe des Quizlernspiels Spot-X absolvierte. Mithilfe der Nutzerstudie konnte eine Verbesserung des „Performance-Scores“ von durchschnittlich 26.8 % durch die adaptive Anpassung der Lerninhalte erzielt werden. Zusätzlich konnte eine positive Korrelation zwischen der Selbsteinschätzung und der $Perf^R$ festgestellt werden. Hingegen war keine Korrelation zwischen der Selbsteinschätzung und der $Perf^P$ oder der $Perf^R$ und der $Perf^P$ festzustellen. Der Grund hierfür ist die abweichende Umsetzung der Lernpfade zum theoretischen Konzept der $Perf^P$.

Zukünftige Arbeiten könnten den GPS durch Erweiterung und Optimierung der Umsetzung der $Perf^P$ Berechnung verbessern. Hierbei wäre insbesondere eine weiterführende, ausführlichere Evaluation interessant, da die Fehler in der Vorarbeit zu $Perf^P$ erst während der Evaluation auffielen, weshalb nicht alle Daten nutzbar sind. Daneben wäre es informativ,

die Beziehung zwischen der $Perf^P$ und der $Perf^R$ sowie der Selbsteinschätzung darzustellen. Folgende Arbeiten könnten die E-Learning-Plattformen ILIAS und Moodle erweitern, um die Generierung der nötigen xAPI-Statements und damit auch den Zugriff auf den GPS zu ermöglichen.

Anschließend an die vorliegende Arbeit wäre es darüber hinaus denkbar, Erweiterungen in der $Perf^R$ zu realisieren. Um eine Korrelation für die $Perf^R$ zu berechnen, wurde in dieser Thesis die Selbsteinschätzung verwendet. Hier wäre es interessant zu sehen, ob eine Möglichkeit besteht, die Selbsteinschätzung für die GPFA Berechnung zu nutzen. Beispielsweise könnte sich eine Abfrage nach der Selbsteinschätzung am Anfang des Systems positiv auf das Kaltstartproblem auswirken. Damit würde die Selbsteinschätzung als initialer Wissensstand fungieren. Weiterhin ist es möglich, dass die Selbsteinschätzung als ein Hinweis für Berechnungsfehler, wie z.B. der „Guess“ oder „Slip“ Parameter bei KT, dient. Daraus könnte eine weitere Nutzerstudie entstehen, in welcher das GPFA Modell dieser Arbeit, mit dem GPFA mit Selbsteinschätzung verglichen wird, um zu erfahren, ob sich das Einbinden der Selbsteinschätzung positiv auf die Performance mit adaptiven Anpassungen auswirkt.

Literatur

- Al-Ajlan, Ajlan u. a. (2008). „Why Moodle“. In: *2008 12th IEEE International Workshop on Future Trends of Distributed Computing Systems*, S. 58–64.
- Albers, Sönke u. a. (2009). *Methodik der empirischen Forschung*. Bd. 3. Springer.
- Alshammari, Mohammad u. a. (2014). „Adaptivity in E-Learning Systems“. In: *2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems*, S. 79–86.
- Athaya, Hisyam u. a. (2021). „Moodle Implementation for E-Learning: A Systematic Review“. In: *Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology*. SIET '21. Malang, Indonesia: Association for Computing Machinery, S. 106–112. URL: <https://doi.org/10.1145/3479645.3479646>.
- Baker, Ryan SJ d u. a. (2008). „More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing“. In: *International conference on intelligent tutoring systems*. Springer, S. 406–415.
- Bamberg, Günter u. a. (2012). *Statistik - Eine Einführung für Wirtschafts- und Sozialwissenschaftler*.
- Battaglia, Peter W. u. a. (2018). *Relational inductive biases, deep learning, and graph networks*. URL: <https://arxiv.org/abs/1806.01261>.
- Baur, Rafael (2022). „Graph Analytics für Performance Messungen und adaptive Lernpfade“. In: *KIT, Fraunhofer IOSB, Bachelor Thesis*.
- Beck, Joseph E. u. a. (2007). „Identifiability: A Fundamental Problem of Student Modeling“. In: *User Modeling 2007*. Hrsg. von Cristina Conati u. a. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 137–146.
- Becker, Torsten u. a. (2016). *Stochastische Risikomodellierung und statistische Methoden*. Springer.
- Benning, Valerie (Juli 2020). *Korrelationskoeffizient*. de-DE. URL: <https://www.scribbr.de/statistik/korrelationskoeffizient/> (besucht am 05. 05. 2023).
- Briskorn, Dirk u. a. (2020). *Graphentheorie*. Springer, S. 291–401.
- Brooke, John u. a. (1996). „SUS-A quick and dirty usability scale“. In: *Usability evaluation in industry* 189.194, S. 4–7.
- Brusilovsky, Peter (Mai 1996). „Adaptive hypermedia: An attempt to analyze and generalize“. In: S. 288–304.

- Brusilovsky, Peter u. a. (Jan. 1996). „ELM-ART: An intelligent tutoring system on world wide web“. In: S. 261–269.
- Cen, Hao, Kenneth Koedinger u. a. (2006). „Learning factors analysis—a general method for cognitive model evaluation and improvement“. In: *Intelligent Tutoring Systems: 8th International Conference, ITS 2006, Jhongli, Taiwan, June 26-30, 2006. Proceedings 8*. Springer, S. 164–175.
- Cen, Hao, Kenneth R. Koedinger u. a. (2007). „Is Over Practice Necessary? –Improving Learning Efficiency with the Cognitive Tutor through Educational Data Mining“. In: *Proceedings of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*. NLD: IOS Press, S. 511–518.
- Chatterjee, Samprit u. a. (2015). *Regression analysis by example*. John Wiley & Sons.
- Corbett, Albert T u. a. (1994). „Knowledge tracing: Modeling the acquisition of procedural knowledge“. In: *User modeling and user-adapted interaction 4*, S. 253–278.
- Csikszentmihalyi, Mihaly (2009). *Flow*. HarperCollins.
- De Sande, Brett van (2013). „Properties Of The Bayesian Knowledge Tracing Model.“ In: *Journal of Educational Data Mining 5.2*, S. 1–10.
- Delaunay, Alexis (2022). „Operationalisierung Graph-basierter Lernpfade für adaptive Serious Games“. In: *KIT, Fraunhofer IOSB, Bachelor Thesis*.
- Diestel, Reinhard (2017). *Extremal graph theory*. Springer, S. 173–207.
- Dörner, Ralf u. a. (2016). „Introduction“. In: *Serious Games: Foundations, Concepts and Practice*. Hrsg. von Ralf Dörner u. a. Cham: Springer International Publishing, S. 1–34. URL: https://doi.org/10.1007/978-3-319-40612-1_1.
- Falakmasir, Mohammad Hassan u. a. (2013). „A Spectral Learning Approach to Knowledge Tracing.“ In: *EDM*, S. 28–34.
- Falmagne, Jean-Claude u. a. (2013). *Knowledge spaces: Applications in education*. Springer Science & Business Media.
- Fang, Ying u. a. (2019). „A meta-analysis of the effectiveness of ALEKS on learning“. In: *Educational Psychology 39.10*, S. 1278–1292. eprint: <https://doi.org/10.1080/01443410.2018.1495829>. URL: <https://doi.org/10.1080/01443410.2018.1495829>.
- Frutos-Pascual, Maite u. a. (2017). „Review of the Use of AI Techniques in Serious Games: Decision Making and Machine Learning“. In: *IEEE Transactions on Computational Intelligence and AI in Games 9.2*, S. 133–152.
- Garland, Ron (1991). „The mid-point on a rating scale: Is it desirable“. In: *Marketing Bulletin*, S. 66–70.
- Göbel, Kerstin u. a. (2023). „Der Übergang zur digitalen Lehre an den Universitäten Duisburg-Essen und Basel in Zeiten der Corona-Pandemie“. In: *Wie Corona die Hochschullehre verändert:*

- Erfahrungen und Gedanken aus der Krise zum zukünftigen Einsatz von eLearning*. Springer, S. 365–391.
- Göbel, Stefan u. a. (2016). „Personalization and Adaptation“. In: *Serious Games: Foundations, Concepts and Practice*. Hrsg. von Ralf Dörner u. a. Cham: Springer International Publishing, S. 161–210. URL: https://doi.org/10.1007/978-3-319-40612-1_7.
- Gong, Yue, Joseph Beck u. a. (Jan. 2011). „How to Construct More Accurate Student Models: Comparing and Optimizing Knowledge Tracing and Performance Factor Analysis.“ In: *I. J. Artificial Intelligence in Education* 21, S. 27–46.
- Gong, Yue, Joseph E Beck u. a. (2010). „Comparing knowledge tracing and performance factor analysis by using multiple model fitting“. In: *ITS2010 Intelligent Tutoring Systems. LNCS 6094*, S. 35–44.
- Gross, Sebastian u. a. (2015). „Learning Feedback in Intelligent Tutoring Systems: Report of the FIT Project, Conducted from December 2011 to March 2015“. In: *KI-Künstliche Intelligenz* 29, S. 413–418.
- Gunderson, Lance H u. a. (2002). *Panarchy: understanding transformations in human and natural systems*. Island press.
- Guy, Rebecca F. u. a. (1977). „The Neutral Point on a Likert Scale“. In: *The Journal of Psychology* 95.2, S. 199–204.
- Han, Jun u. a. (1995). „The influence of the sigmoid function parameters on the speed of backpropagation learning“. In: *From Natural to Artificial Neural Computation: International Workshop on Artificial Neural Networks Malaga-Torremolinos, Spain, June 7–9, 1995 Proceedings* 3. Springer, S. 195–201.
- Hauger, David u. a. (2007). „State of the Art of Adaptivity in E-Learning Platforms.“ In: *LWA*, S. 355–360.
- Heiss, Raffael u. a. (Mai 2018). „Operationalisierung und Messung“. In: S. 63–78.
- ILIAS (Nov. 2022). de. Page Version ID: 227673301. URL: <https://de.wikipedia.org/w/index.php?title=ILIAS&oldid=227673301> (besucht am 04.04.2023).
- ILIAS (2023). URL: <https://www.ilias.de/open-source-lms-ilias/> (besucht am 04.04.2023).
- Ilias Documentation* (2023). URL: https://docu.ilias.de/goto_docu_wiki_wpage_5334_1357.html (besucht am 23.04.2023).
- Joshi, Ankur u. a. (2015). „Likert scale: Explored and explained“. In: *British journal of applied science & technology* 7.4, S. 396.
- Katsaris, Iraklis u. a. (2021). „Adaptive e-learning systems through learning styles: A review of the literature“. In: *Advances in Mobile Learning Educational Research* 1.2, S. 124–145.

- Kevan, Jonathan M u. a. (2016). „Experience API: Flexible, decentralized and activity-centric data collection“. In: *Technology, knowledge and learning* 21.1, S. 143–149.
- Kleinbaum, David G u. a. (2002). *Logistic regression*. Springer.
- Koedinger, Kenneth R u. a. (2012). „The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning“. In: *Cognitive science* 36.5, S. 757–798.
- Lane, H. Chad (2012). „Cognitive Models of Learning“. In: *Encyclopedia of the Sciences of Learning*. Hrsg. von Norbert M. Seel. Boston, MA: Springer US, S. 608–610. URL: https://doi.org/10.1007/978-1-4419-1428-6_241.
- Lang, Charles u. a. (2017). *Handbook of learning analytics*. SOLAR, Society for Learning Analytics und Research New York.
- Leszczenski, James M u. a. (2007). „What’s in a word? Extending learning factors analysis to model reading transfer“. In: *13th International Conference on Artificial Intelligence in Education, Educational Data Mining Workshop*.
- Liu, Qi, Zhenya Huang u. a. (2021). „EKT: Exercise-Aware Knowledge Tracing for Student Performance Prediction“. In: *IEEE Transactions on Knowledge and Data Engineering* 33.1, S. 100–115.
- Liu, Qi, Shuanghong Shen u. a. (2021). „A Survey of Knowledge Tracing“. In: *CoRR abs/2105.15106*. arXiv: 2105.15106. URL: <https://arxiv.org/abs/2105.15106>.
- Mann-Whitney-U-Test* (2023). de. URL: http://www.methodenberatung.uzh.ch/de/datenanalyse_spss/unterschiede/zentral/mann.html (besucht am 11. 05. 2023).
- Mao, Ye (2018). „Deep Learning vs. Bayesian Knowledge Tracing: Student Models for Interventions.“ In: *Journal of educational data mining* 10.2.
- Matayoshi, Jeffrey u. a. (2021). „Are we there yet? Evaluating the effectiveness of a recurrent neural network-based stopping algorithm for an adaptive assessment“. In: *International Journal of Artificial Intelligence in Education* 31.2, S. 304–336.
- Mirvis, Philip H (1991). *Flow: The psychology of optimal experience*.
- Moodle (Nov. 2018). de-DE. URL: <https://moodle.com/de/news/try-moodle-3-6-vor-der-zeit/> (besucht am 04. 04. 2023).
- Moodle Documentation* (2023). URL: https://moodle.org/plugins/logstore_xapi (besucht am 23. 04. 2023).
- Moodle LMS* (2023). de-DE. URL: <https://moodle.com/de/solutions/moodle-lms/> (besucht am 04. 04. 2023).
- Moon, T.K. (Dez. 1996). „The expectation-maximization algorithm“. In: *Signal Processing Magazine, IEEE* 13, S. 47–60.

- Nakagawa, Hiromi u. a. (2019). „Graph-based knowledge tracing: modeling student proficiency using graph neural network“. In: *IEEE/WIC/ACM International Conference on Web Intelligence*, S. 156–163.
- Needham, Mark u. a. (2019). *Graph algorithms: practical examples in Apache Spark and Neo4j*. O'Reilly Media.
- Neo4j (2023). en. URL: <https://neo4j.com/> (besucht am 12. 04. 2023).
- Ning, Huansheng u. a. (2021). „A Review on Serious Games in E-learning“. In: *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, S. 01–08.
- Nye, Benjamin D u. a. (2018). „SKOPE-IT (Shareable Knowledge Objects as Portable Intelligent Tutors): overlaying natural language tutoring on an adaptive learning system for mathematics“. In: *International journal of STEM education* 5, S. 1–20.
- Oproiu, Gabriela Carmen (2015). „A Study about Using E-learning Platform (Moodle) in University Teaching Process“. In: *Procedia - Social and Behavioral Sciences* 180. The 6th International Conference Edu World 2014 “Education Facing Contemporary World Issues”, 7th - 9th November 2014, S. 426–432. URL: <https://www.sciencedirect.com/science/article/pii/S187704281501486X>.
- Pardos, Zachary u. a. (Juni 2010). „T.: Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing“. In: S. 255–266.
- Pavlik Jr, Philip I u. a. (2009). „Performance Factors Analysis—A New Alternative to Knowledge Tracing.“ In: *Frontiers in Artificial Intelligence and Applications*.
- Pesce, David (2022). *Moodle Logstore*. https://github.com/xAPI-vle/moodle-logstore_xapi.
- Piech, Chris u. a. (2015). „Deep knowledge tracing“. In: *Advances in neural information processing systems* 28.
- Piotrowski, Michael (2010). „What is an e-learning platform?“ In: *Learning management system technologies and software solutions for online teaching: Tools and applications*. IGI Global, S. 20–36.
- Pu, Shi u. a. (2021). „Deep Performance Factors Analysis for Knowledge Tracing“. In: *Artificial Intelligence in Education*. Hrsg. von Ido Roll u. a. Cham: Springer International Publishing, S. 331–341.
- Pu, Yanjun u. a. (Dez. 2018). „Parallelizing Bayesian Knowledge Tracing Tool For Large-scale Online Learning Analytics“. In: S. 3245–3254.
- Pustovojtovskij, German (2021). „Anwendung von Regressionsmodellen für adaptive digitale Lernspiele in der Bildauswertung“. In: *KIT, Fraunhofer IOSB, Bachelor Thesis*.
- Radha, R u. a. (2020). „E-Learning during lockdown of Covid-19 pandemic: A global perspective“. In: *International journal of control and automation* 13.4, S. 1088–1099.

- Regression* (2023). en. URL: <https://www.javatpoint.com/linear-regression-vs-logistic-regression-in-machine-learning> (besucht am 15.04.2023).
- Ritterfeld, Ute u. a. (2009). *Serious games: Mechanisms and effects*. Routledge.
- Romero, C. u. a. (2007). „Educational data mining: A survey from 1995 to 2005“. In: *Expert Systems with Applications* 33.1, S. 135–146. URL: <https://www.sciencedirect.com/science/article/pii/S0957417406001266>.
- Romero, Cristobal u. a. (2020). „Educational data mining and learning analytics: An updated survey“. In: *WIREs Data Mining and Knowledge Discovery* 10.3, e1355. eprint: <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1355>. URL: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1355>.
- Sampayo-Vargas, Sandra u. a. (2013). „The effectiveness of adaptive difficulty adjustments on students’ motivation and learning in an educational computer game“. In: *Computers & Education* 69, S. 452–462. URL: <https://www.sciencedirect.com/science/article/pii/S0360131513001711>.
- Sapountzi, Anni u. a. (Aug. 2019). „Dynamic Knowledge Tracing Models for Large-Scale Adaptive Learning Environments“. In: 12, S. 93–110.
- Schmucker, Robin u. a. (2022). „Transferable Student Performance Modeling for Intelligent Tutoring Systems“. In: *CoRR* abs/2202.03980. arXiv: 2202.03980. URL: <https://arxiv.org/abs/2202.03980>.
- Serious Games* (März 2016). en-US. URL: <https://flowleadership.org/serious-games/> (besucht am 20.01.2021).
- Sharma, Neel u. a. (2017). „Adaptive learning in medical education: the final piece of technology enhanced learning?“ In: *The Ulster medical journal* 86.3, S. 198.
- Shute, Valerie u. a. (2018). „Adaptive e-learning“. In: *Educational psychologist*. Routledge, S. 105–114.
- Shute, Valerie J. u. a. (2012). „Adaptive Educational Systems“. In: *Adaptive Technologies for Training and Education*. Hrsg. von Paula J. Durlach u. a. Cambridge: Cambridge University Press, S. 7–27.
- Siemens, George (2013). „Learning analytics: The emergence of a discipline“. In: *American Behavioral Scientist* 57.10, S. 1380–1400.
- Stefan Schneider, Uwe Kohnle (2018). *ILIAS XapiCmi5 Plugin*. https://github.com/interne_tlehrer/XapiCmi5.
- Streicher, Alexander, Sebastian Leidig u. a. (2018). „Eye-Tracking for User Attention Evaluation in Adaptive Serious Games“. In: *Lifelong Technology-Enhanced Learning*. Hrsg. von Viktoria Pammer-Schindler u. a. Cham: Springer International Publishing, S. 583–586.

- Streicher, Alexander und Wolfgang Roller (2015). „Towards an interoperable adaptive tutoring agent for simulations and serious games“. In: *International Conference on Theory and Practice in Modern Computing, MCCSIS*, S. 194–197.
- Streicher, Alexander, Wolfgang Roller und Christian Biegemeier (2017). „Application of adaptive game-based learning in image interpretation“. In: *European Conference on Games Based Learning*. Academic Conferences International Limited, S. 975–978.
- Streicher, Alexander, Rainer Schönbein u. a. (2021). „Graph-Based Modeling for Adaptive Control in Assistance Systems“. In: *Advances in Artificial Intelligence, Software and Systems Engineering*. Hrsg. von Tareq Z. Ahram u. a. Cham: Springer International Publishing, S. 39–46.
- Streicher, Alexander und Jan D. Smeddinck (2016). „Personalized and Adaptive Serious Games“. In: *Entertainment Computing and Serious Games*. Hrsg. von Ralf Dörner u. a. Bd. 9970. Lecture Notes in Computer Science. Cham: Springer International Publishing, S. 332–377.
- IBM (2023). *Was sind rekurrente neuronale Netze?* | IBM. de-de. URL: <https://www.ibm.com/de-de/topics/recurrent-neural-networks> (besucht am 17. 04. 2023).
- Weber, Gerhard u. a. (2016). „Elm-art—an interactive and intelligent web-based electronic textbook“. In: *International Journal of Artificial Intelligence in Education* 26, S. 72–81.
- Wiemeyer, Josef u. a. (2016). „Performance assessment in serious games“. In: *Serious games: foundations, concepts and practice*, S. 273–302.
- Wirth, Werner u. a. (Jan. 2008). *Die Brücke zwischen Theorie und Empirie: Operationalisierung, Messung und Validierung in der Kommunikationswissenschaft*.
- xAPI (2023). en-US. URL: <https://xapi.com/overview/> (besucht am 06. 04. 2023).
- Yadav, Vivek u. a. (Sep. 2020). „User Assistance for Serious Games Using Hidden Markov Model“. In: S. 380–385.
- Yen, Jin Y (1971). „Finding the k shortest loopless paths in a network“. In: *management Science* 17.11, S. 712–716.
- Zhang, Jiani u. a. (2017). „Dynamic key-value memory networks for knowledge tracing“. In: *Proceedings of the 26th international conference on World Wide Web*, S. 765–774.

Abkürzungsverzeichnis

<i>Perf^P</i>	Progress-Performance. 3, 4, 22, 24–30, 44, 55–57, 59, 65–70, 72–77, 81, 84–87, 89–91,
<i>Perf^R</i>	Result-Performance. 3, 4, 22, 23, 26, 30, 31, 37, 38, 40, 44, 45, 55–59, 65–70, 72, 73, 77, 84–86, 89–91,
BKT	Bayesian Knowledge Tracing. 32, 34, 41, 45, 46, 48, 50,
DKT	Deep Knowledge Tracing. 46, 47, 50,
DKVMN KT	Dynamic Key-Value Memory Networks für Knowledge Tracing. 47, 51,
DPFA	Deep Performance Factor Analysis. 50, 51,
ELAI	E-Learning Artificial Intelligence. 4, 5, 8, 18, 19, 70, 72, 73,
GKT	Graph Knowledge Tracing. 23, 47, 49, 51, 52,
GPFA	Graph Performance Factor Analysis. v, 3, 4, 45, 52, 57–59, 67–70, 73, 77, 89, 91,
HMM	Hidden Markov Modell. 32, 33, 45,
KT	Knowledge Tracing. v, 1–3, 5, 18, 31–34, 38, 39, 45–48, 50, 89, 91,
LFA	Learning Factor Analysis. 5, 23, 24, 34, 35, 37–40, 46, 48, 50,
LRS	Learning Record Store. 18, 20, 55, 59, 60, 62,

PFA	Performance Factor Analysis. v, 1-3, 5, 11, 18, 22, 24, 31, 34, 35, 38-41, 46, 48, 50-52, 57-59, 69, 78, 89,
SG	Serious Games. 2, 4, 5, 7-13, 18-20, 22, 27, 31, 43, 50-52, 55, 60-62, 65, 70-72, 77-79, 89, 90,
WK	Wissenskomponente. 30-34, 37-40, 44-48, 50, 57-61, 67-69, 72, 77, 78, 89,
xAPI	Experience API. 2, 4, 18-22, 28, 30, 31, 55, 57, 59-63, 65-67, 71, 73, 77, 78, 90, 91,

Spot-X Umfrage zur Nutzbarkeit

In dieser Umfrage werden Sie darum gebeten, mehrere Spieldurchläufe von Spot-X zu spielen. Anschließend werden Sie darum gebeten, einige Fragen zu beantworten. Ich freue mich über ihre Teilnahme und wünsche Ihnen viel Spaß!

In dieser Umfrage sind 24 Fragen enthalten.

Allgemeine Fragen

Haben Sie bereits Videospiele gespielt? (PC, Konsole, Mobile)

Bitte wählen Sie nur eine der folgenden Antworten aus:

- Ja
 Nein

Haben Sie bereits Erfahrungen mit Quizlernspielen?

Bitte wählen Sie nur eine der folgenden Antworten aus:

- Ja
 Nein

Wie alt sind Sie?

i In dieses Feld dürfen nur Zahlen eingegeben werden.

Bitte geben Sie Ihre Antwort hier ein:

Welchen Bildungsabschluss haben Sie oder planen Sie zu erreichen?

🗳 Bitte wählen Sie eine der folgenden Antworten:

Bitte wählen Sie nur eine der folgenden Antworten aus:

- Schulabschluss (Gymnasium, Realschule, Werkrealschule, Hauptschule, o.Ä.)
- Ich verfüge über keinen Abschluss
- höherer Abschlussgrad (PhD, Professur, ...)
- akademischer Abschluss (Bachelor, Master, Diplom, Fachhochschulreife, Ausbildung)

Benutzbarkeit von Spot-X

Im Folgenden werden Fragen zum Quizlernspiel Spot-X gestellt. Markieren Sie dazu bitte auf den folgenden Skalen, in welchem Maße Sie der Benutzbarkeit des Quizlernspiels zustimmen oder nicht zustimmen. *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1	2	3	4	5
Ich kann mir sehr gut vorstellen, das System regelmäßig zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich empfinde das System als unnötig komplex.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich empfinde das System als einfach zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich denke, dass ich technischen Support brauchen würde, um das System zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich finde, dass die verschiedenen Funktionen des Systems gut integriert sind.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich finde, dass es im System zu viele Inkonsistenzen gibt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann mir vorstellen, dass die meisten Leute das System schnell zu beherrschen lernen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich empfinde die Bedienung als sehr umständlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	1	2	3	4	5
Ich habe mich bei der Nutzung des Systems sehr sicher gefühlt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich musste eine Menge Dinge lernen, bevor ich mit dem System arbeiten konnte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Die Skalierung ist wie folgt:

1 - Stimme überhaupt nicht zu

5 - Stimme voll zu

Ich kann mir sehr gut vorstellen, das System regelmäßig zu nutzen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Ich empfinde das System als unnötig komplex.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Ich empfinde das System als einfach zu nutzen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Ich denke, dass ich technischen Support brauchen würde, um das System zu nutzen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Ich finde, dass die verschiedenen Funktionen des Systems gut integriert sind.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Ich finde, dass es im System zu viele Inkonsistenzen gibt.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Ich kann mir vorstellen, dass die meisten Leute das System schnell zu beherrschen lernen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Ich empfinde die Bedienung als sehr umständlich.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Ich habe mich bei der Nutzung des Systems sehr sicher gefühlt.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Ich musste eine Menge Dinge lernen, bevor ich mit dem System arbeiten konnte.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Stresslevel und kognitive Last (NASA-TLX)

Geben Sie jetzt für jede der unten stehenden Dimensionen an, wie hoch die Beanspruchung war. Markieren Sie dazu bitte auf den folgenden Skalen, in welchem Maße Sie sich in den sechs genannten Dimensionen von der Aufgabe beansprucht oder gefordert gesehen haben: *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1	2	3	4	5
<p>Geistige Anforderungen</p> <p>Wie viel geistige Anstrengung war bei der Informationsaufnahme und -verarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erforderte sie hohe Genauigkeit oder war sie fehlertolerant?</p>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<p>Körperliche Anforderungen</p> <p>Wie viel körperliche Aktivität war erforderlich (z.B. Ziehen, Drücken, Drehen, Steuern, Aktivieren,...)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholsam oder mühselig?</p>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	1	2	3	4	5
Zeitliche Anforderungen Wie viel Zeitdruck empfanden Sie hinsichtlich der Häufigkeit oder dem Takt, mit dem Aufgaben oder Aufgabenelemente auftraten? War die Abfolge langsam und geruhsam oder schnell und hektisch?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Leistung Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Anstrengung Wie hart mussten sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Frustration Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Die Skalierung ist wie folgt:

1 - Stimme überhaupt nicht zu

5 - Stimme voll zu

Geistige Anforderungen

Wie viel geistige Anstrengung war bei der Informationsaufnahme und -verarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erforderte sie hohe Genauigkeit oder war sie fehlertolerant?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Körperliche Anforderungen

Wie viel körperliche Aktivität war erforderlich (z.B. Ziehen, Drücken, Drehen, Steuern, Aktivieren,...)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholsam oder mühselig?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Zeitliche Anforderungen

Wie viel Zeitdruck empfanden Sie hinsichtlich der Häufigkeit oder dem Takt, mit dem Aufgaben oder Aufgabenelemente auftraten? War die Abfolge langsam und geruhsam oder schnell und hektisch?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Leistung

Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht?
Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Anstrengung

Wie hart mussten sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Frustration

Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

Feedback

In folgendem Freitextfeld haben Sie die Möglichkeit noch eigene Anmerkungen zum Quizlernspiel Spot-X zu schreiben.

Bitte geben Sie Ihre Antwort hier ein:

z.B. Geschwindigkeit, Größe der Fragen, Übersichtlichkeit, generelle Erfahrung,...

In folgendem Freitextfeld haben Sie die Möglichkeit noch eigene Anmerkungen zur Durchführung der Nutzerstudie zu schreiben.

Bitte geben Sie Ihre Antwort hier ein:

z.B. Freundlichkeit, Interaktion, Tipps für die nächste Nutzerstudie, Verbesserungsvorschläge, ...

Geschafft! Vielen Dank für deine Teilnahme!

Übermittlung Ihres ausgefüllten Fragebogens:
Vielen Dank für die Beantwortung des Fragebogens.

Spot-X Umfrage zur Nutzbarkeit

3 | 2023-03-20 12:13:09 | 4 | de | 1084604568 | 2023-03-20 12:01:42 | 2023-03-20 12:13:09 | 24 | AO10 | Y | Y | 1 | 1 | 5 | 2 | 4 | 3 | 5 | 2 | 5 | 1 | 3 | 1 | 2 | 4 | 3 | 2 | Verwirrend, dass zwei verschiedene Sprachen verwendet wurden (Englisch, Deutsch) Aufgabenstellung z.T. unklar (besonders Freitextaufgaben) | "Stresslevel & kognitive Last": verwirrende Skalierung der Antwortmöglichkeiten (unpassend zu Fragen) |

4 | 2023-03-28 19:29:50 | 4 | de | 462706006 | 2023-03-28 19:27:26 | 2023-03-28 19:29:50 | 53 | AO10 | N | N | 5 | 2 | 4 | 2 | 3 | 1 | 5 | 2 | 4 | 1 | 4 | 1 | 2 | 4 | 3 | 5 | Aufgaben nicht ganz klar | |

5 | 2023-04-03 11:07:38 | 4 | de | 372837696 | 2023-04-03 11:00:11 | 2023-04-03 11:07:38 | 28 | AO10 | Y | Y | 3 | 2 | 4 | 5 | 4 | 4 | 1 | 2 | 3 | 2 | 3 | 1 | 2 | 3 | 3 | 3 | Fragestellung, zu sehr auf bestimmte Spielkategorien bezogen | top |

6 | 2023-04-04 08:02:09 | 4 | de | 649740721 | 2023-04-04 07:56:06 | 2023-04-04 08:02:09 | 31 | AO10 | Y | Y | 2 | 1 | 3 | 2 | 5 | 3 | 4 | 2 | 4 | 1 | 4 | 2 | 2 | 4 | 3 | 3 | Ladezeit der neuen Fragen, Fragestellung unpassend zu Bild, Fragestellung unklar bei Videospielcharakter | |

7 | 2023-04-09 10:18:06 | 4 | de | 980332284 | 2023-04-09 10:10:16 | 2023-04-09 10:18:06 | 26 | AO10 | Y | Y | 4 | 1 | 5 | 2 | 4 | 1 | 5 | 1 | 5 | 1 | 4 | 3 | 2 | 3 | 3 | 3 | Fragestellungen hinzufügen | Im Voraus noch sagen, dass auf die Nutzerfreundlichkeit geachtet werden soll |

8 | 2023-04-09 11:29:54 | 4 | de | 573946866 | 2023-04-09 11:26:17 | 2023-04-09 11:29:54 | 24 | AO10 | N | Y | 3 | 1 | 5 | 1 | 4 | 1 | 5 | 1 | 1 | 1 | 4 | 1 | 2 | 4 | 1 | 3 | ist Übersichtlich hätten mehr unterschiedliche Fragen sein können | German bester Mann |

9 | 2023-04-10 10:41:30 | 4 | de | 1673654307 | 2023-04-10 10:37:16 | 2023-04-10 10:41:30 | 20 | AO10 | Y | Y | 1 | 2 | 4 | 1 | 3 | 3 | 5 | 2 | 4 | 1 | 3 | 1 | 3 | 3 | 3 | 3 | Langsam | Hilfsbereit |

10 | 2023-04-10 11:26:45 | 4 | de | 1397211687 | 2023-04-10 11:22:27 | 2023-04-10 11:26:45 | 23 | AO10 | Y | Y | 4 | 1 | 5 | 1 | 5 | 2 | 5 | 1 | 5 | 1 | 3 | 1 | 2 | 4 | 3 | 3 | Das Laden zur nächsten Frage könnte ein bisschen schneller sein. Sonst alles super | Alles wurde ausführlich erklärt und Fragen zu jeder Zeit beantwortet. |

11 | 2023-04-10 13:53:04 | 4 | de | 557390458 | 2023-04-10 13:50:46 | 2023-04-10 13:53:04 | 27 | AO10 | Y | Y | 1 | 1 | 5 | 1 | 3 | 2 | 5 | 1 | 5 | 1 | 2 | 1 | 2 | 4 | 2 | 1 | | |

13 | 2023-04-12 17:33:29 | 4 | de | 142425373 | 2023-04-12 17:32:14 | 2023-04-12 17:33:29 | 22 | AO10 | Y | Y | 3 | 1 | 4 | 1 | 3 | 2 | 5 | 1 | 4 | 2 | 3 | 1 | 2 | 5 | 2 | 1 | | |

14 | 2023-04-13 20:45:18 | 4 | de | 843829104 | 2023-04-13 20:41:39 | 2023-04-13 20:45:18 | 27 | AO10 | Y | Y | 5 | 1 | 5 | 1 | 5 | 3 | 5 | 1 | 5 | 1 | 2 | 1 | 1 | 4 | 2 | 3 | mehr Fragen | |

15 | 2023-04-15 18:54:11 | 4 | de | 1615049585 | 2023-04-15 18:51:00 | 2023-04-15 18:54:11 | 25 | AO10 | Y | Y | 2 | 1 | 5 | 1 | 4 | 2 | 5 | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 3 | 4 | | |

16 | 2023-04-15 21:12:00 | 4 | de | 80207496 | 2023-04-15 21:09:12 | 2023-04-15 21:12:00 | 26 | AO10 | Y | Y | 1 | 1 | 5 | 1 | 1 | 5 | 5 | 1 | 5 | 1 | 3 | 1 | 1 | 4 | 2 | 1 | Fragestellung unklar | |

17 | 2023-04-27 00:50:05 | 4 | de | 473564247 | 2023-04-27 00:47:22 | 2023-04-27 00:50:05 | 32 | AO09 | Y | Y | 2 | 1 | 4 | 1 | 3 | 1 | 5 | 1 | 4 | 1 | 2 | 1 | 2 | 4 | 1 | 2 | mehr Fragen, Fragen sehr spezielle Kategorien (z.B. Star Wars und andere Filme). Mehr Mathematikaufgaben oder Geografie wären besser gewesen. | |

18 | 2023-04-27 07:41:36 | 4 | de | 1583737841 | 2023-04-27 07:38:28 | 2023-04-27 07:41:36 | 22 | AO10 | Y | N | 3 | 1 | 4 | 1 | 3 | 2 | 5 | 2 | 5 | 1 | 1 | 1 | 1 | 3 | 1 | 3 | mehr Fragen (teilweise wiederholen sich Fragen) Fragestellung manchmal unklar weniger Fragen zu Spielen | hilfsbereit :-) |

19 | 2023-04-27 08:20:26 | 4 | de | 397749903 | 2023-04-27 08:17:36 | 2023-04-27 08:20:26 | 27 | AO10 | Y | Y | 4 | 2 | 4 | 1 | 2 | 4 | 4 | 2 | 5 | 1 | 3 | 1 | 2 | 4 | 1 | 1 | Mehr Aufgaben, zweiter Durchlauf, ein paar Wiederholungen | Alles gut |