

Wissenschaftlichkeit vs. Praxisbezug

Das Vier-Komponenten Instructional Design-Modell zur Entwicklung eines innovativen Data Science Studiengangs

Simone Opel, Christian Beecks, Andrea Linxen, Karin Elbrecht

Abstract: Datengetriebene Anwendungen nehmen einen immer größeren Raum in Wissenschaft und Technik ein. Dafür ihre Entwicklung und Pflege eine Vielzahl von Spezialist:innen benötigt wird, bietet auch die FernUniversität in Hagen einen Master-Studiengang Data Science an. Wie sich zeigt, ist es für viele Studierenden schwierig, die erlernten Fähigkeiten und wissenschaftlichen Konzepte zur Lösung datenbasierter Probleme in einer Anwendungsdomeine anzuwenden. Um diese Kompetenzen zu fördern, wird das Modul »Einführung in Data Science« lernendenzentriert weiterentwickelt. Neben einer strikten Ausrichtung an einem Data Science Life Cycle werden die Lehr-Lern-Materialien im Sinne des 4C/ID-Modells entwickelt. In diesem Beitrag werden der Entwicklungsprozess und die exemplarische Umsetzung in Lehr-Lern-Szenarien vorgestellt.

Data-driven solutions are becoming increasingly important in science and technology. As many specialists are required for its development and maintenance, the FernUniversität in Hagen also offers a master's programme in Data Science. As it turns out, many students have difficulties applying the skills and scientific concepts they have acquired to solve data-based problems in a domain of application. To promote these skills, the »Introduction to Data Science« module is being further developed in a learner-centred way. In addition to a strict alignment with a data science life cycle, the teaching-learning materials are being developed in accordance with the 4C/ID model. This article presents the development process and exemplary learning and teaching material.

Keywords: Data Science; Asynchrone Lehre; Instructional Design; Data Science Life Cycle; 4C/ID-Modell; Lehr-Lern-Szenarien; Distance Learning; Instructional Design; Teaching-learning Scenarios

1. Einleitung

Aktuelle Technologien und KI-Innovationen basieren in immer höherem Maße auf einer Vielzahl unterschiedlicher Daten, die sach- und fachgerecht gesammelt, vorbereitet, analysiert und anschließend anwendungsbezogen zur Verfügung gestellt und genutzt werden. Die hierfür notwendigen Kompetenzen stammen aus den Disziplinen Mathematik, Statistik sowie Informatik, aber auch aus der jeweiligen Anwendungsdomäne.

Um den Anforderungen gerecht zu werden, komplexe Probleme zu erkennen und datenbasiert lösen zu können, bieten inzwischen viele Hochschulen und Universitäten unterschiedlich ausgestaltete Data Science-Studiengänge¹ an, die ihre Absolventen dazu befähigen sollen, datengetriebene Lösungen zu entwickeln und einzusetzen.

An der FernUniversität in Hagen können Studierende seit dem Wintersemester 2022/23 einen *Master-Studiengang Data Science*² als Fernstudium durchführen, in dessen Rahmen sie die notwendigen Kompetenzen erwerben können. Der Studiengang der FernUniversität besteht aus vollständig digitalen, asynchronen Lehr-Lern-Szenarien, die durch synchrone Elemente wie Videosprechstunden ergänzt werden. Die Studierenden des Studiengangs unterscheiden sich in vielen Aspekten von Studierenden an Präsenzhochschulen. So liegt das Durchschnittsalter der Studierenden bei 41 Jahren (WS2023/24; Altersbereich von 22 bis 71 Jahren; eigene Erhebung), wobei die Mehrheit nebenberuflich studiert und neben hoher intrinsischer Motivation oft schon praktische Erfahrungen in der Auswertung großer Datenmengen hat.

Eines der grundlegenden Module des Studiengangs stellt *Einführung in Data Science* dar, das von den Studierenden in der Regel als eines der ersten Module belegt wird und für viele den ersten Kontakt mit hochschulischem Lernen seit langem darstellt. Das Lehr-Lern-Material besteht aus Lehrvideos, Folien, interaktiven Aufgaben, Miro-Boards, Fachtexten und online umzusetzenden Programmieraufgaben in einer Moodle-Umgebung, was von den Studierenden als sinnvoll, aber auch anspruchsvoll wahrgenommen wird. Angeregt durch das Feedback der Studierenden und der Erkenntnisse der Begleitforschung wurde daher entschieden, das Modul weiterzuentwickeln, um die Online-Lehr-Lernszenarien noch lernendenzentrierter zu gestalten und die Verbindung zwischen der Wissenschaftlichkeit des universitären Studiums und beruflicher Praxis für die Studierenden sichtbarer zu machen.

¹ Stand Dez. 2023: 78 Studiengänge »Data Science« oder Schwerpunkt Data Science, davon 32 Master-Studiengänge (eigene Erhebung, Quelle: <https://www.studycheck.de/>)

² https://www.fernuni-hagen.de/mi/studium/msc_datascience/index.shtml

2. Fachliche und didaktische Aspekte

Das in diesem Beitrag beschriebene Modul *Einführung in Data Science* beinhaltet einige Herausforderungen sowohl für Studierende als auch für Lehrende. Für viele stellt es den ersten Kontakt mit einem Fernstudium dar, auch ist für eine Vielzahl von Studierenden dies das erste Studium nach einer Phase der Berufstätigkeit. Da zudem rund 75 % (Hochschulstatistik, 2024) der Studierenden berufliche oder Care-Verpflichtungen haben, verfügen sie gegenüber Vollzeitstudierenden nur über ein eingeschränktes Zeitbudget, das die meisten von ihnen versuchen effektiv zu nutzen, weshalb sie auf Lehr-Lern-Material, das ihren Bedürfnissen gerecht wird, angewiesen sind.

Das Modul *Einführung in Data Science* gibt den Studierenden einen umfassenden Überblick über die Wissenschafts- und Fachdisziplin und dient dazu, später zu erwerbende vertiefte Kenntnisse im Fach einordnen zu können.

Hierzu gehört auch, die schon einigen Lernenden bekannten praktischen Anwendungen der zu erlernenden Konzepte und Methoden mit ihrer wissenschaftlichen Fundierung zu verknüpfen. Dieser Zusammenhang ist im bisherigen Lehr-Lern-Material für die Studierenden oft nicht ausreichend erkennbar, wie die Rückmeldungen während der regelmäßigen Video- und Chatsprechstunden zeigten. Zudem ist es mit diesem Material noch nicht ausreichend möglich, eine Binnendifferenzierung oder Adaptierung nach Vorkenntnissen und Lernverhalten zu gestalten.

2.1 Grobkonzept

Während bisher das Lernmaterial in thematische Einheiten strukturiert wurde und damit die einzelnen Elemente zwar fachlich, aber nicht im Hinblick auf die Prozessabläufe eines Data Science-Projekts aufeinander bezogen sind, sollen im neu konzipierten Material alle Kompetenzen durch das Erarbeiten eines vollständigen Data Science-Prozesses erworben werden. Dies bedingt jedoch eine enge Verzahnung mit einem durchgehenden didaktischen Konzept, das für die Anwendung in digitalen, asynchronen Lehr-Lern-Szenarien geeignet ist. Ein geeignetes Konzept sollte die Lehrenden bei der Entwicklung neuer und Überarbeitung existierender Materialien in einer Weise unterstützen, dass verschiedene Niveaustufen, eine Mischung aus informierendem und Übungsmaterial sowie die Begleitung durch einen vollständigen Prozess abgebildet werden können.

Ziel dieses Projekts zur Weiterentwicklung des Moduls *Einführung in Data Science* ist damit, Lehr-Lern-Szenarien zu entwickeln, die es den Studierenden erlauben, berufliche Handlungskompetenz im Bereich der Umsetzung von Data Science Projekten zu entwickeln. Berufliche Handlungskompetenz wird in diesem Zusammenhang (vgl. Rauner, 2004) als domänenspezifisches Professions- und Handlungswis-

sen verstanden, das durch Kennen grundlegender Fachkonzepte und deren Anwendung in realen beruflichen Handlungssituation erlangt wird.

2.2 Der Data Science Life Cycle

Data Science Projekte werden häufig entlang eines sog. *Data Science Life Cycles* entwickelt, der sich von der Erfassung der benötigten Daten über die Vorbereitung und Analyse bis hin zum praktischen Einsatz und ihrer eventuellen Nachnutzung erstreckt (vgl. bspw. Berman et al., 2018 oder Stodden, 2020). Um Anforderungen von Praxisprojekten in Aufgaben des Data Mining und damit der Data Science zu übersetzen, wurden in den letzten Jahren hieraus *Vorgehens- und Prozessmodelle* mit unterschiedlichen Schwerpunkten entwickelt.

Eines der wichtigsten ist der sog. CRISP-DM³ (Wirth & Hipp, 2000). In diesem Prozessmodell werden die Schritte vom Business Understanding bis zum Deployment der entwickelten Modelle mit möglichen Zyklen und Artefakten beschrieben. Wegen seiner hohen praktischen Relevanz wird dieses Modell in der aktuellen Version des Moduls vorgestellt, aber nicht eingesetzt. Neben weiteren spezielleren Verfahren wie beispielsweise dem *Knowledge Discovery in Databases* (KDD) (Fayyad et al., 1996) bietet sich ein weiteres, junges Prozessmodell an. Das DASC-PM⁴ (Schulz et al., 2020) an hat den Vorteil, dass die einzelnen Prozessschritte direkt aus Schlüsselbereichen der Data Science entwickelt wurden und sich damit eine enge Verbindung der praktischen Anwendung in einer Domäne zur Bezugswissenschaft und Forschung des Bereichs zeigt. Das Modell besteht aus fünf iterativ aufeinander aufbauenden Schritten (Projektauftrag, Datenbereitstellung, Analyse, Nutzbarmachung und Nutzung), die beliebig oft wiederholt werden können. Im Gegensatz zu manch anderen Prozess- und Vorgehensmodellen sind hier der Projektauftrag sowie die Nutzung des entwickelten Artefakts nicht genuiner Bestandteil, werden im Rahmen des Gesamtmodells jedoch ausführlich modelliert. Alle Prozessschritte werden im Spannungsbereich zwischen Wissenschaftlichkeit und der Verfügbarkeit von IT-Ressourcen durchgeführt, was bspw. über die Anwendung von Data Science-Methoden hinaus Effizienz-, Security- und Datenschutzbetrachtungen erfordert, aber auf Seiten der Wissenschaft auch die Neu- und Weiterentwicklung von Methoden und Verfahrensweisen unterstützt (vgl. Schulz et al., 2022, S. 14ff.).

Nach Abwägung verschiedener Kriterien, unter anderem die Verbreitung, praktische Nutzbarkeit im Beruf sowie die gewünschte sichtbare Verbindung von Wissenschaftlichkeit und Praxis wurde entschieden, das DASC-PM als Basis für die Überarbeitung des Moduls zu verwenden. Um den Studierenden jedoch weiterhin einen Einblick in den verbreiteten CRISP-DM zu geben, wird dieser Bestandteil

³ CRISP-DM: Cross Industry Standard Process for Data Mining

⁴ DASC-PM: Data Science Process Model

eines umfassenden Anwendungsszenarios mit Gamification-Elementen, das in einem parallelen Arbeitsprozess entwickelt wird.

2.3 Das 4C/ID-Modell als Instructional Design Konzept

Sucht man nach Instructional Design-Modellen, findet man viele unterschiedliche Konzepte in variierender Komplexität und Vielschichtigkeit. Konzepte wie das AD-DIE-Modell (vgl. Review von Spatioti, 2022) oder TPACK (Mishra & Koehler, 2006) unterstützen die Lehrenden in der Vorbereitung der Materialentwicklung, helfen jedoch wenig bei der Entwicklung von adaptierbaren Materialien oder der Ermöglichung von selbstgesteuertem Scaffolding (vgl. Bliss et al., 1996) durch die Studierenden.

Ein Modell, das diese Unterstützung anbietet, ist das Vier-Komponenten Instructional Design- (4C/ID) Modell (Van Merriënboer, 1997). Dieses wissenschaftlich gut untersuchte Modell ist variabel einsetzbar und unterstützt die Entwicklung unterschiedlicher Lehr-Lern-Materialien, die die Studierenden beim Erwerb der komplexen Fähigkeiten fördern, die sie benötigen, um die erworbenen Kompetenzen auf neue Situationen, insbesondere auf berufliche Anforderungssituationen, zu übertragen (Frèrejean, 2019). Die im 4C/ID-Modell verwendeten Komplexitätsstufen können auf die Stufen der Lernzieltaxonomie nach Anderson und Krathwohl (Anderson & Krathwohl, 2001) gemappt werden, so dass die Formulierung und Einordnung der zu entwickelnden Kompetenzen in bekannter Form durchgeführt werden kann.

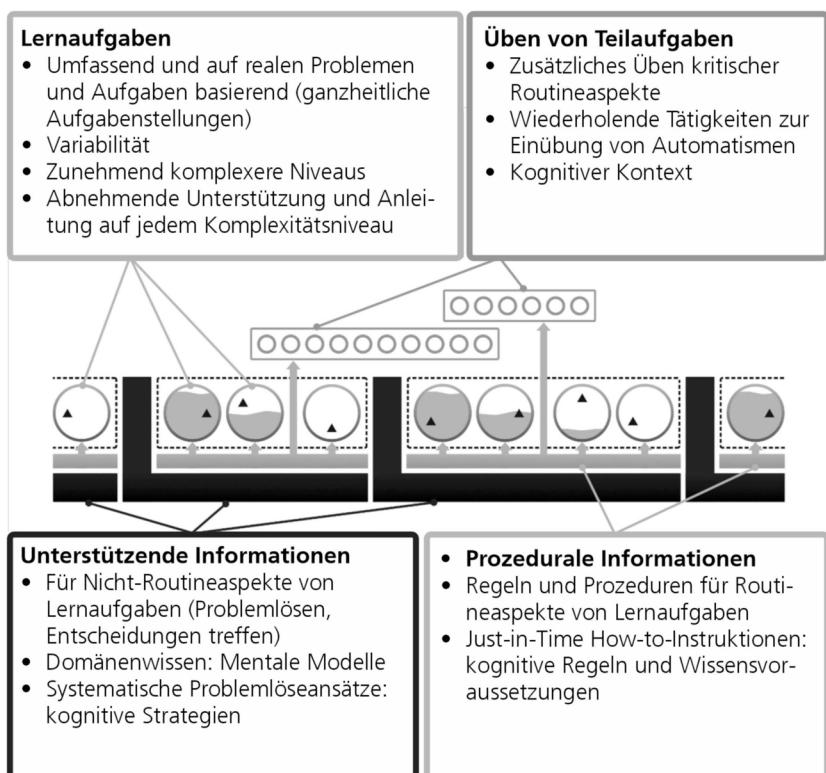
Das Modell selbst besteht aus 4 Komponenten (vgl. Abb. 1):

- Die *Lernaufgaben* (Learning Tasks) stellen den Kern des Lehr-Lern-Materials dar. Sie dienen den Lernenden zum Erarbeiten der Inhalte und zum Erwerb der notwendigen Kompetenzen. Lernaufgaben zu einem Themenbereich werden von Aufgabenset zu Aufgabenset komplexer, innerhalb eines Aufgabensets dagegen wird die Unterstützung und Führung der Lernenden von Aufgabe zu Aufgabe geringer, so dass sie schrittweise Autonomieerfahrungen (vgl. Deci & Ryan, 2012 oder Prenzel, 1993) beim Lernen machen können.
- Um die Lernaufgaben bearbeiten zu können, erhalten die Studierenden *Unterstützende Informationen*, die das notwendige Fachwissen zur Bewältigung der Aufgaben enthalten und dem Aufbau mentaler Modelle und von Heuristiken dienen. Die unterstützenden Informationen bieten die wissenschaftliche Basis der Lehrveranstaltung. Das Material zu einem Lernaufgabenset muss – im Gegensatz zur fachlogischen Gestaltung von Lernmaterialien – jedoch nicht zwangsläufig nur Fachwissen aus einem Bereich, zum Beispiel der »Deskriptiven Datenanalyse« oder »Clustering von Daten« enthalten, sondern kann auch

für diese Lernaufgaben notwendige Informationen zum Datenrecht oder über die Fachdomäne aufweisen.

- *Prozedurale Informationen* zeigen und erklären Verfahren und Vorgehensweisen. Dies könnten beispielsweise Videos sein, in denen deskriptive Verfahren berechnet und durch »Lautes Denken« (Konrad, 2020) erklärt werden. Sie kommen insbesondere bei Lernaufgaben niedriger bis mittlerer Komplexität zum Tragen, da hier besonders häufig wichtige Methoden und Denkweisen verstanden und eingeübt werden müssen (nach Anderson & Krathwohl (2001): Niveau des Anwendens).
- Das letzte Element ist das *Üben von Teilaufgaben*, um insbesondere Routineaufgaben, beispielsweise häufig benötigte Berechnungsverfahren, sicher und effizient durchführen zu können.

Abb. 1: Elemente des 4C/ID-Modells (Van Merriënboer, 1997)



Für ein Modul, das nach dem 4C/ID-Modell gestaltet ist, müssen demnach Materialien aus allen vier Elementen vorhanden sein, um die Lernenden in ihren individuellen Lernprozessen so gut wie möglich zu unterstützen.

3. Verknüpfung des Prozessmodells mit dem 4C/ID-Modell

Das beschriebene 4C/ID-Modell wurde für die Entwicklung gesamter Studiengänge konzipiert. Allerdings können auch einzelne Module eines vorhandenen Studiengangs überarbeitet und dabei vorhandene Lehr-Lern-Materialien wiederverwendet werden. In diesem Fall müssen einige der Designprinzipien in kleinerem Maße umgesetzt werden:

Im Gegensatz zur Neukonzeption eines Studiengangs können die Kompetenzziele nicht frei erarbeitet und auf mehrere Module und Lehrveranstaltungen aufgeteilt werden, sondern zumindest die übergeordneten Kompetenzziele des Moduls sind vorhanden. Frei davon abgeleitet werden können jedoch die Teilkompetenzen, die im Anschluss ausdifferenziert und mit Lernmöglichkeiten gefüllt werden müssen.

Eine zusätzliche Anforderung der in diesem Beitrag beschriebenen Modulüberarbeitung stellt die Verknüpfung mit dem Data Science-Prozessmodell (DASC-PM) dar. Um von Beginn an einen Roten Faden bei der Entwicklung zu haben, wurden die folgenden *Festlegungen* getroffen:

- Der gesamte Data Science-Prozess wird mittels eines durchgehenden Beispiels niedriger Komplexität eingeführt. Dieses Beispiel wird in allen Prozessschritten weiterverfolgt. Durch den Fokus auf niedrige Komplexität können sich die Studierenden zunächst auf die Schritte des Prozess-Modells konzentrieren und können sich zu einem späteren Zeitpunkt mit anspruchsvoller Methoden und Inhalten beschäftigen.
- Lernaufgaben höherer Komplexität greifen unterschiedliche Anwendungsfälle ausgewählter Projekte und Szenarien auf. Nicht in jedem Data Science-Projekt können alle wichtigen Methoden, Verfahren und Erkenntnisse abgebildet werden, so dass das Einführen mehrerer Lehr-Lern-Szenarios unumgänglich ist.
- Bei der Konzeption aller Lernaufgaben und Szenarien wird darauf geachtet, nicht nur die praktische Anwendung zu zeigen bzw. durchführen zu lassen, sondern immer auch Bezüge zur wissenschaftlichen Fundierung herzustellen. Es reicht also nicht aus, eine Analyse durchführen zu können, sondern auch die mathematischen oder statistischen Grundlagen werden in dem Maße erarbeitet, wie es für die Wissenschaftlichkeit der Methode notwendig ist.

- Wiederkehrende Szenarien sorgen durch ihren Wiedererkennungswert bei den Studierenden für einen niedrigeren Mental Load (vgl. Cognitive Load Theory von Chandler & Sweller, 1991).
- Die verwendeten Datensätze sollen auf Open Data (Open Data, 2023) basieren, um sie frei nutzen zu können.

3.1 Vorarbeiten

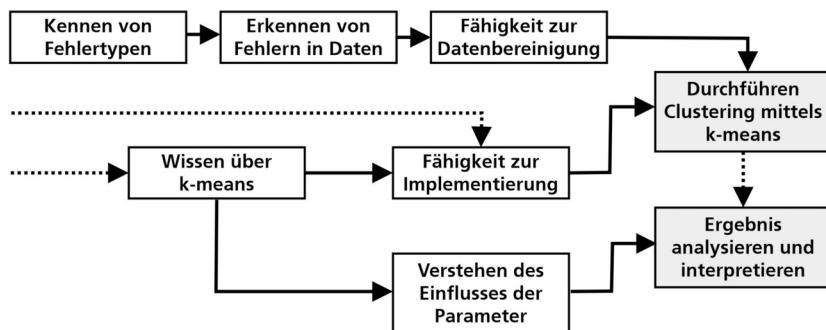
Um Lernaufgaben entlang der fünf Prozessschritte des DASC-PM zu entwickeln, muss zunächst analysiert werden, welche Kompetenzen notwendig zur Erledigung einer Lernaufgabe sind. Hierzu wurden für exemplarische Data Science-Projekte die verschiedenen während der einzelnen Projektpasen entstandenen Artefakte analysiert. Aus diesen Ergebnissen konnten die Kompetenzen abgeleitet werden, die zur Anfertigung der Artefakte notwendig waren. Die so gefundenen Kompetenzen und Wissenselemente wurden strukturiert und in Abhängigkeit zueinander gesetzt. Das entstandene *Kompetenznetz* (siehe Abb. 2) erlaubt es anschließend, Lernaufgaben zu entwickeln, die aufeinander aufbauen und einen schrittweisen Kompetenzaufbau ermöglichen.

Beispielsweise benötigen Lernende zum *Durchführen des Clusterings von Daten mittels des k-means-Algorithmus* (Lernaufgabe) zunächst das *Wissen*, was Clustering bedeutet und welche unterschiedliche Verfahren es gibt. Anschließend benötigen sie das *Wissen*, wie dieser Algorithmus funktioniert und die *Fähigkeit*, dieses Wissen in ausführbaren Code umzusetzen. Hinsichtlich der zu verwendenden Daten müssen Sie die *Fähigkeit* besitzen, die Rohdaten zu bereinigen. Hierzu müssen Sie verschiedene Fehlerarten in Daten *kennen* und in der Folge im Rohdatensatz *erkennen*. Um das Ergebnis des Algorithmus (eigene Lernaufgabe) zu interpretieren, müssen Sie zunächst die Bedeutung und den Einfluss der verschiedenen Parameter des Algorithmus *verstanden* haben, um anschließend das Ergebnis zu *analysieren* und *interpretieren* zu können.

Daraus folgt, dass die Studierenden sowohl Vorkenntnisse über den Algorithmus als auch über die Vorbereitung der Daten benötigen, bevor sie das Verfahren fachgerecht durchführen können. Abb. 2 zeigt einen Ausschnitt aus dem zugehörigen, vereinfachten Kompetenznetz. Derartige Abhängigkeitsnetze wurden für die verschiedenen Kompetenzen des gesamten Moduls erarbeitet.

Im Anschluss daran wurden die Kompetenzen und Wissenselemente den Prozessschritten des DASC-PM zugeordnet: So gehören die Fähigkeiten und Wissen über das Bereinigen der Daten zum Prozessschritt *Datenbereitstellung*, das Durchführen des Clusterings gehört zum Schritt der *Analyse*, während die Analyse und Interpretation des Ergebnisses je nach Projektauftrag schon zur *Nutzbarmachung* oder noch zur *Analyse* gehören.

Abb. 2: Ausschnitt aus dem Kompetenznetz des Moduls »Einführung in Data Science« (eigene Darstellung)



Eine weitere Zuordnung wurde – wo möglich – zu den vier verschiedenen Elementen des Instruktionsmodells getätigt: Während das Fachwissen über Clusteringverfahren und den k-means-Algorithmus in den Bereich der *Unterstützenden Informationen* gehört, sollten Hilfestellungen zur Implementierungen als *Prozedurale Informationen* angeboten werden. Auch das Wissen über Fehlertypen erarbeiten sich die Studierenden anhand von *Unterstützenden Informationen*, während der Erwerb zu Kompetenzen zur Datenbereinigung durch weitere Übungen zur Durchführung des Verfahrens ergänzt wird.

3.2 Entwicklung der Lernaufgaben

Aus all diesen Inhalten und Informationen wurden in den nächsten Schritten Lernaufgaben entwickelt. Für die wenig komplexen Lernaufgaben wurde ein durchgehendes Datenprojekt gewählt, die komplexeren Lernaufgaben werden aus unterschiedlichen Datenprojekten entnommen. Als durchgehende Anwendung niedriger Komplexität wurden Ein- und Verkaufsdaten eines UK-Retailers⁵ verwendet, die zur Entwicklung einer *optimierten Einkaufsplanung* des Unternehmens verwendet werden sollen. Exemplarisch wird die Entwicklung der Lehr-Lern-Materialien des ersten Prozessschrittes *Projektauftrag* dargestellt. Die Lernaufgaben sollen die Studierenden dazu befähigen, die (Teil-)Kompetenzen zu entwickeln, die sie benötigen, um anhand eines vorgegebenen Projektauftrags eine *Projektskizze* zu erstellen:

- Sie können das Szenario und generelle Anforderungen beschreiben.
- Sie können Use-Cases definieren und priorisieren.
- Sie können einen einfachen Datenmanagementplan (DMP) erstellen.
- Sie können eine einfache Risikobewertung vornehmen.

5 Verfügbar als Open Data unter <https://www.kaggle.com/datasets/carrie1/ecommerce-data>

- Sie können Meilensteine und Aufwände abschätzen.
- Sie können die Ziele des Data Science Projekts beschreiben.

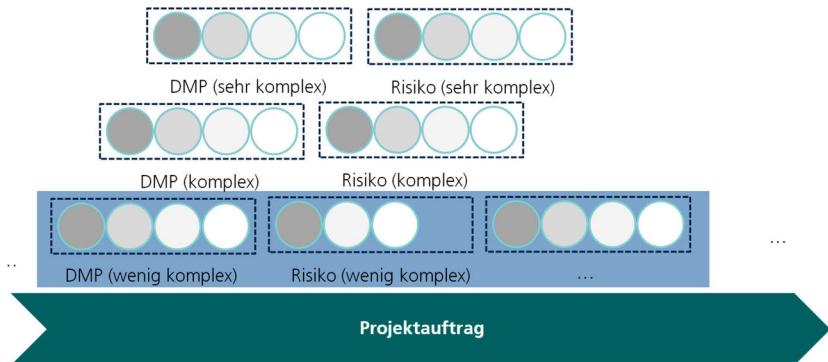
Für jede dieser Teilkompetenzen wurde ein eigenes Set an Lernaufgaben entwickelt. Für das Szenario der *optimierten Einkaufsplanung* sehen die Aufgabenstellungen zur Entwicklung eines Datenmanagementplans (DMP) wie folgt aus:

- *Hohes Maß an Unterstützung und Anleitung:*
»Analysieren Sie den gegebenen DMP nach den folgenden Aspekten: ...«
 - Der DMP ist vollständig gegeben.
 - Die Leitfragen zur Analyse sind gegeben (»Im DMP sind organisatorische Vorkehrungen zur Nachnutzung der Daten beschrieben. Überlegen Sie sich, welche technischen Voraussetzungen aus Ihrer Sicht dafür notwendig sind.«; »Sie finden im DMP unterschiedliche Rollen, die Projektbeteiligte einnehmen können. Analysieren Sie, welche Rolle Zugriff auf welche Daten benötigt.«; »Welche Informationen finden Sie in den Metadaten aller Datendateien?«).
 - Umsetzung in Moodle mit H5P-Elementen (Zuordnungen, MC-Fragen).
- *Mittleres Maß an Unterstützung und Anleitung:*
»Skizzieren Sie, welche Dokumente Sie für die Sicherstellung der Datensicherheit im Bereich ›Lager‹ benötigen ...«
 - Die Dokumente müssen aktiv beschrieben werden.
 - Extraktion der benötigten Dokumente ist aus der Aufgabenstellung möglich, es sollten weitere Recherchen durchgeführt werden, um das Ergebnis zu vervollständigen. Hinweise zur Suche sind gegeben.
 - Umsetzung als Moodle-Aufgabe mit Peer-Review.
- *Geringes Maß an Unterstützung und Anleitung:*
»Formulieren Sie die Anforderungen an die Datensicherheit im Bereich ›Einkauf Abt. 1‹.
 - Extraktion des Ergebnisses ist nicht mehr aus den gegebenen Informationen möglich.
 - Es müssen weitere Quellen genutzt werden, die selbst gefunden und ausgewertet werden müssen; eine Weiterverwendung der Quellen der vorherigen Lernaufgabe ist teilweise möglich.

Die Studierenden haben bei der Bearbeitung der Fragestellungen von Schritt zu Schritt mehr Freiheitsgrade, allerdings wird auch die Bearbeitung anspruchsvoller und das Maß an Hilfestellung geringer, so dass die Studierenden schrittweise dazu befähigt werden, die erlernten Methoden auf die jeweilige Fragestellung zu übertragen.

Die Lernaufgaben höherer Komplexität, die aus anderen Data Science-Projekten stammen, sind ebenso strukturiert: zunächst Aufgaben mit wenig Freiheitsgraden, aber einem hohen Maß an Unterstützung, anschließend schrittweise weniger Führung und Unterstützung, dafür mehr Freiheit in der Bearbeitung.

Abb. 3: Konzeption der Lernaufgaben des Prozessschrittes »Projektauftrag«; Kreise: Lernaufgaben; je dunkler die Färbung, desto höher ist die Unterstützung und Führung. Gestrichelter Kasten: Lernaufgabenset gleicher Komplexität und gleichen Inhaltsbereichs; Blau hinterlegt: durchgehendes Data Science Projekt mit Lernaufgaben niedriger Komplexität (eigene Darstellung)



3.3 Entwicklung der weiteren Materialien

Folgend dem Ansatz des 4C/ID-Modells wurden die weiteren Materialien zur Ergänzung der Lernaufgaben entwickelt.

Unterstützende Informationen beantworten die Fragen »Was ist das? Wie ist es aufgebaut? Wie funktioniert es?« und stellen den wissenschaftlichen Anteil des Lehr-Lern-Materials dar. Hierfür wurden viele der schon vorhandenen Lehrvideos und Texte weiterverwendet, angepasst und ergänzt.

Prozedurale Informationen unterstützen die Studierenden bei Routineaufgaben, beim Durchdringen von Verfahren und beim Methodeneinsatz. Hierfür wurden und werden Erklärvideos und Tutorials produziert, die diese Unterstützung bieten.

Zur *vertiefenden Übung* von Teilaufgaben werden die schon existierenden und erprobten Übungsaufgaben verwendet, zudem werden für die verschiedenen Szenarios und Komplexitätsniveaus weitere Aufgaben laufend ergänzt und im aktuellen Modul evaluiert.

4. Intendierte Lernprozessverbesserungen

Es wird erwartet, dass die neue Strukturierung der Lehr-Lern-Materialien dazu beiträgt, dass die Studierenden ihr Lernen stärker an ihre Vorkenntnisse und Interessen anpassen können, was der Heterogenität der Lernenden entgegenkommen sollte. Da die Wege durch die Lernaufgaben nicht fest vorgegeben sind, haben die Studierenden ein hohes Maß an Autonomie bei der Bearbeitung des Materials:

Möchten sie sich zunächst einen Überblick über den gesamten Data Science-Prozess verschaffen, können sie zunächst das gesamte Projekt »optimierte Einkaufsplanung« bearbeiten, um sich anschließend mit den verschiedenen Phasen intensiver zu beschäftigen. Dies entspricht einem *spiralcurricularen Aufbau* (Bruner, 1960), der mehrere Begegnungen mit den jeweiligen Lerngegenständen in steigender Komplexität und Anforderungsniveau ermöglicht.

Die Studierenden können sich aber auch in die einzelnen Bereiche direkt vertiefen, indem sie zunächst die Lernaufgaben aller Komplexitätsstufen beispielsweise zum Datenmanagementplan bearbeiten, und anschließend an den Lernaufgaben zum Thema Risikoabschätzung weiterzuarbeiten. Auf diese Weise begegnen sie dem Stoff nur einmal, dies jedoch sehr intensiv.

Je nach Vorkenntnissen und Interessen können sich die Studierenden jedoch auch *eigene Lernpfade* durch das Material erschaffen. Sie könnten beispielsweise entscheiden, sich in die Lernaufgaben zum Projektauftrag zu vertiefen, um anschließend durch die Phasen der Datenbereitstellung und Analyse entlang von Szenarien oder auf Basis ihrer Vorkenntnisse zu gehen, um die letzten beiden Phasen wieder Komplexitätsniveau für Komplexitätsniveau zu bearbeiten, je nach ihren Vorkenntnissen und ihrer Motivation.

Da die Sets an Lernaufgaben entlang des DASC-PM angeordnet wurden, können die Studierenden nicht nur abstraktes Wissen über die Prozessschritte in einem Data Science-Projekt erwerben, sondern auch erkennen, wie die einzelnen Schritte, die darin enthaltenen Fragestellungen sowie die verwendeten Methoden zusammenhängen. Bei dieser Art des Aufbaus von Lehr-Lern-Szenarien wird häufig zurecht angemerkt, dass gerade durch das Implementieren von mathematischen oder statistischen Methoden die Gefahr einer reinen »Methodenschulung« bestünde. Um dem zu entgegnen, werden Teile der Lernaufgaben auf die Reflexion der theoretischen Grundlagen und der Anwendung von wissenschaftlichen Methodiken genutzt und diese durch hierzu passendes unterstützendes Material, das den wissenschaftlichen Ansprüchen eines universitären Studiums genügt, fundiert.

5. Ausblick

In diesem Beitrag wurde die Weiterentwicklung eines Moduls des Master-Studiengangs *Data Science* der FernUniversität in Hagen vorgestellt, die den Anspruch hat, die Entwicklung der beruflichen Handlungskompetenz der Studierenden zu verbessern. Da die Transformation der Lehrveranstaltung noch in vollem Gange ist, liegen noch keine empirischen Befunde darüber, wie erfolgreich die Weiterentwicklung war, vor. Jedoch werden seit Beginn des Studiengangs unter Einsatz von Learning Analytics-Werkzeugen sämtliche Aktivitäten der Studierenden im Moodle-Kurs aufgezeichnet. Wir sind gespannt, welche Änderungen im allgemeinen Lernverhalten, dem Kompetenzzuwachs, dem Verfolgen individueller Lernpfade und auch der Zufriedenheit mit dem Lehr-Lern-Material sich ergeben werden.

Das in diesem Beitrag beschriebene Vorgehen zur Weiterentwicklung eines Moduls kann dann als Vorlage für andere Module des Studiengangs, aber auch für Module anderer Studiengänge dienen. Wie es sich im Laufe der Entwicklung immer wieder gezeigt hat und noch immer zeigt, ist die Qualität der Vorarbeiten – das Kompetenznetzwerk der zu erwerbenden Kompetenzen einschließlich der jeweils zu erreichenden Niveaustufen – essentiell für die Entwicklung der Lernaufgaben. Wurde dieses Netzwerk nicht vollständig entwickelt, besteht die Gefahr, durch das Lehr-Lern-Material nicht alle Kompetenzfacetten und Wissenselemente, die erworben werden sollen, abzudecken. Eine Besonderheit des beschriebenen Moduls war, die verschiedenen Lernaufgaben entlang eines beruflichen Prozessmodells auszurichten. Ist eine derartige Ausrichtung nicht relevant, werden im Sinne des 4C/ID-Modells typische Handlungsprozesse des Faches betrachtet. Diese zu finden und zu beschreiben wäre ein zweiter Schritt in der Entwicklung des jeweiligen Moduls, um anschließend die Kompetenzen des Kompetenznetzwerks mit diesen Handlungsprozessen in Einklang zu bringen. Die anschließende Entwicklung der Lernaufgaben und weiteren Materialien kann dann entlang dieser Handlungsprozesse durchgeführt werden, so dass dieses Verfahren in sehr vielen Modulen, in denen theoretisches Wissen mit Handlungskompetenz verknüpft werden soll, erfolgreich zur Anwendung kommen kann.

Literaturverzeichnis

- Anderson, L.W., & Krathwohl, D.R. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. Longman.
- Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., Raghavan, P., Stodden, V., & Szalay, A.S. (2018).

- Realizing the potential of data science. *Commun. ACM*, 61(4), 67–72. <https://doi.org/10.1145/3188721>
- Bliss, J., Askew, M., & Macrae, S. (1996). Effective Teaching and Learning: scaffolding revisited. *Oxford Review of Education*, 22(1), 37–61. <https://doi.org/10.1080/0305498960220103>
- Bruner, J. (1960). *The Process of Education*. Berlin-Verlag.
- Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, 8(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2
- Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. *Handbook of theories of social psychology*, 1(20), 416–436.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37
- Frèrejean, J., van Merriënboer, J.J.G., Kirschner, P.A., Roex, A., Aertgeerts, B., & Marcellis, M. (2019). Designing instruction for complex learning: 4C/ID in higher education. *European Journal of Education*, 54, 513–524. <https://doi.org/10.1111/ejed.12363>
- Hochschulstatistik (2024). *FernUniversität in Hagen*. <https://www.fernuni-hagen.de/uniiintern/organisation/statistik/index.shtml>
- Konrad, K. (2020). Lautes Denken. In G. Mey & K. Mruck (Hg.), *Handbuch Qualitative Forschung in der Psychologie: Band 2: Designs und Verfahren* (S. 373–393). Springer.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Open Data (2023). Bundesministerium des Inneren und für Heimat. <https://www.bmi.bund.de/DE/themen/moderne-verwaltung/open-government/open-data/open-data-node.html>
- Prenzel, M. (1993). Autonomie und Motivation im Lernen Erwachsener. *Zeitschrift für Pädagogik*, 39(2), 239–253.
<http://dx.doi.org/10.25656/01:11174>
- Rauner, F. (2004). *Praktisches Wissen und berufliche Handlungskompetenz*. (ITB-Forschungsberichte, 14). Universität Bremen. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-366049>
- Schulz, M., Neuhaus, U., Kaufmann, J., Badura, D., Kühnel, S., Badewitz, W., Dann, D., Kloker, S., Alekozai, E. M., & Lanquillon, C. (2020). Introducing DASC-PM: A Data Science Process Model. *ACIS 2020*, 45, 1–12.
- Schulz, M., Neuhaus, U., Kaufmann, J., Kühnel, S., Alekozai, E. M., Rohde, H., Hosseini, S. & Theuerkauf, R. & Badura, D. et al. (2022). DASC-PM v1. 1-Ein Vorgehensmodell für Data-Science-Projekte. <http://dx.doi.org/10.25673/85296>
- Spatioti, A. G., Kazanidis, I., & Pange, J. (2022). A comparative study of the ADDIE instructional design model in distance education. *Information*, 13(9), 402.

- Stodden, V. (2020). The data science life cycle: a disciplined approach to advancing data science as a science. *Communications of the ACM*, 63(7), 58–66. <https://doi.org/10.1145/3360646>
- Van Merriënboer, J.J.G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Educational Technology Publications.
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. Practical Applications of Knowledge Discovery and Data Mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29–40.

