

Fähigkeitsprofile im Physikdidaktischen Wissen mithilfe von Machine Learning

Jannis Zeller, Josef Riese

Vortrag auf der GDCP – Jahrestagung 2023

Problemstellung / Vision

Situation:

- Der Stand des **Fachdidaktischen Wissens** von Studierenden soll ermittelt werden.
- Z. B. zu Feedback Zwecken oder als formative Diagnose



Schwierigkeiten:

- Wie valide formative Diagnose vornehmen?
- Wie kann diese über „Note“ oder „Score“ hinausgehen?

Ansatz:

- Nutzung eines validierten Testinstruments zur Einordnung
- Ableiten **inhaltlicher Aussagen** aus vorhandenem Datensatz

Theorie und Ausgangslage

Fachdidaktisches Wissen

- Fachdidaktisches Wissen (FDW) neben Fachwissen (FW) und Pädagogischem Wissen (PW) als Teil des **Professionswissens** von Lehrkräften bereits seit längerem im Fokus der fachdidaktischen Forschung (Shulman, 1986; Baumert & Kunter, 2006; Riese, 2009; Hume et al., 2019; Sorge et al., 2019)
- **Arbeitsdefinition:** FDW = „Dasjenige Wissen welches zur adressatengerechten Aufbereitung des FW notwendig ist“ (Baumert & Kunter, 2006)
- FDW wird hierzulande häufig **dreidimensional operationalisiert** (Tepner et al., 2012; Kröger, 2019; Gramzow, 2015)
- Itementwicklungsmodell im Projekt ProfiLe-P in den Dimensionen ***Inhaltsbereiche, Facetten & Kognitive Anforderungen*** (nach Gramzow, 2015, S. 104)

Fachdidaktisches Wissen

Konforme Ergebnisse unterschiedlicher Projekte (z. B. ProwiN, ProfiLe-P, KiL) **bei Betrachtung quantitativer Scores:**

→ signifikante **Zuwächse** des FDW im Studium und Vorbereitungsdienst sowie Identifikation bedeutsamer **Prädiktoren** (CP, SWS, allg. kog. Fähigkeit) (z. B. Riese & Reinhold, 2012; Kirschner, 2013; Kröger, 2019)

Jenseits der Gesamtscores:

- *Gibt es prototypische Wissensausprägungen bzw. „Fähigkeitsprofile“?*
- *Was zeichnet einen hohen Wissensstand (inhaltlich / kriterienorientiert) aus?*

Entwicklung und kriterienorientierte Beschreibung des FDW

Erste datenbasierte inhaltliche Beschreibungen von Niveaustufen in Kooperationsanalyse der Projekte KiL/KeiLa und ProfiLe-P(+) aufbauend auf Schiering et al. (2023) und Zeller et al. (2022):



FDW ist unabhängig von der konkret zugrundeliegenden Operationalisierung in* ...

- ..niedrigen Ausprägungen auf reproduktive Aspekte beschränkt und...
- ...erweitert sich in höheren Ausprägungen hin zu evaluierenden und kreierenden Elementen.

*siehe hoffentlich bald als Artikel

Entwicklung und kriterienorientierte Beschreibung des FDW

- Bestehende Ergebnisse (noch) nicht ausreichend für formatives inhaltliches Feedback
- Aussagen bleiben auf Scores beschränkt
 - Auch Kategorisierungen von Scores primär theoretisch begründet
- Auch Niveaustufen nur für stark generalisierte & hierarchische Einordnung geeignet
 - Z. B. keine Unterscheidung zwischen Stärken im Evaluieren oder im Kreieren

Ziel der Analyse

Vision:

Generalisierbare, inhaltlich-kriterienorientierte, empirisch begründete Beschreibung von Ausprägungen des FDW, z. B. zur Ermöglichung von inhaltlich aussagekräftigem **Feedback / formativem Assessment über hierarchische Einordnungen hinaus.**

FF1: Welche Fähigkeitsprofile des FDW lassen sich in den Score-Daten aus ProfiLe-P+ mithilfe von Clusteranalysen finden?

FF2: Durch welchen Sprachgebrauch im Testinstrument zeichnen sich die Personengruppen zu den Fähigkeitsprofilen aus?

Methodik und Ergebnisse

- ProfiLe-P+ (Vogelsang et al., 2019):
Daten in Quer- und
Längsschnitten zu
Professionswissen und
Performanz in standardisierten
Handlungs-Situationen
- Genutztes FDW-Testinstrument
nach Gramzow (2015) mit 4
geschlossenen und 20 offenen
Aufgaben
- Insgesamt $N = 846$ digitalisierte
& bepunktete Bearbeitungen
dieses Instruments vorhanden

Aufgabe 1 [9]

Ein Lehrer hat das Wechselwirkungsprinzip „Actio=Reactio“ (3. Newtonsches Axiom) in einer 9. Klasse eingeführt. Nachfolgend spielt sich folgende Szene ab.

Lehrer: Stellt euch jetzt einmal vor, ein Apfel hängt an einem Baum. Wo haben wir hier jetzt Actio und Reactio?

Schüler A: Na, ist doch klar, der Apfel zieht am Ast und der Ast hält den Apfel oben!

Die Klasse signalisiert Zustimmung

Lehrer: Ja richtig – schön, ihr habt es verstanden! Was ist denn dann, wenn der Apfel jetzt herunterfällt? Also während des Fallens, wo ist da Actio und Reactio?

Ein Gemurmel stellt sich ein

Schüler B: Ja gilt das denn dann überhaupt noch? Ich meine, ist doch immer nur ideal, dass das gilt!?

Schüler A: Klar hast du noch Actio und Reactio, nur Actio wird halt immer größer, der Apfel wird ja schließlich schneller beim Fallen!

Schüler B: Ich dachte, die müssen gleich sein? Wo willst du überhaupt Reactio haben, der fällt doch frei und wird nicht mehr gehalten!?!

Schüler A: Hm. Na Actio hast du auf jeden Fall schon mal, er bewegt sich ja. Und er wird ja auch nicht beliebig schnell, die Luftreibung bremst ihn ja. Das ist deine Reactio!

- a) Offensichtlich haben die Schüler die Ausführungen des Lehrers nicht richtig verstanden, die Übertragung auf die Situation mit dem frei fallenden Apfel gelingt nicht. Analysieren Sie die Szene: Inwiefern ist das Vorgehen des Lehrers nicht optimal?

- b) In den Aussagen der Schüler werden einige typische, fachlich nicht korrekte Vorstellungen deutlich. Welche können Sie jeweils bei den Schülern entdecken? Nennen Sie zwei.

1. _____

2. _____

Score-Preprocessing

- Fähigkeitsprofile ~ prototypisches Antwortverhalten ~ Cluster in den Score-Daten (+ Prototypischer Sprachgebrauch?)
- Clusteranalysen als unsupervised Learning – Verfahren bisher eher selten in Naturwissenschaftsdidaktik eingesetzt (Zhai et al., 2020a, 2020b)
- Problem: Clusteranalyse auf Basis der FDW-Score-Daten (> 24 Dimensionen) ist kaum interpretierbar → systematisches Problem computerbasierter (psychometrischer) Analyse (siehe Nelson, 2020)
- Methodologische Vorlage Computational Grounded Theory nach Nelson (2020): Iteratives / zyklisches Verbinden von menschlichem Expertenwissen mit computerbasierten Analysen

Cluster-Workflow (siehe auch Zeller & Riese, 2023)

- Induktiv und Deduktiv (siehe Andersen & Krathwol, 2001) ermittelte **Anforderungsdimensionen**: *Erinnern/Verstehen, Anwenden, Analysieren, Evaluieren, Kreieren, Unterrichtssituation*

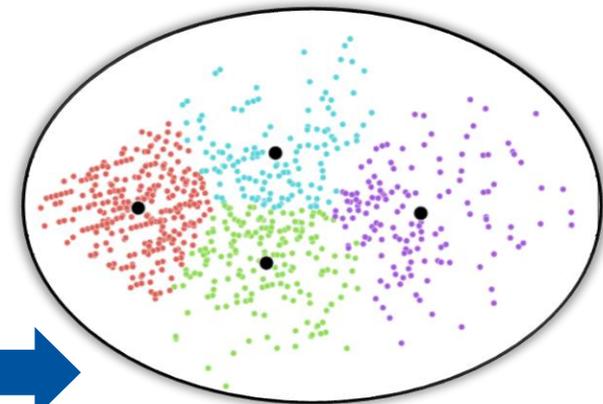
A1 → Analysieren und Evaluieren
 A2 → Anwenden und Unterrichtssituation
 ...



Pers.	A1	A2	A3	A4	A5	A6	A7	A8	...
1	1	1	0	1	0	1	1	1	...
2	0	0	0	1	0	1	0	0	...

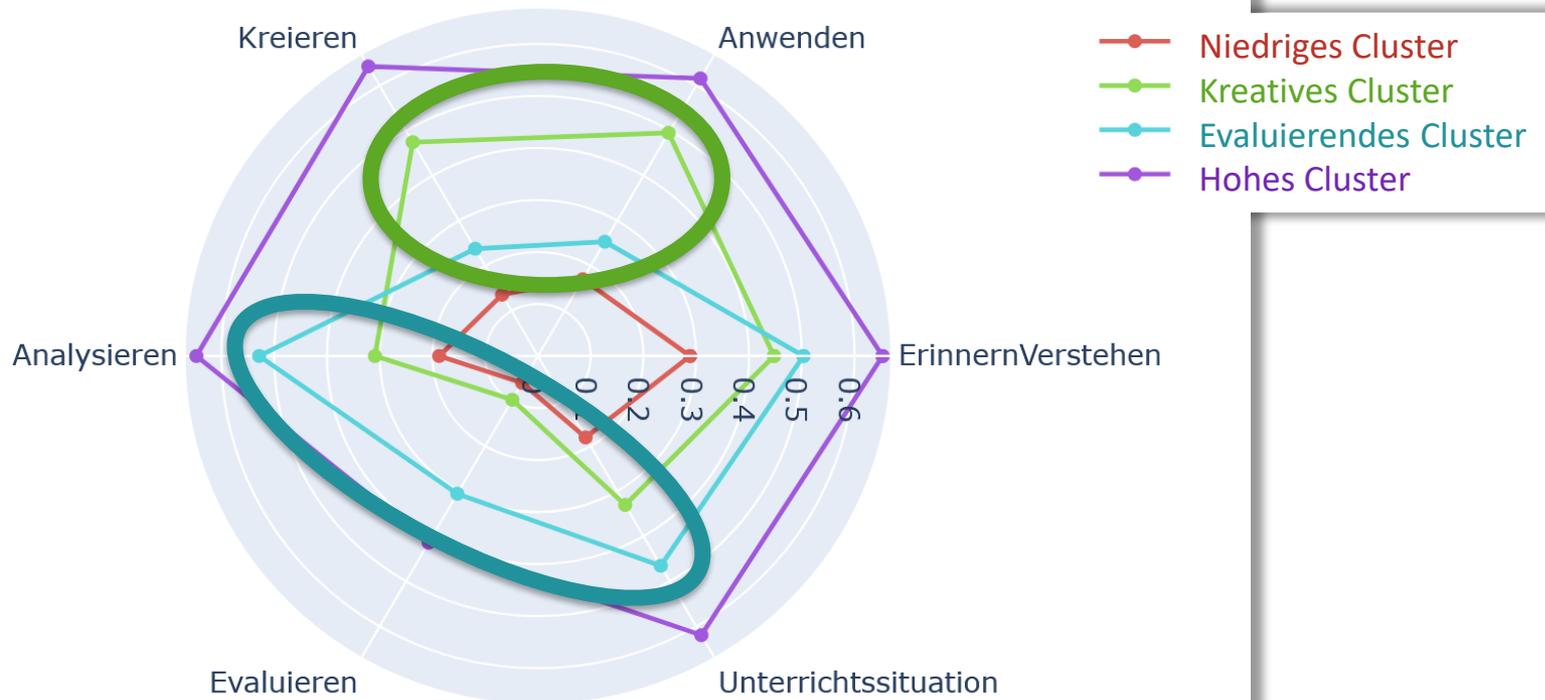


Pers.	Erinnern/Verstehen	Anwenden	Analysieren	...
1	6	2	3	...
2	3	1	0	...
...				



Ergebnisse

Netzdiagramme der k-Means Cluster-Zentren



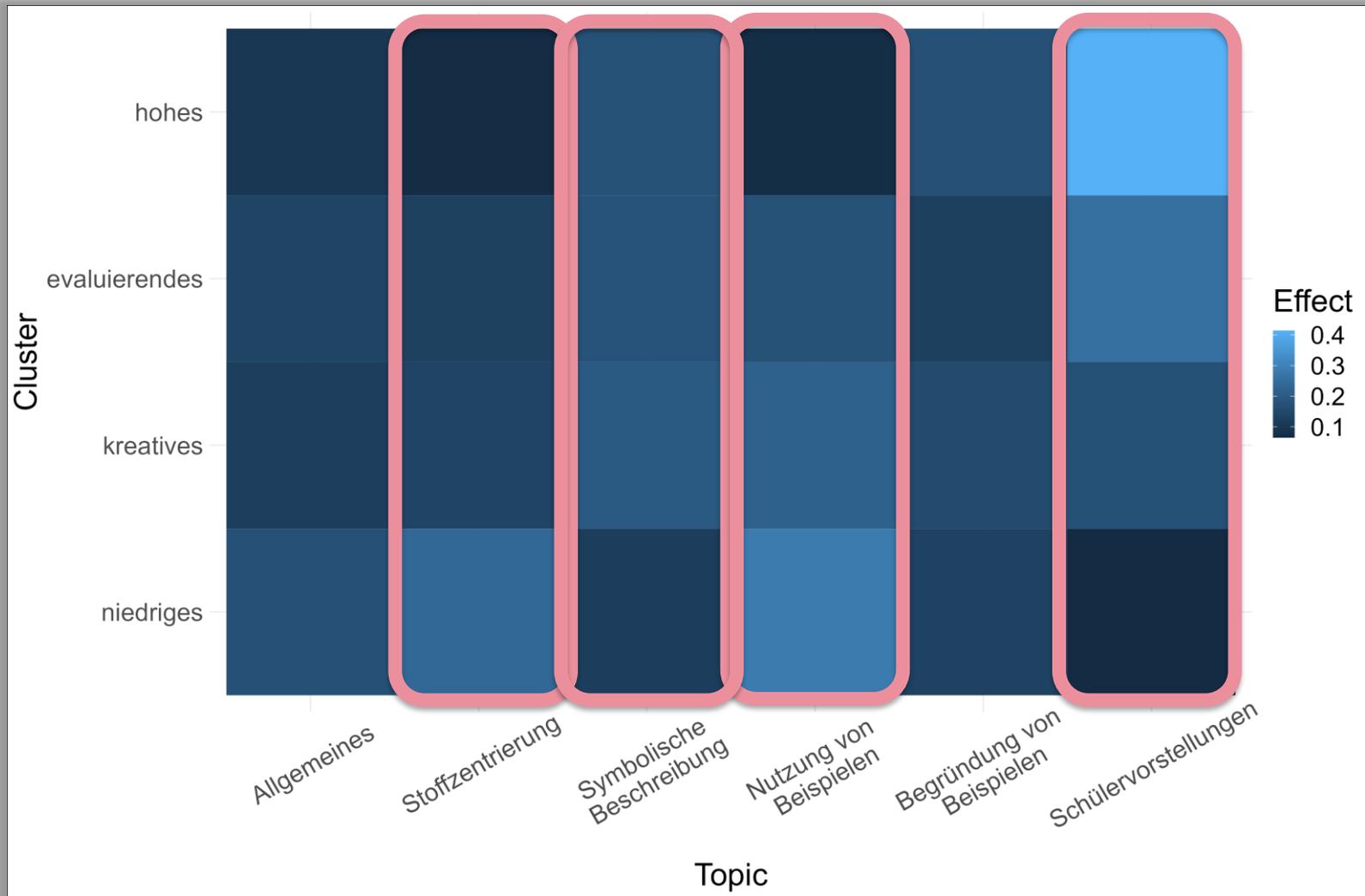
Structural Topic Models

- **Topic Models** (Blei, 2012) erstellen mithilfe eines Mathematischen Modells eine Charakterisierung von „Themen“ (Wort-Wahrscheinlichkeiten) auf Basis des Auftretens von Worten in einem Text Fundus
 - Keine Annahme von „Wortlisten“ o. Ä. im Vorfeld
 - **Structural Topic Models = STM** (Roberts et al., 2016, 2019) ermöglichen es, Informationen über Kovariaten (hier: Score-Cluster) direkt mit in die Modellierung einfließen zu lassen.
- **Score-Cluster** werden mit zusätzlichen Informationen über typischen Sprachgebrauch und verwendetem Vokabular angereichert

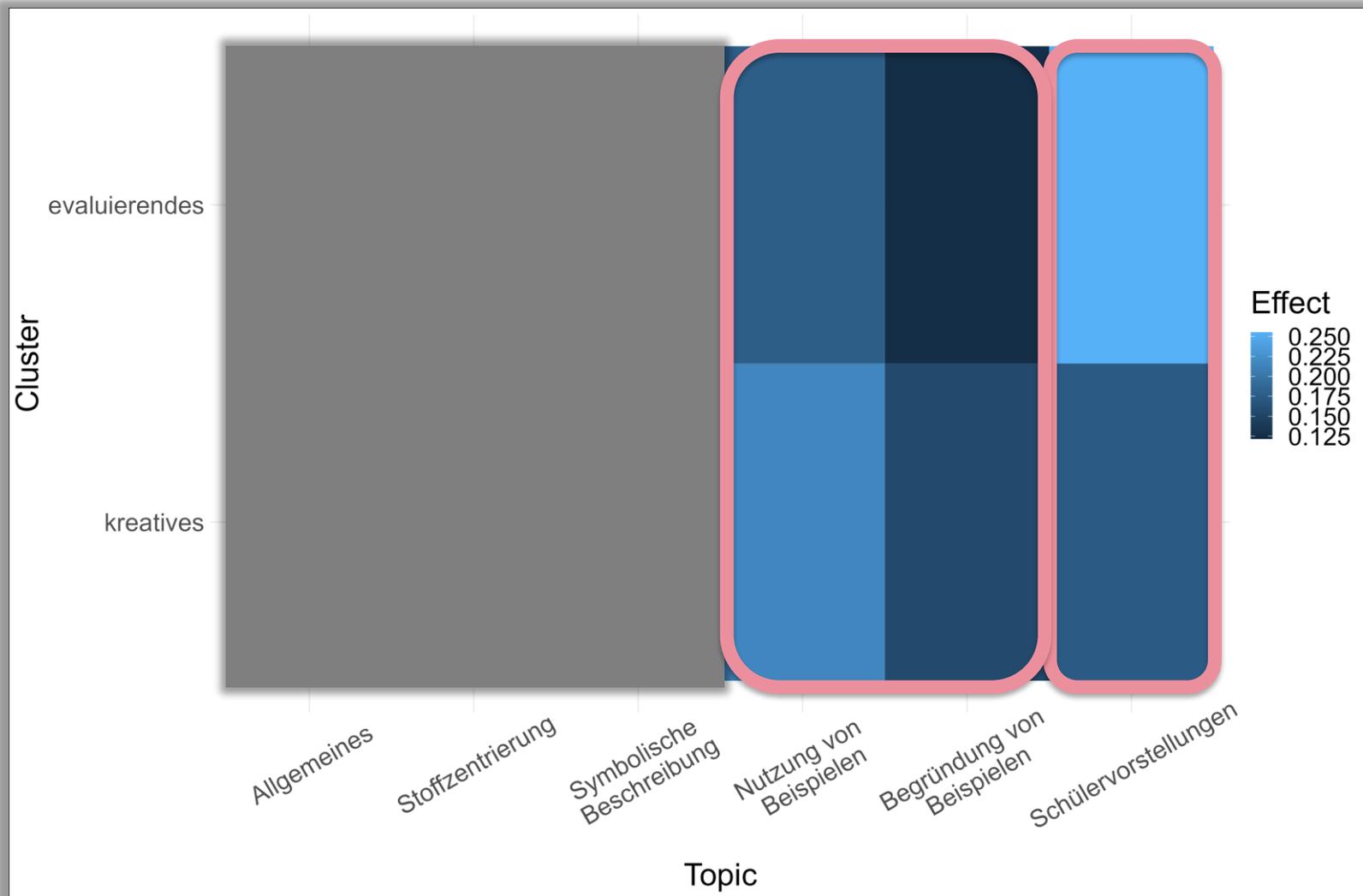
Charakterisierung der Topics

Topic	Charakteristische Worte (Auswahl, unterschiedliche Metriken)	„Kurzbeschreibung“
1	vorstellung, lehrer, zusammenhang, system, ergebnis, vernachlässigung, gelernt	Allgemeine Konzepte
2	auto , erklären, fährt , lehrer, berg , kind, kurv , anschauungsmateri , bus	Nutzung von Beispielen
3	lehrer, vorstellung, impuls , grundkenntnis , unterrichtsstoff , vergisst	„Stoffzentrierung“
4	körper , actio , reactio , v , f , 0 , sim , p , pmv	Symbolische Beschreibung
5	schülervorstellungen , alltagserfahrungen , kognitiven , kraftbegriff , konflikt , alltagsbezüg , begriffwechsel , arbeitsweisen, fachmethoden	Schülervorstellungen
6	veranschaulichung , vorstellung, auto , billiardkugel , denkanstoß , sensibilität , zeigend	Begründung von Beispielen

Zusammenhang zw. Topics & Clustern



Zusammenhang zw. Topics & Clustern



Diskussion

Diskussion

Es lassen sich mithilfe der vorgestellten Methode **Personengruppen** ~ **Fähigkeitsprofile** finden, unterscheiden und verorten, die in der Tendenz **prototypisches Sprachverhalten** zeigen:

1. Personen mit **Stärken im Kreieren** von Unterrichtselementen (besonderer Fokus auf Beispielen)
2. Personen mit **Stärken im Evaluieren** von beobachteten Unterrichtssituationen (besonderer Fokus auf Schülervorstellungen)

„Eigenheiten“ / Limitationen:

- Enge Verzahnung von theoriebasierter Dimensionierung / Interpretation und „objektiver“ Auswertung
- Viele Design-Entscheidungen im Workflow (Score-Cluster aber robust)
- Gefundene Fähigkeitsprofile (Score-Cluster + Sprachgebrauch) sind eher *Tendenzausprägungen als latente Gruppen*.
- Teilweise nur wenige Abstufungen pro Dimension

→ Als Explorative Analyse zu betrachten.

→ Ergebnisse haben Hypothesencharakter

Diskussion

Implikationen:

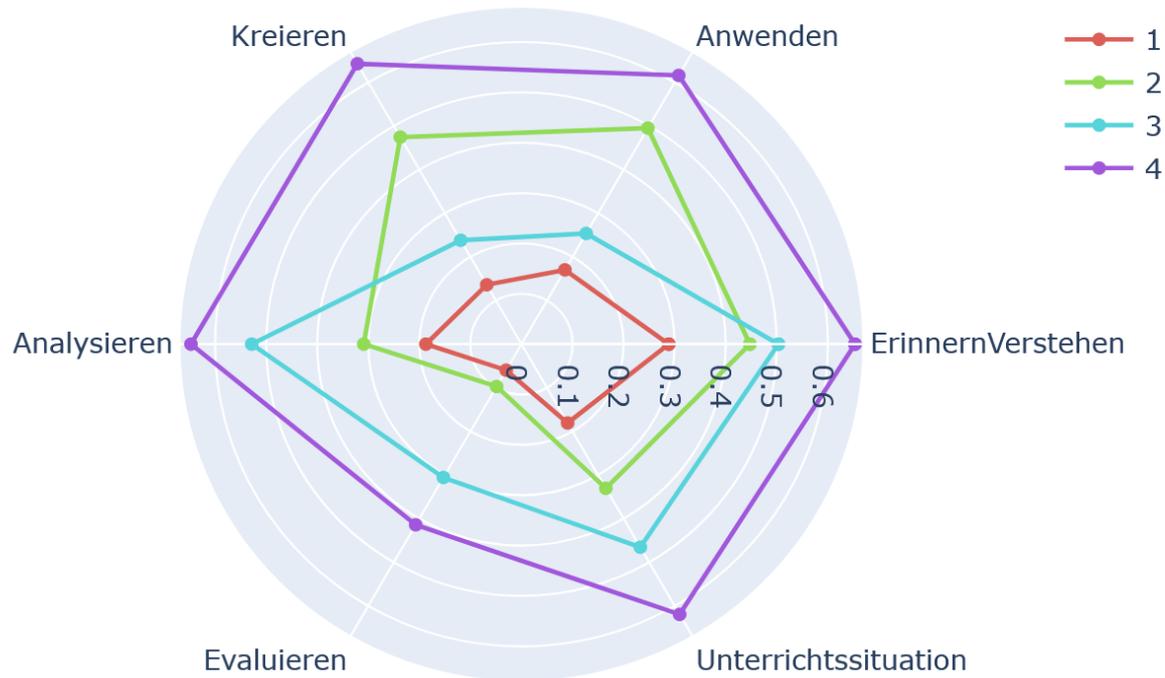
- Schülervorstellungen auch sprach-inhaltlich als Kernkonstrukt festgestellt
- Bei Beispielen scheint „Qualität vor Quantität“ zu gelten
- Verortung von Proband:innen im Rahmen der Fähigkeitsprofile:
 - Möglichkeit für gezielteres Feedback in Bezug auf Stärken bzgl. **Evaluieren/Analysieren** vs. **Kreieren/Anwenden**
 - Möglichkeit für gezieltere Auswahl von Lehrinhalten / Fördermöglichkeiten z. B. **Analyse von Unterrichtsvideografien** vs. **Erstellen von Unterrichtseinheiten**

Ausblick (unterstrichene im Anhang)

- Längsschnitte: Betrachtung der Fähigkeitsprofil-Durchläufe im Längsschnitt
- Zusammenhänge zwischen Fähigkeitsprofilen und demographischen Merkmalen & Unis sowie anderen Professionswissensdimensionen
- ***Automatisierte Auswertung*** mithilfe von NLP & ML
 - Zuordnung zu Fähigkeitsprofilen vs. Feedback in Dimensionen denkbar
 - Transfer in die Lehrpraxis sowie leichteren Zugang zu weiterer Forschung ermöglichen

Vielen Dank für Ihre Aufmerksamkeit!

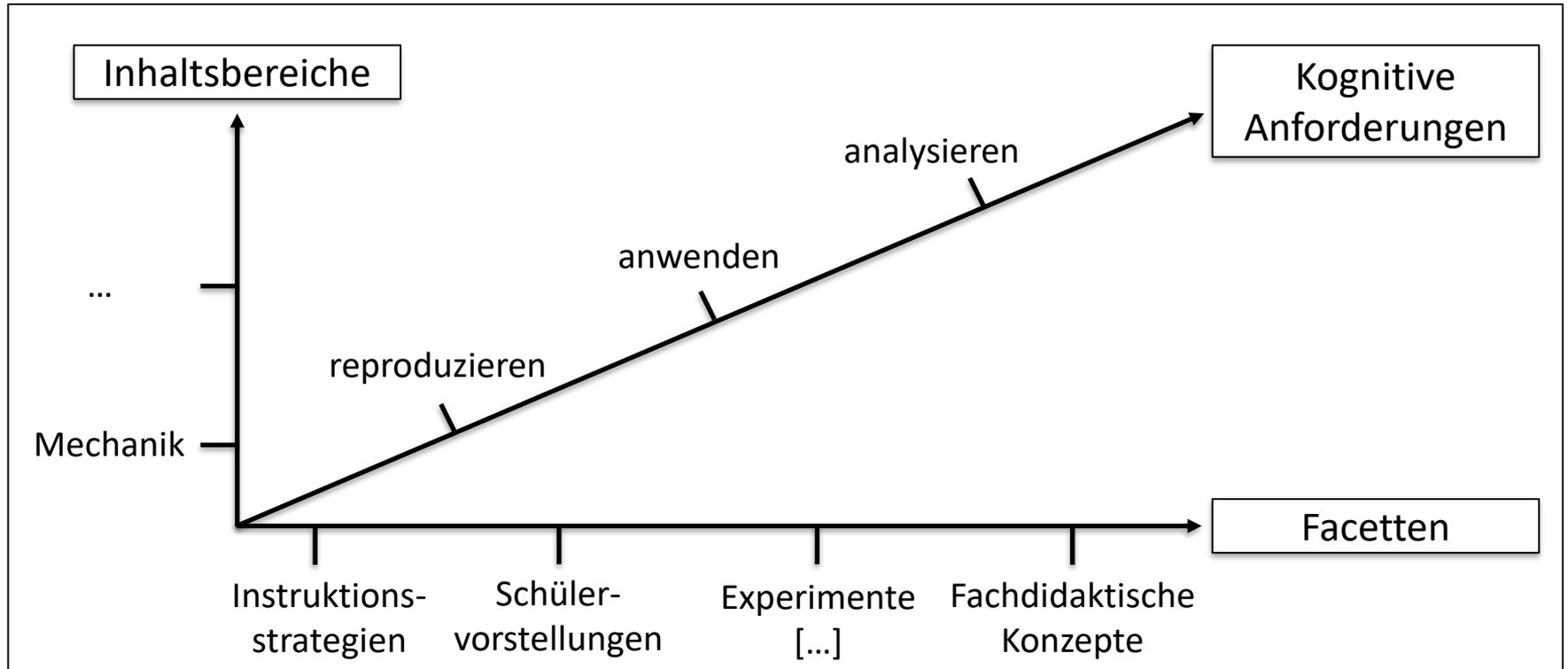
Netzdiagramme der k-Means Cluster-Zentren



Kontakt: jannis.zeller@upb.de

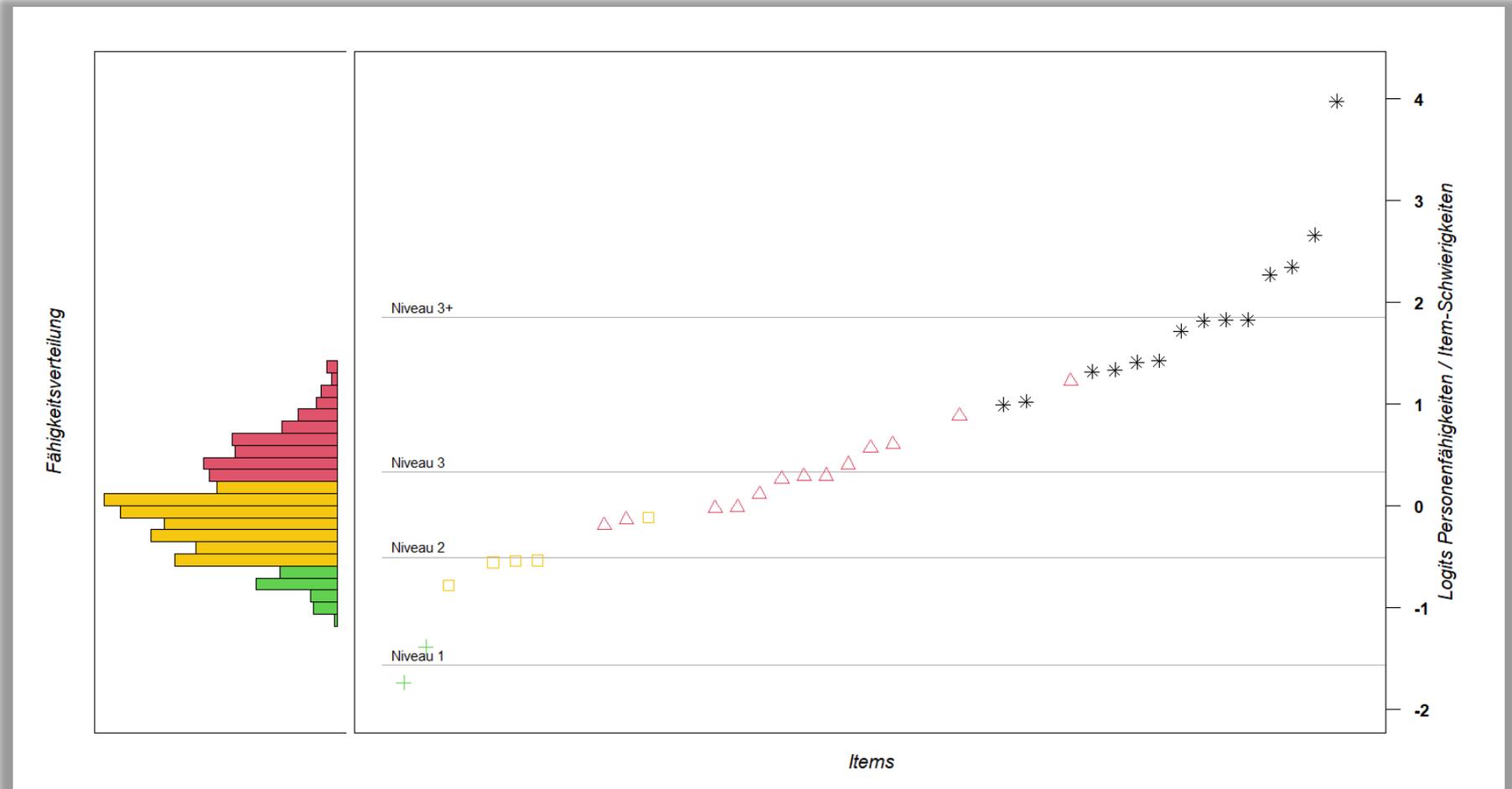
Backup Folien

Modellierung des FDW in Profile-P

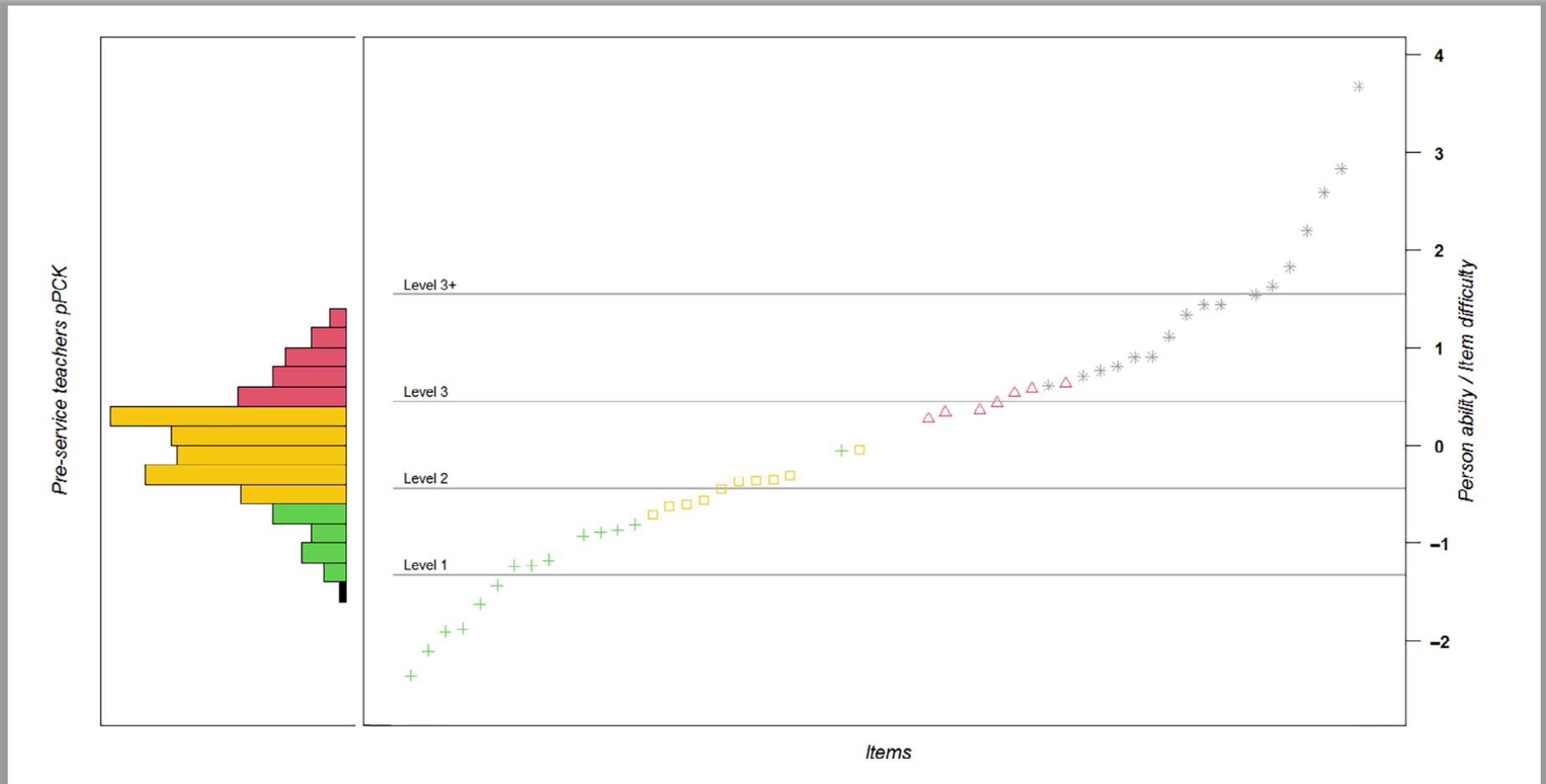


Scale-Anchoring – Wright Map Profile-P

Scale-Anchoring-Verfahren siehe z. B. Mullis et al. (2015)



Scale-Anchoring – Wright Map KiL / KeiLa



nach Schiering et al. (2023, S. 15)

Scale-Anchoring – Niveaubeschreibungen

(grob anhand der Operatoren)

Niveau	ProfiLe-P	KiL / KeiLa (Schiering et al., 2023)
1	Experimente: nennen Schülervorstellungen: nennen	Curriculum: unterscheiden, verstehen, Instruktionsstrategien: kennen SV: unterscheiden, charakterisieren
2	Fachdid. Konzepte: nennen SV: erkennen	Curr: unterscheiden SV: wissen, planen, fördern
3	Exp: entwickeln, nennen Inst: evaluieren SV: rekonstruieren	Assessment: bewerten Curr: arrangieren Inst: kennen
3+	Exp: entwickeln Inst: evaluieren SV: rekonstruieren	Ass: definieren, identifizieren Curr: begründen Inst: erstellen SV: identifizieren, fördern

Score-Cluster: Anforderungsdimensionen (1)

Erinnern:

- Etwas wiederzuerkennen oder abzurufen und dies nennen bzw. wiederzugeben, ist Kernbestandteil der Aufgabe.
- Weite Teile der Aufgabe sollten allein durch Erinnern an fachdidaktische Inhalte lösbar sein.
- Es wird nach „typischen“ Aspekten (z. B. Schülervorstellungen) gefragt, was impliziert, dass es um konsens-Wissen geht, welches explizit in Lehrveranstaltungen erworben werden kann.
- Beispiel: Fakten zu bestimmten fachdidaktischen Konzepten nennen
- Gegenbeispiel: Eine Schüleräußerung wird betrachtet.

Verstehen:

- Ein Element fachdidaktischen Wissens verstanden zu haben, bedeutet, dieses Element beschreiben, klassifizieren, vergleichen und erklären zu können, bzw. es in ein Begriffsnetz einordnen zu können.
- Eine Aufgabe wird der Dimension „Verstehen“ zugeordnet, wenn diese Fähigkeiten / Kompetenzen die Bearbeitung der Aufgabe vereinfachen.
- Weite Teile der Aufgabe sollten allein durch das Verstehen fachdidaktischer Inhalte lösbar sein, insbesondere ohne die Konzepte bereits auf Situationen übertragen zu müssen.
- Beispiel: Die Funktionen von Unterrichtselementen (z. B. Einleitung, Sicherung, Experimentieren) erleichtert deren Auflistung.
- Gegenbeispiel: Eine Situation oder ein konkreter Gegenstand wird betrachtet.

Anwenden:

- Fachdidaktisches Wissen, ein Verfahren oder eine Prozedur anzuwenden oder zu ermitteln, wann die Anwendung einer Prozedur legitim ist, ist Kernbestandteil der Aufgabe.
- Konstruktion / geeignete Auswahl von physikalischen Beispielen zu gegebenen Fragestellungen.
- Beispiel: Prognostizieren von typischen Fehlern mithilfe von Wissen über Schülervorstellungen
- Gegenbeispiel: Analyse eines exemplarischen Unterrichtsmaterials

Score-Cluster: Anforderungsdimensionen (2)

Analysieren:

- Einen Aspekt, eine Situation, eine Äußerung zu analysieren, ist Kernbestandteil der Aufgabe und / oder eine Analyse wird explizit in der Aufgabenstellung eingefordert.
- *Beispiel:* Rekonstruktion von Schülervorstellungen aus Äußerungen
- *Gegenbeispiel:* Auswahl eines geeigneten Beispiels zur Vermittlung eines Fachinhalts

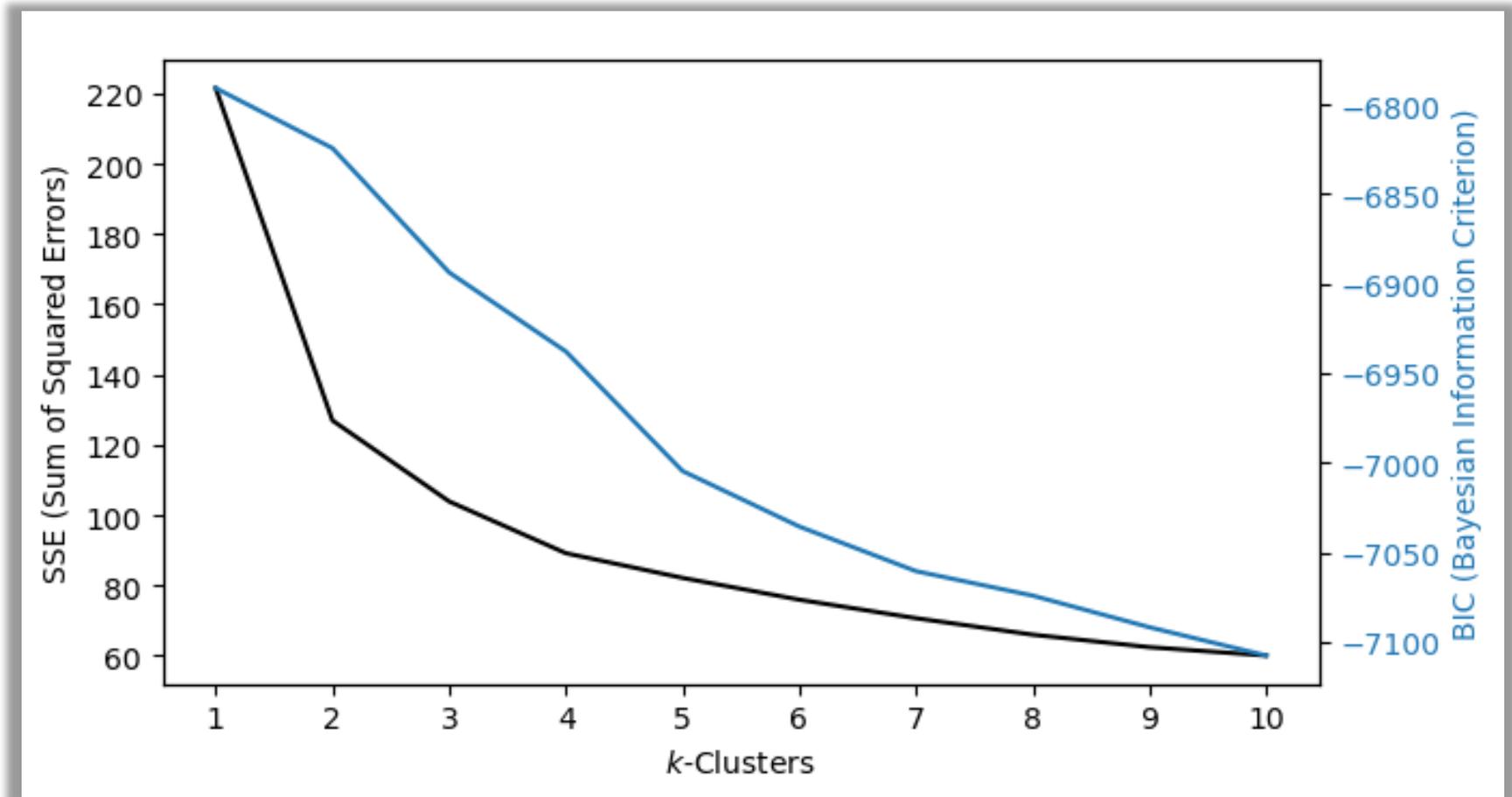
Evaluieren:

- Qualitätsurteile über fachdidaktisch relevante Elemente (z. B. Handlungen, Material, etc.) auf Basis von Kriterien und Standards bzw. des Wissens treffen, d. h. zu überprüfen und kritisieren, ist Kernbestandteil der Aufgabe.
- Auch die Begründung eines (möglicherweise vorgegebenen) Qualitätsurteil fällt unter diese Kategorie.
- Dabei liegt der Fokus auf der Evaluation von fachdidaktisch relevanten Elementen und nicht der Evaluation von Fachwissen beispielsweise in Schüler:innenäußerungen.
- *Beispiel:* Ein beschriebenes Vorgehen einer Lehrkraft bewerten / kommentieren
- *Gegenbeispiel:* Eigenes Vorgehen wird begründet

Kreieren:

- Selbst auf Basis einer Situation oder Beschreibung Elemente fachdidaktisch relevanter Handlungen oder vollständige fachdidaktisch relevante Handlungsketten zu kreieren, ist Kernbestandteil der Aufgabe.
- *Beispiel:* Selbst eine Lösungsstrategie entwickeln oder Alltagsbeispiele unter konkreten Zielsetzungen begründet auswählen

Score-Cluster: K -Means Elbow-Plot



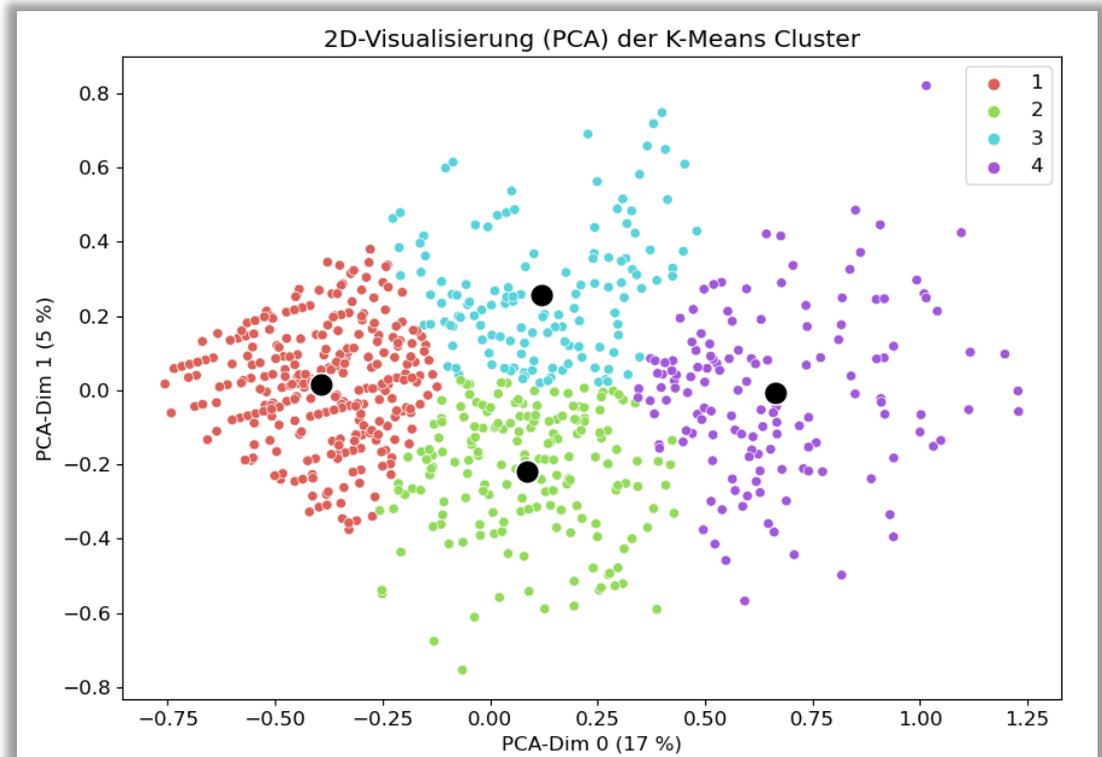
Warum K -Means?

- K -Means (MacQueen, 1967) ist ein „klassischer“ Cluster-Algorithmus
- Modernere Methoden nutzen vor allem „dichte-basierte“ Algorithmen (z. B. DBSCAN und HDBSCAN)
 - Experimente: Datensatz weißt *nicht* systematisch entsprechende „Dichte-Regionen“ auf
- Annahmen Probabilistischer Methoden wie Gaussian Mixture Models (\sim Latente Klassenanalyse) sind nicht erfüllt; z. B. sind die Daten diskretisiert und nicht ausreichend normalverteilt
 - Konvergieren hier nicht / schlecht

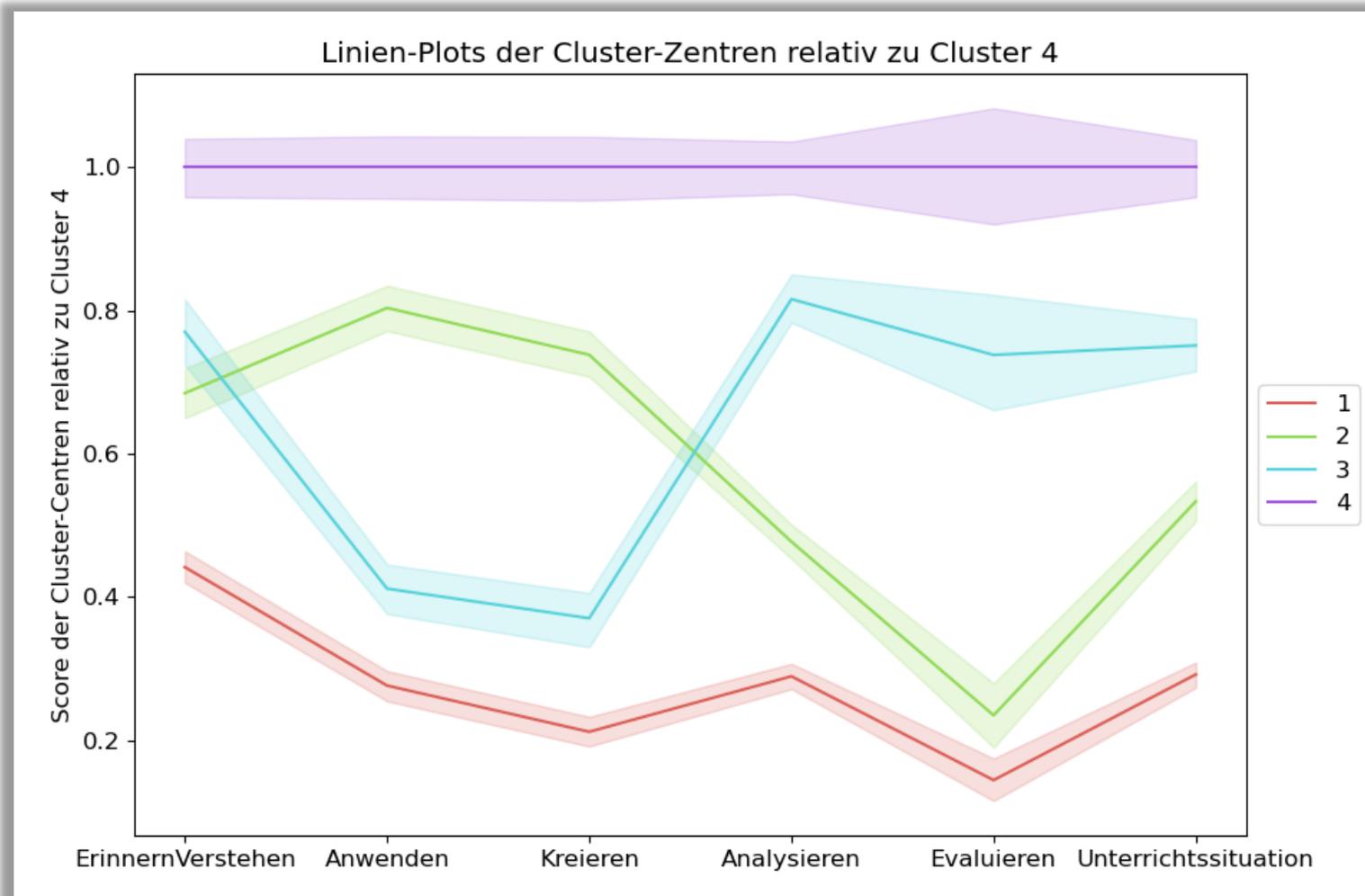
Limitation: Mit K -Means gefundene Cluster sollten nur eingeschränkt als echt „latente“ Gruppen aufgefasst werden, sondern als „Tendenzausprägungen“

Score-Cluster:

- K -Means sorgt für ausgeglichene Cluster-Größen (Ausdehnung im Vektorraum – nicht automatisch Personenanzahl)
- K -Means' „Zentroid“-Vektoren, also Clusterzentren sind als zentrales Ergebnis direkt interpretierbar.



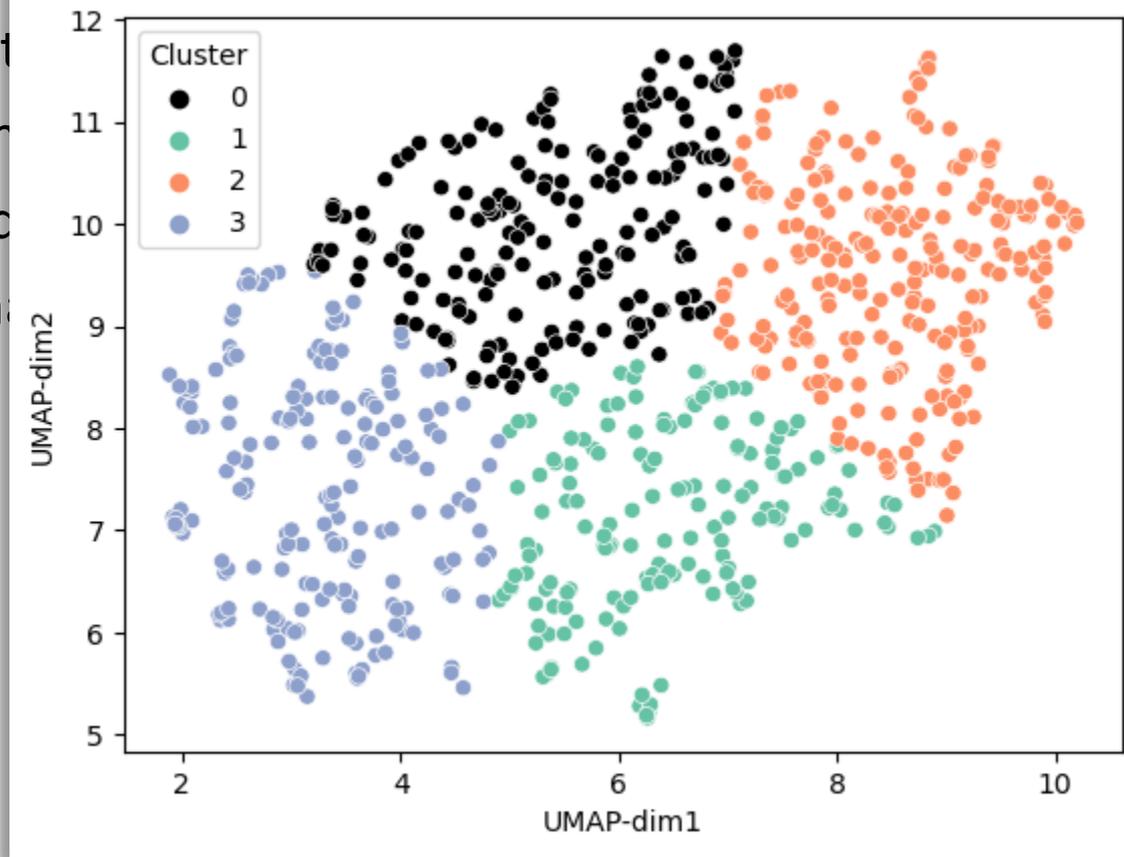
Score-Cluster: Ergebnisse



Validitätsargument für die Dimensionalisierung

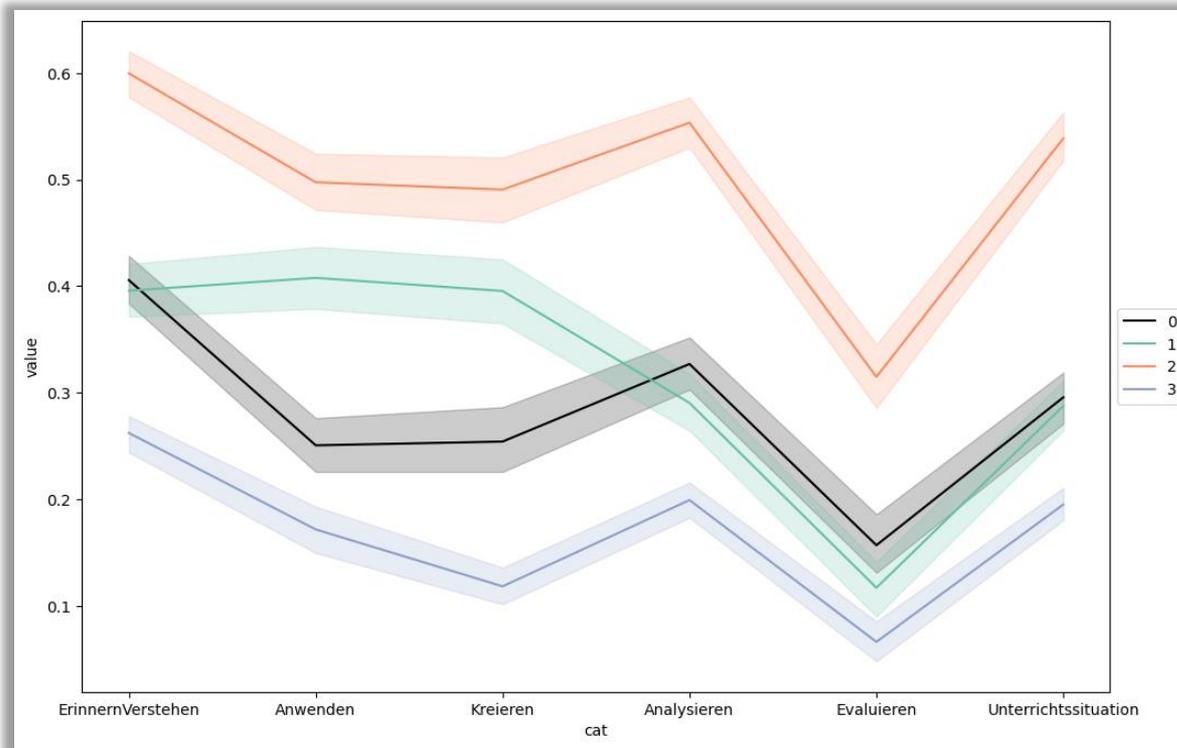
Ansatz: Direkte Analyse des höher-dimensionalen Datensatzes, d. h. direkt auf Basis der It

1. Dimension
2. Clusterbild
3. Transform



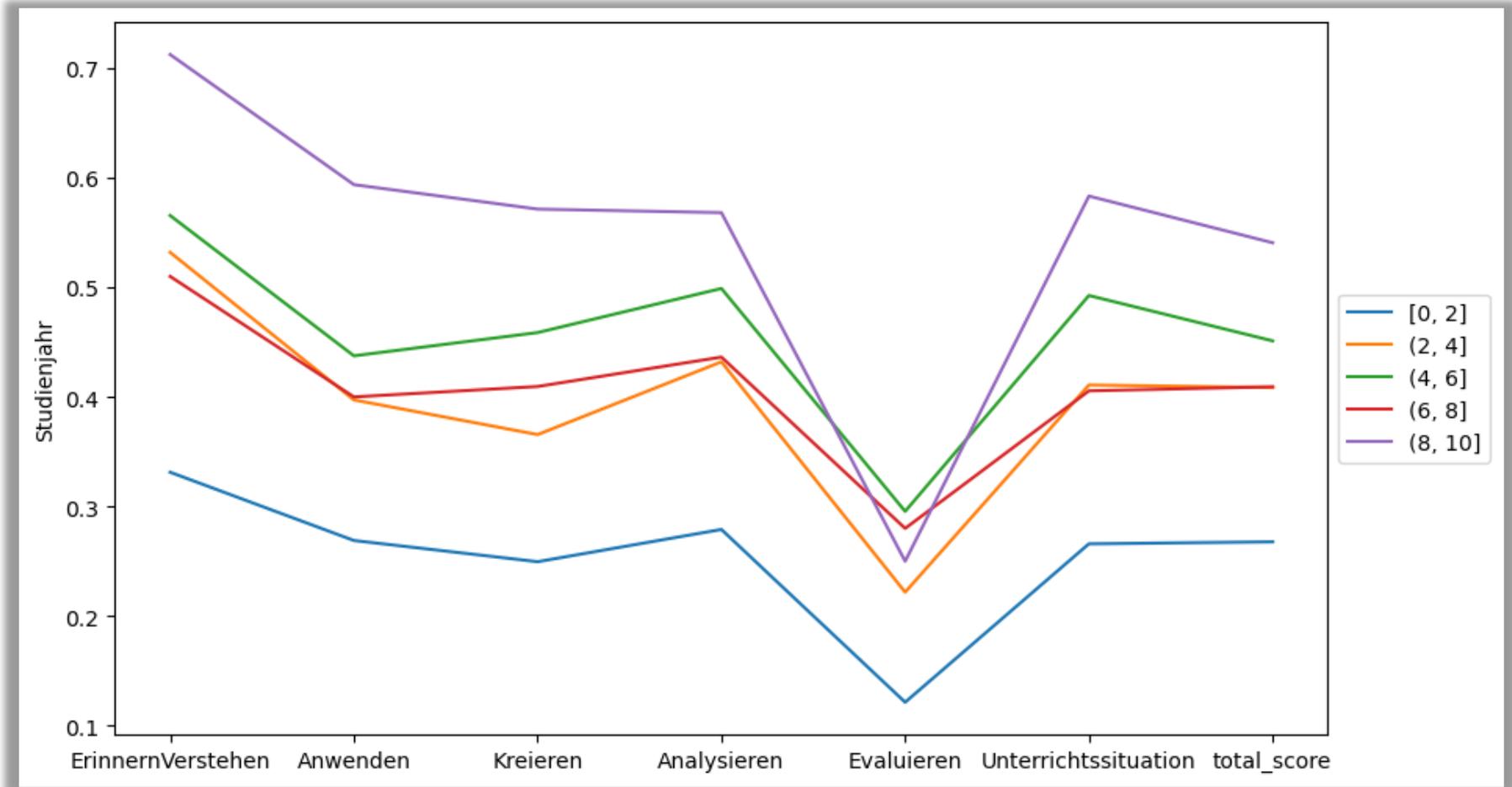
Clusterbildung

Validitätsargument für die Dimensionalisierung

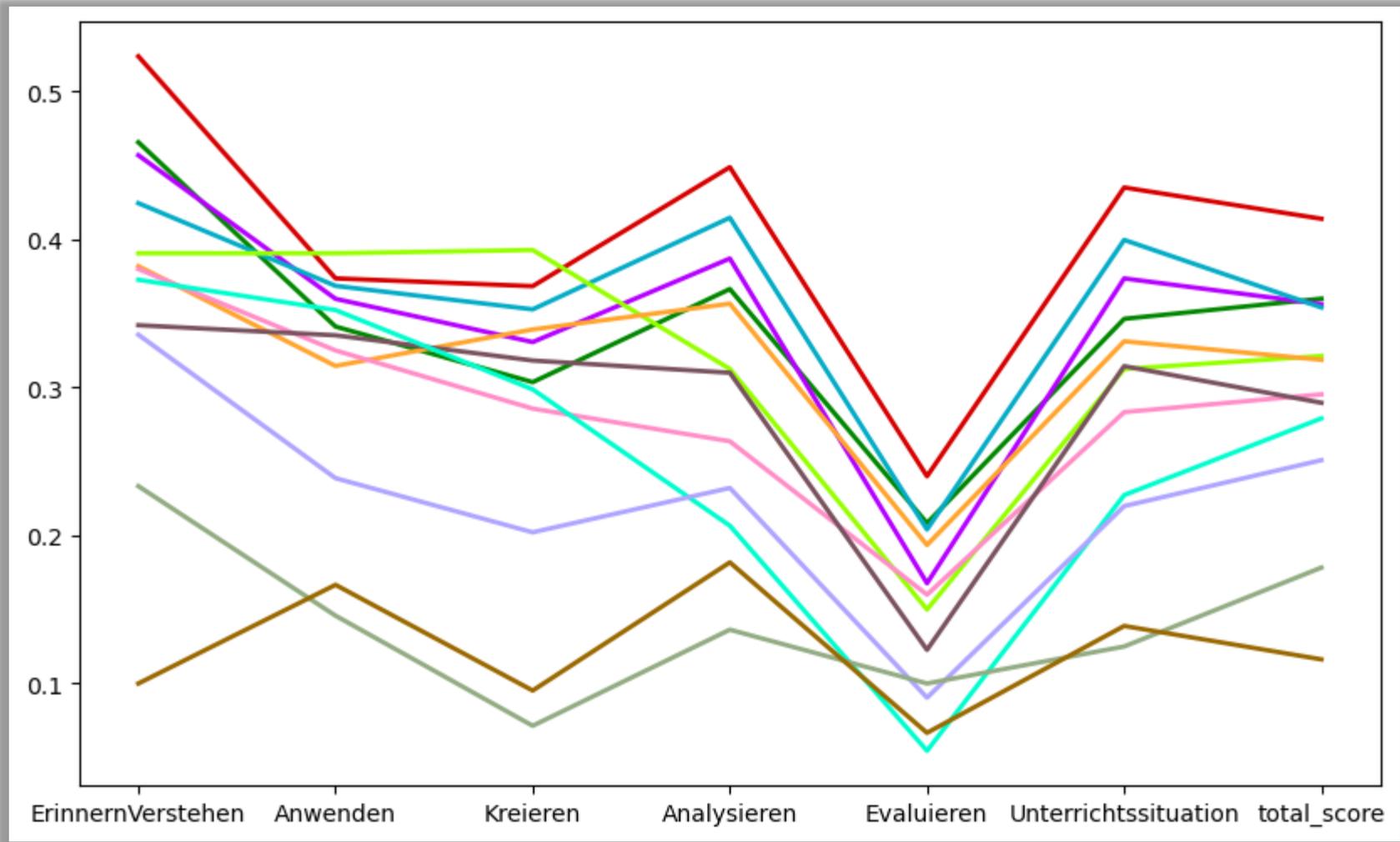


→ Auch wenn die „Kreieren vs. Evaluieren“ Systematik hier nur noch schwach erkennbar ist, so *ist* sie noch erkennbar. Das spricht dafür, dass die theoretisch abgeleitete Dimensionalisierung eine „echte“ Systematik abbildet bzw. hervorhebt.

Dimensionalisierung und Studienjahr



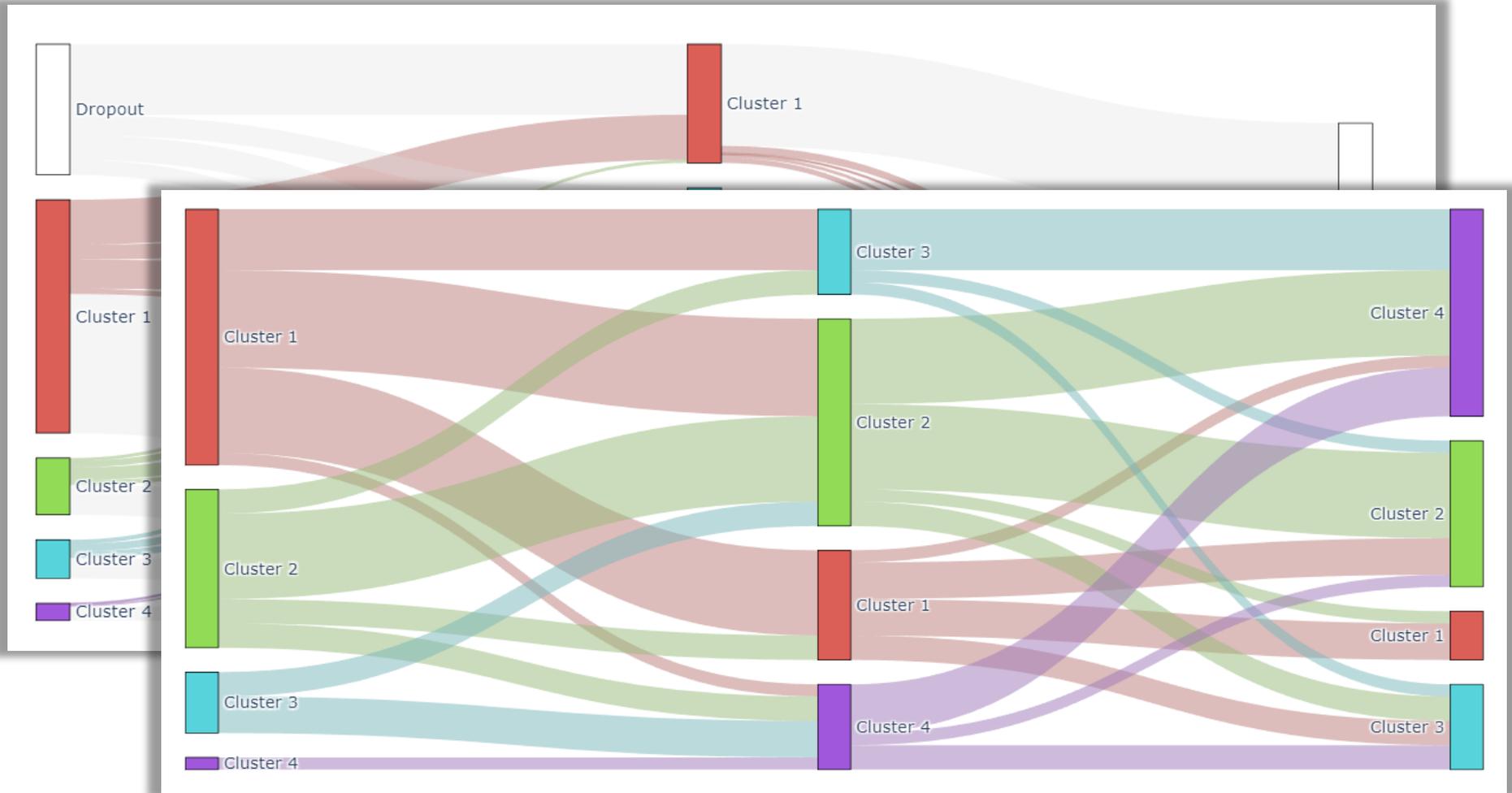
Dimensionalisierung und Universitäten



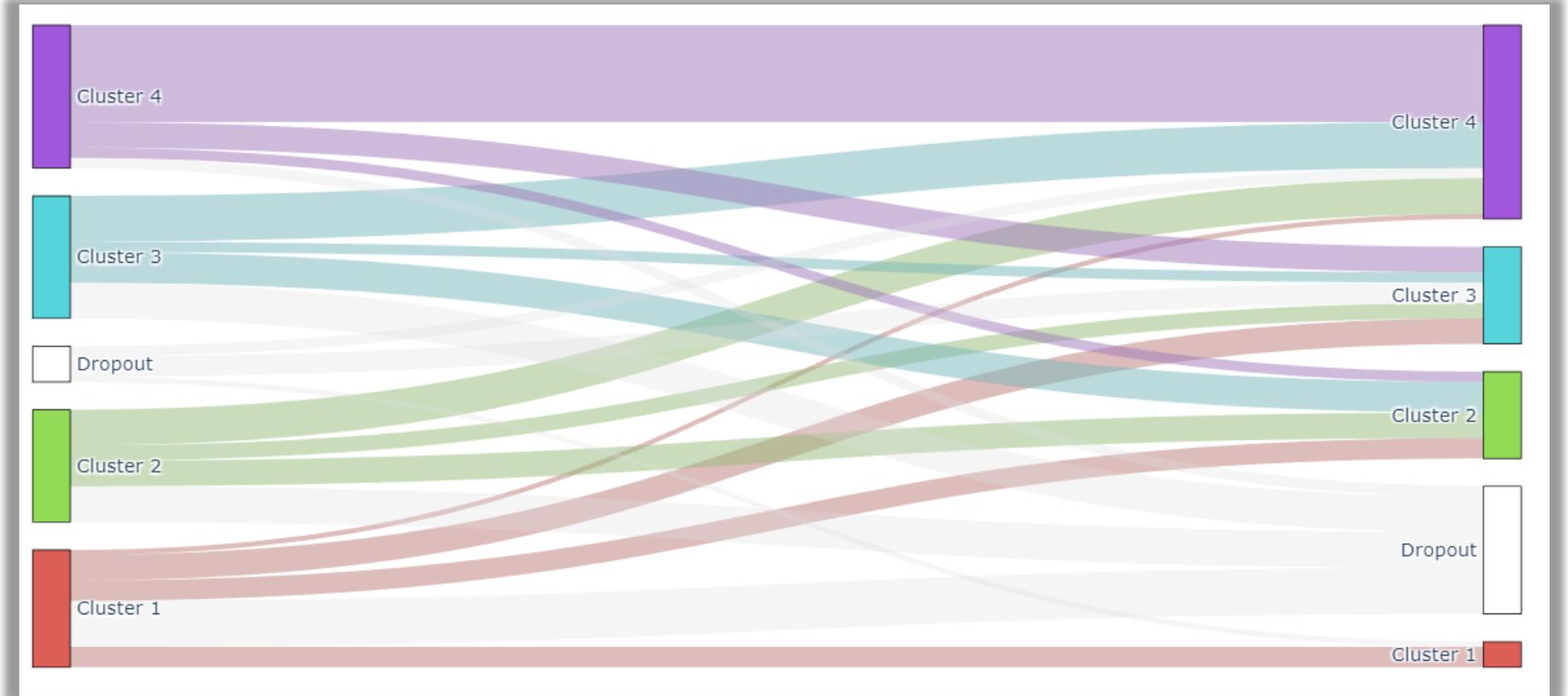
Fähigkeitsprofile Demographie

	„Niedrig“	„Kreativ“	„Evaluierend“	„Hoch“	Gesamt
Größe (N)	380	185	141	140	846
%-weibl.	33	38	31	35	34
Studienjahr	2.7 (2.6)	4.2 (3.2)	5.1 (3.7)	6.9 (4.0)	4.1 (3.6)
Gesamtscore	8.5 (3.4)	15.6 (3.7)	16.8 (3.6)	23.7 (4.0)	14.0 (3.3)

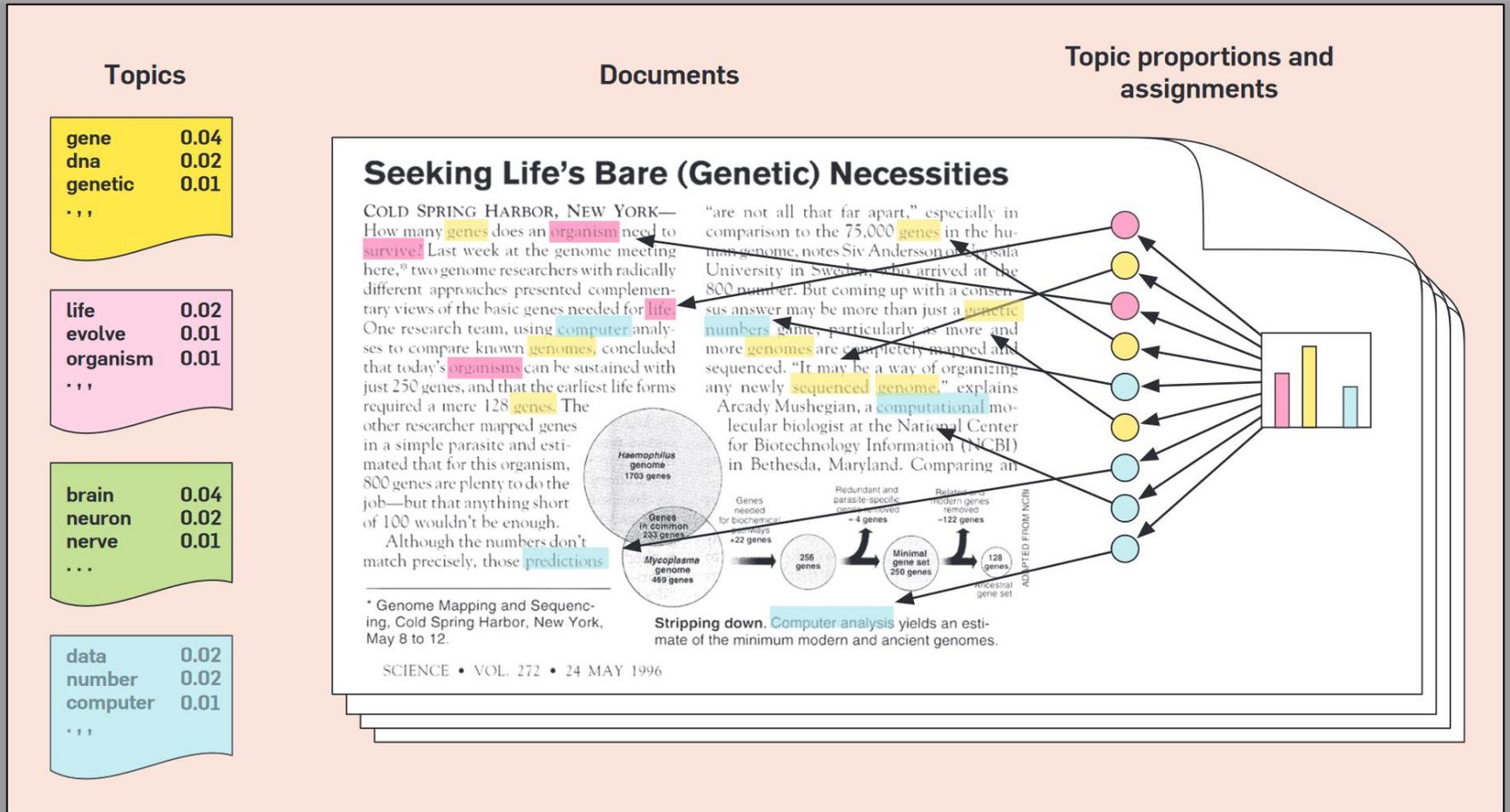
Fähigkeitsprofile Entwicklung Bachelor



Fähigkeitsprofile Entwicklung Master



Topic Models



Topic Models schematisch (Blei, 2012, S. 78)

Structural Topic Models (Roberts et al., 2016)

$$\boldsymbol{\gamma}_k \sim \text{Normal}_P(0, \sigma_k^2 I_P), \quad \text{for } k = 1 \dots K - 1,$$

$$\boldsymbol{\theta}_d \sim \text{LogisticNormal}_{K-1}(\boldsymbol{\Gamma}' \mathbf{x}'_d, \boldsymbol{\Sigma}),$$

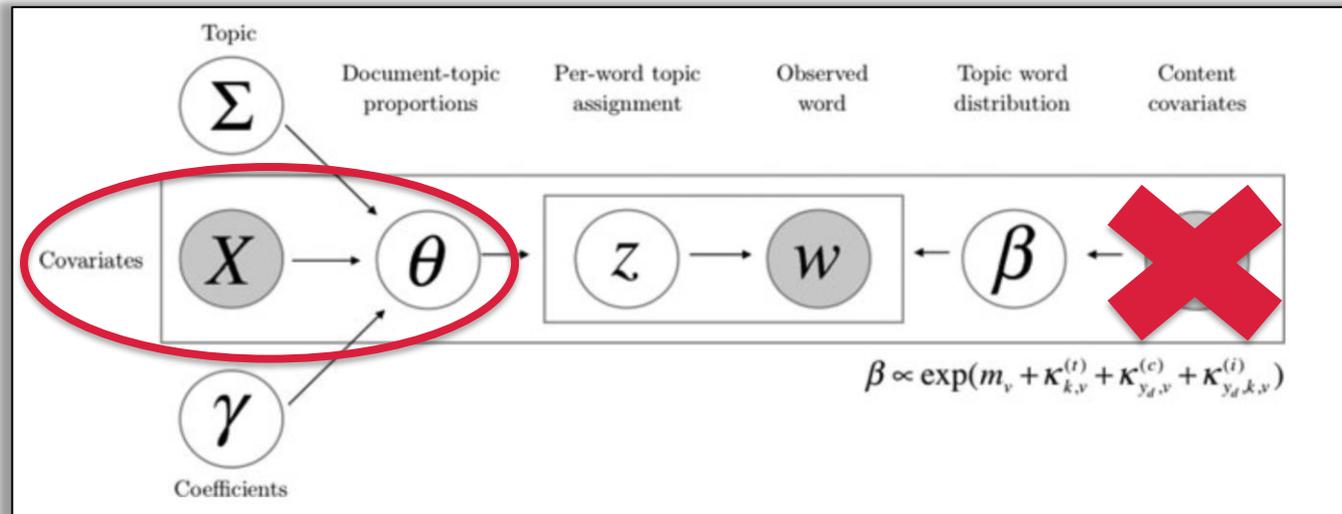
$$\mathbf{z}_{d,n} \sim \text{Multinomial}_K(\boldsymbol{\theta}_d), \quad \text{for } n = 1 \dots N_d,$$

$$\mathbf{w}_{d,n} \sim \text{Multinomial}_V(\mathbf{B} \mathbf{z}_{d,n}), \quad \text{for } n = 1 \dots N_d,$$

$$\beta_{d,k,v} = \frac{\exp\left(m_v + \kappa_{k,v}^{(t)} + \kappa_{y_d,v}^{(c)} + \kappa_{y_d,k,v}^{(i)}\right)}{\sum_v \exp\left(m_v + \kappa_{k,v}^{(t)} + \kappa_{y_d,v}^{(c)} + \kappa_{y_d,k,v}^{(i)}\right)},$$

for $v = 1 \dots V$ and $k = 1 \dots K$,

Math. Bausteine des Generativen Modells (S. 3)



Generatives Modell als Directed Acyclic Graph (S. 990)

STM: Pre-Processing

Input Data

ID	A1	A2	...
001	Foo! und bar...	z. B.

Concat Item-Strings

ID	Text
001	Foo! und bar... z. B.

Remove Abbreviations

ID	Text
001	Foo! und bar... zum Beispiel ...

Lower Casing

ID	Text
001	foo und bar... zum beispiel ...

Stopword removal

ID	Text
001	foo bar... zum beispiel ...

Drop Punctuation

ID	Text
001	Foo und bar... zum Beispiel ...

Stemming

ID	Text
001	foo bar... zum beispi ...

Drop Words due to Freq.

ID	Text
001	foo ... zum beispi ...

STM (mit Score-Cluster)

Zusammenhang zw. Topics & Clustern

- Analyse & Darstellung mittels einer Regression über die Dokumente:

$$P(\text{Topic}_i) \propto \text{Cluster}_k$$

d. h. es wird aufgeklärt, wie die „Fokussierung“ auf Topic i durch die Zugehörigkeit zu Cluster k beeinflusst wird.

- Da die Clusterzugehörigkeit eine nominale Variable ist, ist das letztlich eine **ANOVA**.

A problem has been detected and windows has been shut down to prevent damage to your computer.

The problem seems to be caused by the following file: SPCMDCON.SYS

PAGE_FAULT_IN_NONPAGED_AREA

If this is the first time you've seen this stop error screen, restart your computer. If this screen appears again, follow these steps:

Check to make sure any new hardware or software is properly installed. If this is a new installation, ask your hardware or software manufacturer for any windows updates you might need.

If problems continue, disable or remove any newly installed hardware or software. Disable BIOS memory options such as caching or shadowing. If you need to use Safe Mode to remove or disable components, restart your computer, press F8 to select Advanced Startup Options, and then select Safe Mode.

Technical information:

*** STOP: 0x00000050 (0xFD3094C2,0x00000001,0xFBFE7617,0x00000000)

*** SPCMDCON.SYS - Address FBFE7617 base at FBFE5000, Datestamp 3d6dd67c