

# Future Skills: Data Literacy

---

uni.digital, Berlin

Katharina Schüller

4. April 2019



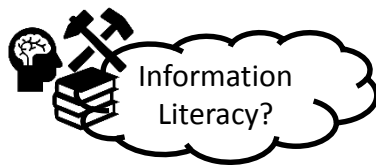
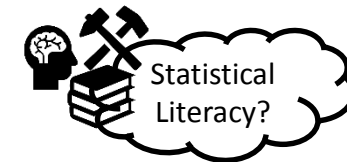
## Data Literacy als 21st century skill

---

Hintergrund und Zielsetzung der Studie

- Kompetenzrahmen für Digitalkompetenzen am Beispiel von Data Literacy entwickeln und umsetzbares Wissen für Hochschulen verfügbar machen
- Messung von Wirkung und Qualität von Lehre und Studium im digitalen Zeitalter wie auch die Testentwicklung für Digitalkompetenzen am Beispiel von Data Literacy vorbereiten

“Wir ertrinken in Informationen und dürsten nach Wissen.” (Naisbitt, 1982)



„Wenn wir mündige Bürger in einer modernen technologischen Gesellschaft möchten, dann müssen wir ihnen drei Dinge beibringen: Lesen, Schreiben und statistisches Denken, das heißt den vernünftigen Umgang mit Risiken und Unsicherheiten.“ (Wells, 1903)

## ... an einen Kompetenzrahmen

- Bildet alle Stufen des Wissens- bzw. Wertschöpfungsprozesses aus Daten ab
- Erfasst alle Kompetenzdimensionen: (a) Wissen, (b) Fertigkeiten, (c) Fähigkeiten, (d) Motivation und (Wert-)Haltung
- Erlaubt es, die erfassten Kompetenzen in konkrete und testbare Lern- oder Kompetenzziele zu überführen
- Reflektiert die Interdisziplinarität der Aufgabe:
  - Was will ich? (Fachexperte)
  - Was kann ich? (Datenexperte)
  - Was darf ich? (Datenschützer)
  - Was soll ich? (Datenethiker)

## ... an Mess- und Testinstrumente

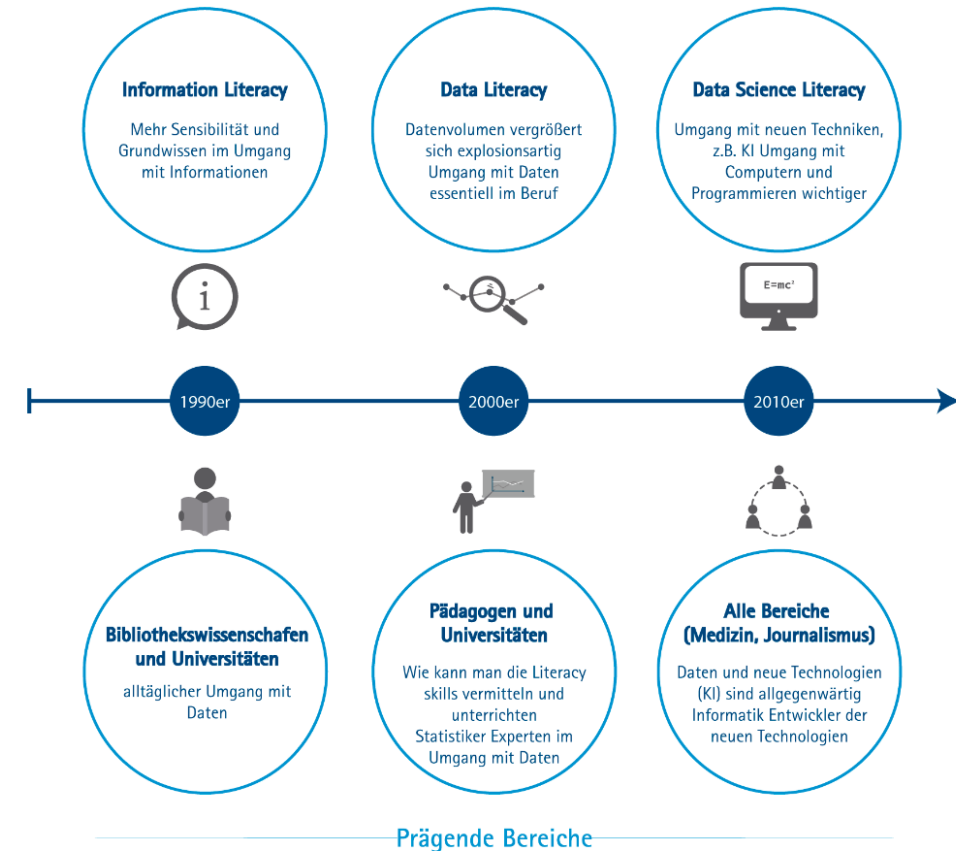
- Erfassen kognitive und affektive Lernbereiche
- Erfassen möglichst viele Lernstufen: (a) Wissen, (b) Lernerfolg, (c) Verhalten, (d) Ergebnis
- Sind transparent bezüglich der Möglichkeiten und Grenzen einer Schlussfolgerung von beobachtbarem Verhalten auf dahinter liegender Kompetenz
- Genügen den Testgütekriterien der Validität, Reliabilität und Objektivität
- Sind mit vertretbarem Aufwand (Geld, Zeit, benötigte Fähigkeiten der Prüfer) durchzuführen

# Der Kompetenzrahmen

---

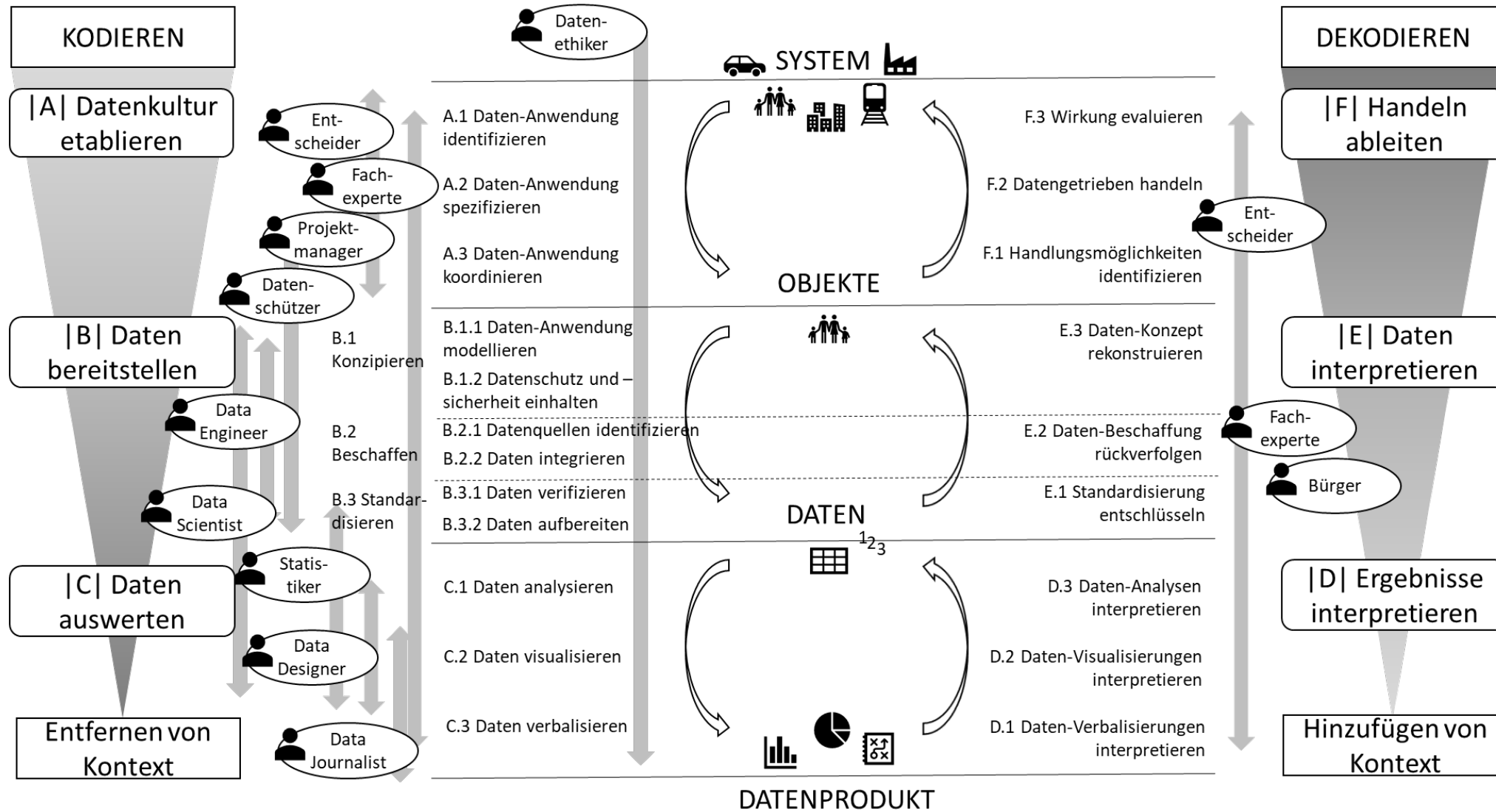
Vom Aufgabenverständnis zu den Kompetenzen

- Data Science Literacy
  - nicht 21st century skill, sondern eher Kompetenzprofil eines neuen Berufes
- Data Literacy
  - sollte weit umfassender verstanden werden als in der Vergangenheit
  - Cluster aller effizienten Verhaltensweisen und Einstellungen für die effektive Erfüllung sämtlicher Prozessschritte zur Wertschöpfung bzw. Entscheidungsfindung aus Daten.
- Haltung („Ethical Literacy“):
  - Fähigkeit, die Bedeutung von Daten zur Entscheidungsfindung vollständig zu erfassen, indem mögliche Interpretationen dieser Daten in unterschiedlicher Kontextualisierung reflektiert und kritisch bewertet werden.



# Vom kompetenzdefinierenden Prozess zum Kompetenzrahmen

Data Literacy als übergreifende Kompetenz und Bezug zu angrenzenden berufsspezifischen Kompetenzprofilen



# Beispiel C.1 Daten analysieren (Kodieren = produktive Seite)

Kompetenz		Kompetenzdimensionen			Kompetenzniveaus
Bezeichnung	Beschreibung	Beispiele für Wissen („Knowledge“)	Beispiele für Fähigkeiten („Skills“)	Beispiele für Haltung („Attitude“)	Beispiele für aufsteigende Niveaus
C.1: Daten analysieren	Setzt Analyseverfahren aus verschiedenen Gebieten (Statistik, Analytics, Machine Learning), unter Zuhilfenahme der geeigneten Werkzeuge sach- und zweckorientiert ein	<ul style="list-style-type: none"> <li>• Kenntnis von Verfahren für unterschiedliche Aufgaben (Deskription, Exploration, Prognose) sowie über deren Voraussetzungen, Stärken und Schwächen</li> <li>• Wissen über Verfahren für gerichtete und ungerichtete Fragestellungen</li> <li>• Wissen über Schätzverfahren und Algorithmen</li> <li>• Wissen über Möglichkeiten der Modelldiagnostik und der Modelloptimierung</li> <li>• Wissen über Maßnahmen, um die Robustheit und allgemeine Gültigkeit der Modelle sicherzustellen (z.B. Kreuzvalidierung)</li> <li>• Wissen über mögliche Ursachen von Artefakten</li> <li>• Wissen über Best Practices und Standards der Modellentwicklung und Datenanalyse, z.B. Validierung</li> <li>• Wissen, dass bei der Analyse von Daten Information verloren geht und dass ein Auswahl- und Bewertungsprozess nötig ist, auf welche Information verzichtet werden kann</li> </ul>	<ul style="list-style-type: none"> <li>• Fähigkeit, messbare Beziehungen in Modellen abzubilden</li> <li>• Fähigkeit, anhand der jeweiligen Fragestellung und der verfügbaren Daten geeignete Analyseverfahren zu identifizieren und auszuwählen</li> <li>• Fähigkeit, das Modell geeignet zu spezifizieren (z.B. durch Festlegung von Optimierungskriterien)</li> <li>• Fähigkeit, das Modell auf Schwächen und Artefakte (z.B. Overfitting, Multikollinearitätsprobleme) zu untersuchen und diesen entgegenzuwirken</li> <li>• Fähigkeit, die Unsicherheit der Modellergebnisse zu beurteilen und die benötigte Genauigkeit zu bestimmen</li> <li>• Fähigkeit, zukünftige Nutzungen der Analyseergebnisse zu antizipieren</li> <li>• Fähigkeit, relevante von irrelevanten Informationen im Analyseprozess zu trennen (z.B. Variablen zu selektieren)</li> </ul>	<ul style="list-style-type: none"> <li>• Bereitschaft, in einem iterativen und oft zeitaufwändigen Prozess Modelle zu implementieren und anzupassen</li> <li>• Skeptische Grundhaltung bei der Datenanalyse</li> <li>• Bereitschaft, Informationsverluste im Prozess der Analyse abzuwägen und zu akzeptieren</li> <li>• Bereitschaft, „good analytics standards“ einzuhalten, selbst wenn diese nicht explizit definiert sind</li> <li>• Bereitschaft, ressourcenschonend zu arbeiten, z.B. das Modell nicht zu „over-engineeren“</li> <li>• Bereitschaft, die benötigte Präzision auch bei knappen Ressourcen und gegen Widerstände durchzusetzen, um Trugschlüsse zu verhindern</li> <li>• „Analytical Fairness“ als Grundhaltung, d.h. Bereitschaft, Analysen nicht durchzuführen, wenn das Risiko eines Missbrauchs hoch ist</li> <li>• Objektivität als Grundhaltung, insbesondere in Situationen, in denen Datenlage und Fragestellung Spielraum für die Analyse lassen</li> </ul>	<ul style="list-style-type: none"> <li>• Kann mit grundlegenden statistischen Methoden wie etwa Mittelwert und Standardabweichung umgehen</li> <li>• Beherrscht und verwendet komplexere Analyseverfahren und erkennt die Grenzen der Analyseergebnisse, kann genau einschätzen, welche Methoden für welche Fragestellungen und Daten die besten Ergebnisse liefern</li> <li>• Beherrscht und verwendet hochgradig komplexe Modelle und erkennt, wenn die Analyse keine sinnvollen Ergebnisse liefern kann bzw. wann die Informationen der Analyse keine Relevanz für die Fragestellung haben und der Analyseprozess beendet werden sollte</li> </ul>



# Beispiel D.3 Datenanalysen interpretieren (Dekodieren = rezeptive Seite)

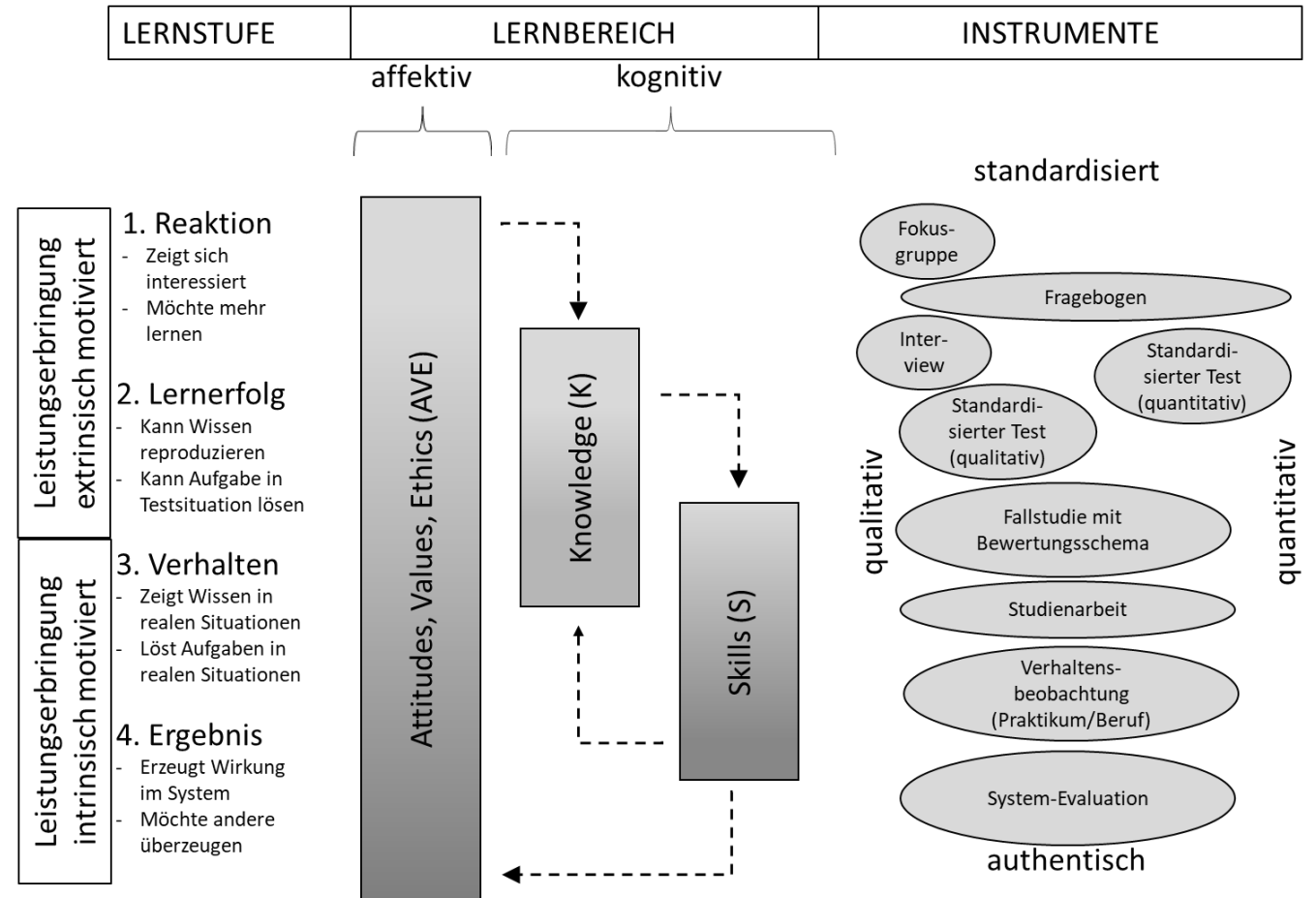
Kompetenz		Kompetenzdimensionen			Kompetenzniveaus
Bezeichnung	Beschreibung	Beispiele für Wissen („Knowledge“)	Beispiele für Fähigkeiten („Skills“)	Beispiele für Haltung („Attitude“)	Beispiele für aufsteigende Niveaus
D.1: Daten-Analysen interpretieren	Interpretiert Datenprodukte (Statistiken, Modellergebnisse) in verbalisierter Form bzw. prüft kritisch die explizit oder implizit gelieferte Interpretation	<ul style="list-style-type: none"> <li>Wissen über statistische Kennzahlen wie Mittelwerte, Prozente, Prozentpunkte, deren Aussagekraft und Einschränkungen</li> <li>Wissen über Zusammenhänge zwischen statistischen Kennzahlen und den zugrundeliegenden Daten</li> <li>Wissen über den Bezug zwischen Kennzahlen (z.B. relative/absolute Häufigkeiten)</li> <li>Wissen, dass die Wahl einer bestimmten Kennzahl das Ergebnis eines bewussten Entscheidungsprozesses sein kann</li> <li>Kenntnis statistischer Fachbegriffe</li> <li>Kenntnis statistischer Fehlschlüsse (z.B. Korrelation vs. Kausalität)</li> <li>Wissen, dass statistische Aussagen i.d.R. keine Einzelfallaussagen darstellen</li> <li>Wissen, dass die Interpretation das Hinzufügen von Kontextinformation erfordert</li> </ul>	<ul style="list-style-type: none"> <li>Kann Schlüsse ziehen, über welche Charakteristika der Daten eine Kennzahl Aussagen trifft</li> <li>Versteht, welche Kennzahlen (auch verbalisiert) äquivalent verwendet werden</li> <li>Kann hinterfragen, ob spezielle Datensituationen das Ergebnis beeinflussen</li> <li>Kann hinterfragen, ob die Darstellungsform die Aussage eines Datenprodukts beeinflusst</li> <li>Kann hinterfragen, inwiefern die Interpretation der Ergebnisse vom eigenen Kontextwissen abhängt</li> <li>Kann hinterfragen, inwiefern die Darstellung der Ergebnisse durch die Wahl der Form (z.B. Prozente) oder die Verbalisierung die Interpretation manipulativ beeinflusst</li> </ul>	<ul style="list-style-type: none"> <li>Bereitschaft, explizit kommunizierte, vorgegebene Interpretationen in Daten-Verbalisierungen zu hinterfragen</li> <li>Bereitschaft, nach implizit kommunizierten Interpretationen zu suchen und diese zu hinterfragen</li> <li>Bereitschaft, das eigene Kontextwissen in Bezug auf dessen Einfluss auf die Interpretation zu hinterfragen</li> <li>Offenheit gegenüber neuen Erkenntnissen, auch wenn diese den bisherigen Überzeugungen widersprechen</li> <li>Bereitschaft, die Bedeutung der Ergebnisse für den Sachverhalt, in den sie gestellt werden, zu hinterfragen</li> </ul>	<ul style="list-style-type: none"> <li>Kann einfache statistische Terminologie verstehen und deren Bezug zu den Daten interpretieren, kennt grundlegende Formen der Manipulation durch Statistiken und Berichte und die zu beachtenden Kriterien</li> <li>Besitzt ein fortgeschrittenes Verständnis der Terminologie und kann zwischen verschiedenen Begriffen sauber differenzieren, kennt Bausteine expliziter Kommunikation und kann prüfen, ob explizite Interpretationen aus den Ergebnissen ableitbar sind</li> <li>Besitzt ein vertieftes Verständnis der Terminologie, kann implizite Aussagen und Interpretationen erkennen, und zueinander sowie in Bezug auf den Sachverhalt abwägen und beurteilen</li> </ul>

## Vorauswahl von Instrumentarien

---

Einordnung und Impulse für die Testentwicklung

- Lernstufen-Modell nach Kirkpatrick
- Problem: Wie lassen sich Lernstufen messen, die gekennzeichnet sind durch eine intrinsische Motivation der Leistungserbringung, in einem System, das durch Noten einen extrinsischen Motivator einführt?
- Standardisierte Instrumente: Validierung und Benchmarking möglich
- Empfehlung: Multimethodischer Ansatz. Kombination von (mehrstufigem) Multiple Choice und Fallstudien mit Bewertungsschema



## Beispielaufgabe zu den Kompetenzen D.1 „Daten-Verbalisierungen interpretieren“ und E.1 „Standardisierung entschlüsseln“

Unter Bezugnahme auf die Allbright-Studie „Die Macht der Monokultur“ titelte das Manager-Magazin:

„Der Frauenanteil in den Vorständen steigt (um 0,7 Prozentpunkte)“. Sie erhalten folgende zusätzlichen Informationen:

1. Der Frauenanteil ist um 9,75% gestiegen.
2. Der Frauenanteil betrug zuvor 7,3%.

*Frage 1: Wie groß ist der Frauenanteil in den Vorständen heute?*

*Frage 2: Wie viele Frauen sind heute in den Vorständen?*

- A) Information 1 reicht aus, um die Frage zu beantworten.*
- B) Information 2 reicht aus, um die Frage zu beantworten.*
- C) 1 und 2 werden zusammen benötigt, um die Frage zu beantworten.*
- D) 1 oder 2 alleine reichen jeweils aus, um die Frage zu beantworten.*
- E) 1 und 2 zusammen reichen nicht aus, um die Frage zu beantworten.*

*Richtig sind D (Frage 1) und E (Frage 2).*

Die Original-Schlagzeile lautete dabei fälschlicherweise: „Der Frauenanteil in den Vorständen steigt (um 0,7 Prozent).“

## Beispielaufgabe zusätzlich zu den Kompetenzen C.3 „Daten verbalisieren“

Eine weitergehende Testfrage könnte also die (falsche) Originalschlagzeile zusammen mit den Rohdaten der Studie präsentieren und fragen:

- A) *Ist die Aussage der Schlagzeile korrekt? Begründen Sie Ihre Antwort.*
- B) *Handelt es sich Ihrer Meinung nach um eine starke oder eine schwache Veränderung? Begründen Sie Ihre Antwort.*
- C) *(Wie) würde sich ihre Antwort zu B) verändern, wenn die Schlagzeile gelautet hätte: „Der Frauenanteil in den Vorständen steigt (um 9,75 Prozent).“*
- D) *Wie würden Sie die Schlagzeile formulieren, um zu signalisieren, dass Sie die Frauenquote befürworten?*
- E) *Wie würden Sie sie formulieren, um zu signalisieren, dass Sie die Frauenquote ablehnen?*

Vielen Dank für Ihre  
Aufmerksamkeit!

---

