

Entwicklung eines Instruments zur Erfassung von Data Literacy

Sebastian Stuppan^{1,2}, Glena Iten¹, Roland Künzle¹ und Markus Wilhelm^{1,2}
Pädagogische Hochschule Luzern¹, Pädagogische Hochschule Heidelberg²
Kontakt: Sebastian Stuppan, sebastian.stuppan@phlu.ch

Hintergrund

Relevanz von Data Literacy

- Die **PISA 2025-Rahmenkonzeption** betont die wachsende **Bedeutung des Umgangs mit Daten** bei der Analyse komplexer Phänomene des Anthropozäns. **Daten sind zentral für das Verständnis globaler Herausforderungen** wie Klimawandel, Wirtschaft und Migration (Schüller & Busch, 2019).
- Die **fortschreitende Digitalisierung** verstärkt **diesen Trend** durch stark zunehmende **Datengenerierung und -verarbeitung** (Brinkmüller & Scheidig, 2024; Cukier & Mayer-Schönberger, 2013).
- Im **Kontext der „Nature of Science“-Diskussion** und des forschenden Lernens (z. B. Citizen Science) **gewinnt Data Literacy im Unterricht weiter an Bedeutung**.
- Studien zeigen, dass (angehende) Lehrpersonen im Bereich Natur und Technik häufig nur über begrenzte Data Literacy verfügen** (Evers & Lauermann, 2019; Michos, Schmitz & Petko, 2023; Reeves & Chiang, 2018; Wu, Xu & Zhang, 2021).

Konzeptualisierung und theoretische Einordnung von Data Literacy

- Data Literacy – Datengrundverständnis** – umfasst die Fähigkeit, Daten kontextbezogen kritisch zu erfassen, zu verwalten, zu bewerten und zu nutzen (Ridsdale et al., 2015).
- Die **Swiss Data Literacy Charta** betont zusätzlich die **Bedeutung digitaler Selbstbestimmung und einer kritisch-reflexiven Datennutzung** (Swiss Data Literacy Charta, 2024).
- Die Forschung zeigt eine **zunehmende Differenzierung theoretischer Modelle von Data Literacy**. Zwei aktuelle **Rahmenmodelle integrieren kognitive, verhaltensbezogene und affektive Komponenten** (Fokus auf K12: Friedrich et al., 2024; Fokus auf Lehrpersonen: Lee et al., 2024).
- Aufgrund der **interdisziplinären Natur von Data Literacy ist sie besonders im MINT-Kontext relevant** (Grillenberger, 2018; Vance et al., 2022).

Rahmenmodell & Fragestellung

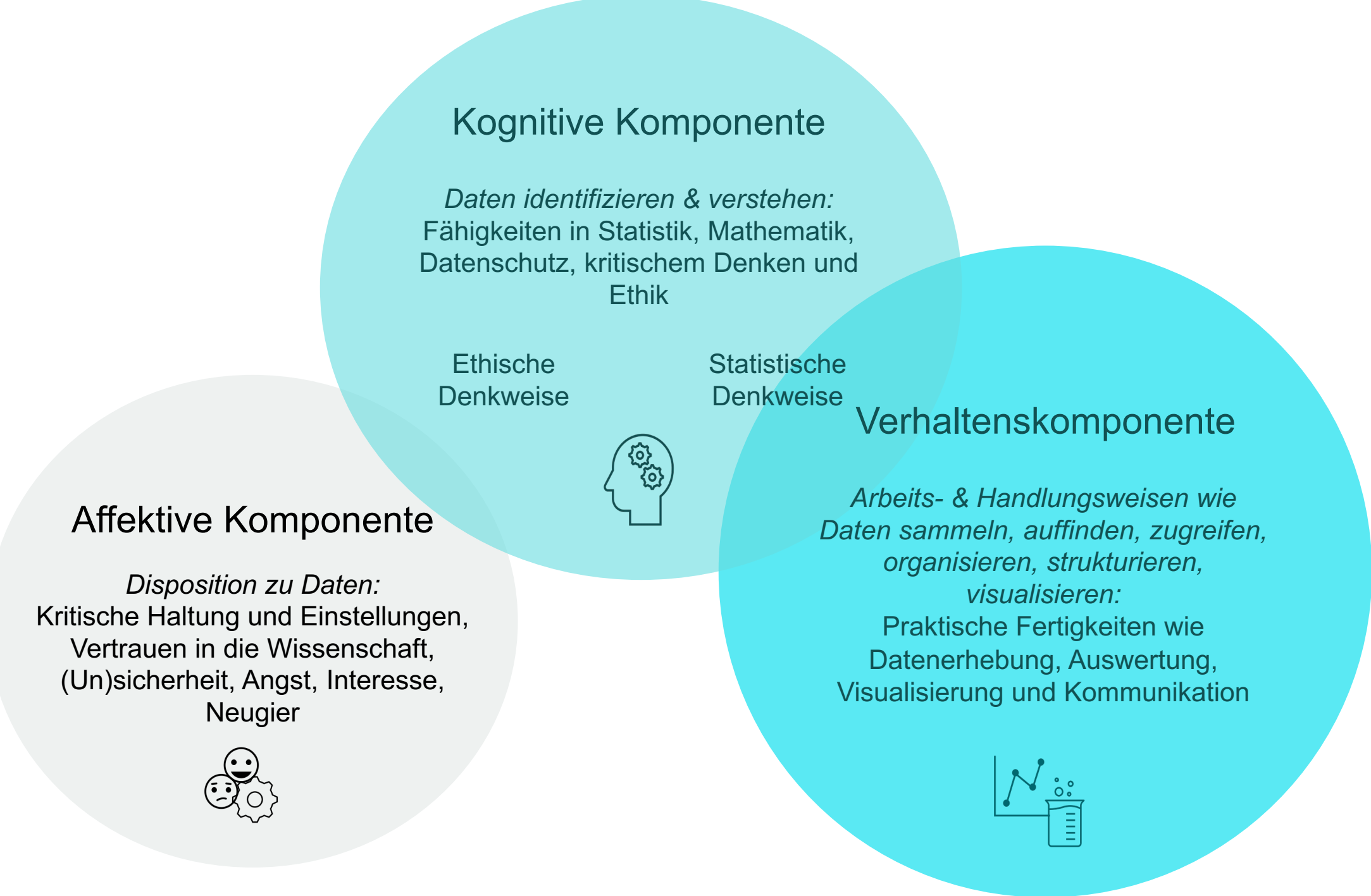


Abb. 1 Rahmenmodell zur Konzeptualisierung von Data Literacy (in Anlehnung an Friedrich et al., 2024, S. 5, Fig. 2; Lee et al., 2024, S. 184, Fig. 3; Lund et al., 2026, S. 4, Fig. 1) mit den zehn Subkomponenten von Kim et al. (2024, 2025) aufgeteilt in die drei Komponenten Kognition, Verhalten und Affekt.



Übergeordnete Ziele:

- Die vorgeschlagenen Data Literacy Komponenten identifizieren
- Unterschiede in diesen Komponenten nach Fächerwahl (Naturwissenschaften vs. andere Bereiche) sichtbar machen
- Passende Förderangebote entwickeln



Fragestellung zur Skalenentwicklung:

- Lässt sich die Selbstbeurteilung der multidimensionalen Data Literacy entsprechend dem Rahmenmodell valide und reliabel erfassen?

Methode

Stichprobe

- 101 Studierende der PH Luzern wurden rekrutiert (71% weiblich; 90% zwischen 18 & 24 Jahren)

Methode zur Skalenentwicklung

- Befragung im Rahmen einer Modulveranstaltung zu Alltag und Wissenschaft (Forschungspraktikum zu quantitativen Methoden).
- Durchführungsort: PH Luzern, **Studierende im ersten Semester PS und S1**, freiwillig und anonym.
- Modellfacetten** (vgl. Komponenten der Data Literacy in Abb. 1) in Form einer online Fragebogenbefragung, randomisierte Item-Reihenfolge, 5-stufige Ratingskala: 0 = stimmt nicht bis 4 = stimmt sehr, **33 Items zur Data Literacy** (angepasst auf Grundlage von Kim et al., 2025) und **5 Items zu Trust in Science and Scientists** (Krüger et al., 2022).

Ausgewählte Ergebnisse

Fokus: Kognitive Komponente

Tab. 1: (N = 101): Reliabilitätswerte und Itemtexte der Subkomponenten

Subskala	Itemtext	M	SD	Schiefe	Schwierigkeits-index	α wenn gelöscht
Data Awareness						
DA11	...definieren, was Daten sind.	2.55	0.87	-0.12	0.64	0.61
DA12	...verschiedenen Datentypen und -formate beschreiben.	2.13	0.87	-0.07	0.53	0.63
DA13	...den aktuellen und zukünftigen Wert von Daten erklären. / Den aktuellen und potenziellen Nutzen von Daten erklären.	2.14	0.86	-0.47	0.53	0.66
Mittlere Inter-Item-Korrelation = 0.47; Cronbachs α = 0.72; McDonalds ω = 0.72						
Data Quality Evaluation						
DQE71	... Datenquellen danach beurteilen, ob sie die definierten Kriterien erfüllen.	2.26	0.80	-0.26	0.56	0.51
DQE72	... beurteilen, ob Daten für einen bestimmten Zweck die erforderliche Qualität haben.	2.36	0.79	-0.49	0.59	0.62
DQE73	... Daten kritisch bewerten, um sicher zu stellen, dass sowohl Datentyp wie -menge für den vorgesehenen Zweck geeignet sind.	2.42	0.77	-0.46	0.60	0.61
Mittlere Inter-Item-Korrelation = 0.41; Cronbachs α = 0.68; McDonalds ω = 0.66						
Data Storage (Understanding, Data Management)						
DS91	... zwischen verschiedenen Datenspeicheroptionen unterscheiden.	1.88	0.98	0.05	0.47	0.61
DS92	... verschiedene Datensicherungsmethoden und -strategien festlegen.	1.91	0.81	-0.06	0.48	0.62
DS93	... die besten Speicheroptionen für die kurz- und langfristige Datenspeicherung bestimmen.	1.88	0.89	0.24	0.47	0.55
Mittlere Inter-Item-Korrelation = 0.43; Cronbachs α = 0.69; McDonalds ω = 0.66						
Data Ethics						
DE101	... Daten sicher, vertraulich und rechtkonform verarbeiten (z. B. DSGVO, Einwilligungserklärung).	2.16	1.02	-0.56	0.54	0.67
DE102	... Techniken zur De-Identifizierung (z. B. Anonymisierung) nutzen, um den Schutz der Persönlichkeit zu wahren. / ... um den Schutz sensibler Daten zu gewährleisten.	2.31	1.04	-0.15	0.58	0.61
DE103	... Eigentums- und Nutzungsrechte von Daten bestimmen (z. B. geistiges Eigentum und Lizenzen).	1.73	0.92	-0.24	0.58	0.57
DE104	... die aktuellen Best Practices zur Datenzitation anwenden (z. B. Verwendung von Digital Object Identifiers, DOI).	1.14	0.76	-0.10	0.38	0.47
Mittlere Inter-Item-Korrelation = 0.36; Cronbachs α = 0.69; McDonalds ω = 0.63						

Anmerkung. Die Items beginnen immer mit folgender Einleitung: «Inwiefern stimmen die folgenden Aussagen für dich? Ich denke, ich kann erfolgreich...». Für die Reliabilitätsanalyse wurden zusätzlich jeweils die McDonalds Omega Werte berechnet, um sie mit den Cronbachs Alpha Werten vergleichen zu können (Robustheit gegenüber Annahmen der testtheoretischen Modelle).

Beispiel einer individuellen Auswertung bezüglich der vorgeschlagenen Data-Literacy-Subkomponenten >> <https://drive.switch.ch/index.php/s/6iypRKHplwJwugu>



Diskussion

- Das Modell zeigt einen **mässigen Fit**; Während RMSEA und SRMR akzeptable Werte erreichen, bleiben CFI und TLI unter dem empfohlenen Grenzwert von .90 (Hu & Bentler, 1999).
- Die **kognitive und Verhaltenskomponente korrelieren stark**, was auf eine fehlende empirische Trennschärfe und mögliche **inhaltliche Überlappungen der Konstrukte** hinweist.
- Zudem zeigen sich **teilweise niedrige Faktorladungen** einzelner Items, die die Modellgüte negativ beeinflussen können (λ = .35–.89).
- Die **kognitive Komponente weist eine akzeptable Reliabilität** auf (α = .68–.72) und kann damit als hinreichend intern konsistent betrachtet werden.
- Die mittleren Inter-Item-Korrelationen liegen im empfohlenen Bereich (r_{ij} = .36–.46) und sprechen für eine grundsätzlich kohärente Operationalisierung der Subskalen.
- Trust in Science and Scientists**, als Stellvertreter der affektiven Komponente, weist **keine signifikanten Zusammenhänge** mit den übrigen Konstrukten auf und ist daher nur begrenzt in das **Modell integriert**.
- Einschränkend** wirkt insbesondere die **kleine Stichprobengrösse** (N = 101).

Fazit: Insgesamt deuten die Befunde auf **strukturelle Unschärfen im Messmodell** hin. Für **zukünftige Analysen** wird daher eine **theoriegeleitete Überarbeitung empfohlen**, insbesondere die **Prüfung einer Zusammenlegung stark korrelierter Konstrukte** sowie die **Optimierung bzw. Reduktion einzelner Items**.