LEHRPULS · FORSCHUNGS-KIT

LehrPuls Forschungs-Kit

Datenmodell, Spaltenvertrag, Kennzahlen, Beispielanalysen und Hinweise zu Ethik und Datenschutz für Hochschuldidaktik-Forschende und Abschlussarbeiten. Stand 8. November 2026, LehrPuls 1.21.1.

1. Wozu dieses Kit

LehrPuls ist ein Programm für Live-Fragen im Hörsaal: Die Lehrperson stellt Fragen vom eigenen Laptop, die Teilnehmenden antworten mit dem Smartphone, und jede Antwort wird mit Zeitpunkt, Bearbeitungsdauer und Bewertung lokal in einer SQLite-Datenbank gespeichert. Daraus entsteht über ein Semester ein Datensatz, der für die hochschuldidaktische Forschung interessant ist: dieselbe Frage in mehreren Durchgängen, erste und zweite Antwort bei Peer Instruction, Antwortzeiten je Person, Konfidenzangaben, Themenzuordnung jeder Frage.

Dieses Kit beschreibt, wie Sie an diese Daten kommen, was jede Spalte bedeutet, wie LehrPuls seine Kennzahlen rechnet (damit Ihre Analyse mit den Zahlen im Programm übereinstimmt) und welche Regeln für Einwilligung, Anonymität und Löschung gelten. Dazu gehören ein synthetischer Beispieldatensatz und je ein Analyse-Skript in R und Python, die ohne LehrPuls-Installation laufen.

Alle Angaben in diesem Dokument stammen aus dem Quelltext von LehrPuls 1.21.1 und dem Handbuch (Abschnitte 4.5, 8.1, 8.12 und 10.9).

2. Drei Wege zu den Daten

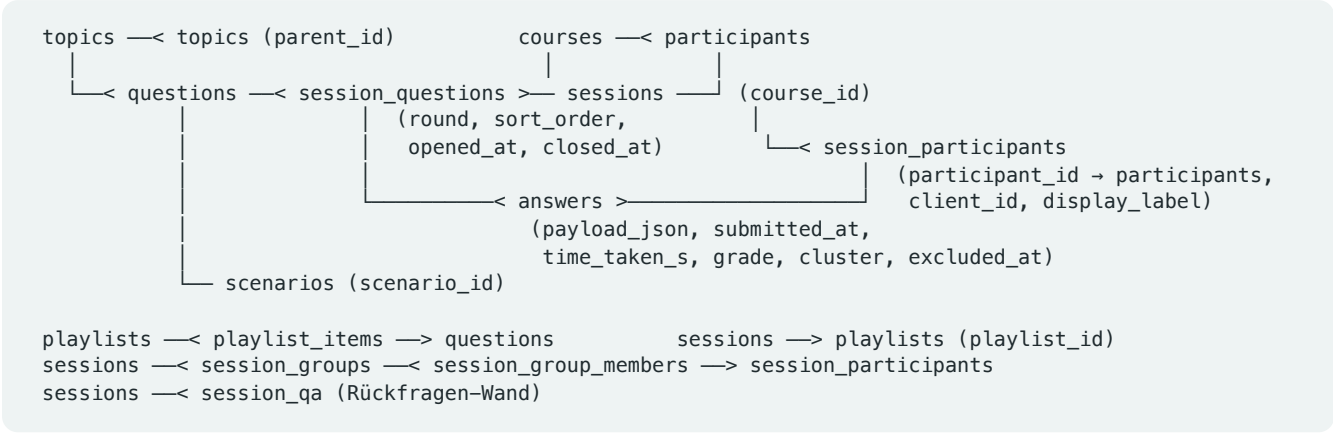
Weg	Was Sie bekommen	Wann sinnvoll
Rohdaten-Export (Datei)	Excel-Datei mit drei Blättern (Antworten, Fragen, Sessions) oder CSV mit den Antworten; eine Zeile je Antwort (Langformat)	Einmalige Übergabe eines Semesters an ein Forschungsprojekt, Archivierung, SPSS, jamovi, JASP
Lese-Schnittstelle `/api/v1`	Dieselben drei Tabellen mit denselben Spaltennamen als JSON oder CSV, dazu drei Kennzahl-Tabellen; nur lesend, befristeter Schlüssel	Laufende Erhebungen, Skripte, die nach jeder Stunde abholen, Dashboards
Beispieldatensatz dieses Kits	Synthetische CSV-Dateien im Format des Rohdaten-Exports	Analyseplan und Skripte entwickeln, bevor echte Daten vorliegen; Lehre der Methoden

Beide echten Wege liefern dieselben Spalten in derselben Reihenfolge; die Spaltennamen sind fest und sprachunabhängig, ein Vertrag, den spätere Versionen nur erweitern, nicht ändern (neue Spalten kommen hinten dazu). Der Rohdaten-Export steht in LehrPuls nur im Expertenmodus zur Verfügung (Einstellungen), ebenso die Lese-Schnittstelle.

Die Datenbankdatei `lehrpuls.db` selbst ist eine gewöhnliche SQLite-Datei und lässt sich mit jedem SQLite-Werkzeug öffnen. Ihr Schema (Abschnitt 3) ist aber kein Vertrag: Es ändert sich mit Programmversionen, Antworten liegen dort als unbewertetes JSON, und die Bewertung (richtig, Punkte) entsteht erst beim Export gegen den aktuellen Stand der Frage. Arbeiten Sie deshalb mit Export oder Schnittstelle.

3. Datenmodell

3.1 Beziehungsdiagramm



Lesart: `A ──< B` heißt „ein A hat viele B“. Die Kette für jede Auswertung lautet `answers → session_questions → questions → topics` (Inhalt) und `answers → session_participants → participants → courses` (Person und Kurs). Themen bilden über `parent_id` einen Baum; eine Themenauswertung summiert immer den Teilbaum.

3.2 Entitäten

Die Spalten folgen dem Datenbankschema von LehrPuls 1.21.1 (Schemaversion 56). Aufgeführt sind die Spalten, die für Auswertungen tragen; interne Spalten (Sortierung, Beispieldaten-Markierung, Übersetzungen `i18n_json`, Staffelstab-Kennung `uid`) sind weggelassen.

topics (Themen, Baum)

Spalte	Bedeutung
`id`	Kennung
`parent_id`	Oberthema; leer bei einem Hauptthema
`name`	Name; Pfad im Export als „Haupt > Unter“
`archived`	1 = archiviert

questions (Fragenpool)

Spalte	Bedeutung
`id`	Kennung; bleibt über alle Einsätze gleich (Grundlage des Längsschnitts)
`topic_id`	Thema; leer bei Sofortfragen ohne Thema
`qtype`	Fragetyp, einer von 19 Werten (Anhang A)
`title`, `body_md`	Titel und Fragetext (Markdown)
`options_json`	Optionen, Lösung, Toleranzen je Fragetyp
`solution_md`	Erklärung der Lösung
`difficulty`	Einschätzung der Lehrperson (ganzzahlig), wird im Programm nicht gegen die gemessene Schwierigkeit gestellt
`time_budget_s`	vorgesehene Dauer in Sekunden

Spalte	Bedeutung
`tags`, `source`	Schlagwörter; Herkunft bei importierten Fragen
`archived`	1 = archiviert; gespielte Fragen sind nur archivierbar, nicht löschbar

courses (Kurse) und **participants** (Kursliste)

Spalte	Bedeutung
`courses.id`, `name`, `semester`	Ein Kurs gehört zu genau einem Semester (Text, z. B. „WS 2026/27“); dieselbe Lehrveranstaltung im nächsten Semester ist ein neuer Kurs
`courses.purge_after`	Löschfrist (Datum): ab dann erinnert LehrPuls daran, die Namen zu entfernen (Abschnitt 9)
`participants.id`, `course_id`, `display_name`	Person der Kursliste; `import_ref` ist die Kennung aus der importierten Liste (z. B. Matrikelnummer)
`participants.active`	0 = ausgetreten

sessions (eine Unterrichtseinheit)

Spalte	Bedeutung
`id`	Kennung
`course_id`	Kurs; leer bei einer Session ohne Kurs
`join_mode`	`anonymous`, `pseudonym` oder `named`; mit `join_verified` = 1 (Name plus Matrikelnummer geprüft) erscheint im Export `verified`
`started_at`, `ended_at`	Beginn und Ende, UTC
`playlist_id`, `playlist_name`	Playlist, aus der die Fragen kamen; der Name bleibt auch nach Löschen der Playlist erhalten
`question_lang`	Sprache, in der die Fragen gestellt wurden
`probe`	1 = Probelauf (erscheint nicht in Auswertungen)

session_participants (Teilnahme einer Person an einer Session)

Spalte	Bedeutung
`id`	Kennung
`participant_id`	Verweis in die Kursliste; leer bei anonymer Teilnahme oder Gästen
`display_label`	Angezeigter Name oder Pseudonym
`client_id`	Gerätekennung; `plenum` ist die Plenumsantwort der Lehrperson, `group-…` eine Gruppe. Beide zählen in keiner Statistik als Person

session_questions (eine gestellte Frage in einer Session)

Spalte	Bedeutung
`id`	Kennung (`session_frage_id` im Export)
`question_id`	die Frage aus dem Pool
`round`	Runde; 1 = erste Abstimmung, 2 = zweite Runde bei Peer Instruction (eigene Zeile mit eigener Position)
`sort_order`	Position in der Session; im Export als `frage_nr` = `sort_order` + 1

Spalte	Bedeutung
`opened_at`, `closed_at`, `revealed_at`	Frage geöffnet, geschlossen, Lösung aufgedeckt (UTC); Standzeit = `closed_at` – `opened_at`
`excluded_at`	gesetzt, wenn die Lehrperson die Frage aus der Auswertung ausgeschlossen hat
`shuffle`	gesetzt, wenn die Antwortoptionen in dieser Session gemischt gezeigt wurden

answers (eine Antwort)

Spalte	Bedeutung
`session_question_id`, `session_participant_id`	Frage und Person; je Paar höchstens eine Antwort (eindeutig)
`payload_json`	Rohantwort als JSON, z. B. `{"choice":"a"}`, `{"choices":["a","c"]}`, `{"value":38.6}`, `{"text":"…"}`
`submitted_at`	Zeitpunkt der Abgabe (UTC)
`time_taken_s`	Bearbeitungszeit, am Gerät der Person gemessen; kann fehlen
`grade`, `grade_comment`, `graded_at`	Wertung von Hand in Prozent mit Begründung (offene Formate)
`cluster`	Antwortgruppe, in die die Lehrperson eine offene Antwort einsortiert hat
`excluded_at`	gesetzt, wenn diese Antwort aus der Auswertung ausgeschlossen wurde

Weitere Tabellen: `playlists` und `playlist_items` (Fragenblöcke und ihre Reihenfolge), `playlist_time_log` (Stoppuhr je Playlist-Knoten), `session_groups` und `session_group_members` (Gruppenphasen), `session_qa` (Rückfragen-Wand mit Zeitstempel, Reichweite, Stimmen), `scenarios` (Fallbeschreibungen, die mehrere Fragen teilen), `einstellungen` (Schlüssel-Wert, darunter die Mindestgröße der Auswertungen).

3.3 Was der Export aus dem Modell macht

Der Export fügt die Kette aus Abschnitt 3.1 zu einer flachen Tabelle zusammen, bewertet jede Antwort gegen den **aktuellen** Stand der Frage (`korrekt`, `punkte`) und ersetzt Personen durch stabile Pseudonyme. Zwei Eigenschaften sind für Forschungsdesigns entscheidend:

- **Bewertung gegen den aktuellen Fragenstand.** Wer eine Lösung nachträglich korrigiert, ändert rückwirkend die Statistik aller früheren Einsätze. Für Vorher-Nachher-Vergleiche einer überarbeiteten Frage legen Sie die neue Fassung als neue Frage an (neue `frage_id`).
- **Pseudonyme sind nur über die Kursliste stabil.** Der Anonymisierer vergibt die Nummern T1, T2, … in der Reihenfolge des ersten Auftretens; Schlüssel ist die Kursliste-Kennung der Person (`participant_id`). Bei anonymer Teilnahme oder Gästen ist der Schlüssel die Teilnahmezeile der Session, dieselbe Person erhält also in jeder Session eine neue Nummer. Längsschnitte **je Person** setzen daher namentliche Teilnahme (`named` oder `verified`) voraus; Längsschnitte **je Frage** funktionieren in jedem Beitrittsmodus.

---

4. Spaltenvertrag des Rohdaten-Exports

Die Spaltennamen sind ein fester Vertrag des Rohdaten-Exports. Dieselben Spalten liefern `/api/v1/antworten`, `/api/v1/fragen` und `/api/v1/sessions`.

4.1 Tabelle „Antworten“ (eine Zeile je Antwort)

Nr.	Spalte	Typ	Bedeutung
1	`nutzer_id`	Text	Zufällige Kennung dieser LehrPuls-Installation (UUID). Sessions, Fragen und Teilnehmende sind nur zusammen mit ihr eindeutig, wenn Sie Exporte mehrerer Lehrender zusammenführen
2	`nutzer`	Text	Name der Lizenz; leer mit „Nutzer anonymisiert“
3	`session_id`	Zahl	Session
4	`session_start_utc`	Text	Beginn der Session, `JJJJ-MM-TT HH:MM:SS` in UTC
5–7	`kurs_id` , `kurs` , `semester`		Kurs; leer ohne Kurs
8–9	`playlist_id` , `playlist`		Playlist; leer bei Sofortfragen
10	`beitrittsmodus`	Text	`anonymous` , `pseudonym` , `named` oder `verified`
11	`teilnehmer_id`	Text	Stabiles Pseudonym `T1` , `T2` , … (anonymisiert) oder Schlüssel `P<participant_id>` beziehungsweise `S<session_participant_id>` (Klarnamen-Export)
12	`teilnehmer`	Text	„Teilnehmer 1“ … oder der angezeigte Name; bei Gruppen der Gruppenname
13	`gruppe`	0/1	1 = die Zeile ist eine Gruppenantwort, keine Person
14	`session_frage_id`	Zahl	Kennung der gestellten Frage in dieser Session (je Runde eigene Kennung)
15	`frage_nr`	Zahl	Position in der Session, ab 1
16	`runde`	Zahl	1 oder 2 (Peer Instruction)
17	`frage_id`	Zahl	Frage im Pool; gleich über alle Sessions
18	`fragetyp`	Text	Anhang A
19	`frage_titel`	Text	Titel
20–21	`thema_id` , `thema`		Thema; `thema` als Pfad „Haupt > Unter“; leer ohne Thema
22	`antwort`	Text	Antwort lesbar (Optionstexte, Zahl mit Einheit, Freitext, Bildpunkt in Prozent)
23	`antwort_zahl`	Zahl	Nur Zahlen-, Schätz- und Skalenfragen: der Zahlenwert
24	`antwort_roh`	Text	Rohantwort als JSON (`payload_json`)
25	`korrekt`	0/1	Alles-oder-nichts; leer bei nicht bewertbaren Fragetypen
26	`punkte`	Zahl	Erfüllungsgrad 0 bis 1 mit Teilpunkten (Mehrfachauswahl, Reihenfolge, Zuordnung, Lückentext); folgt `wertung_hand` , wo gesetzt
27	`dauer_s`	Zahl	Bearbeitungszeit in Sekunden; kann leer sein
28	`gesendet_am_utc`	Text	Zeitpunkt der Abgabe, UTC
29	`ausgeschlossen`	0/1	1 = Frage oder Antwort ist aus der Auswertung ausgeschlossen. Der Export behält die Zeile; wer selbst rechnet, filtert auf 0
30–31	`wertung_hand` , `kommentar`		Wertung von Hand in Prozent und Begründung; leer = keine
32	`skizze_svg`	Text	Skizze einer offenen Antwort als SVG; sonst leer
33	`optionen_reihenfolge`	Text	Bei gemischten Optionen: Kennungen in der gezeigten Reihenfolge, z. B. `c,a,d,b`
34	`datei`	Text	Name einer angehängten Datei; die Datei selbst liegt in der Medienablage

Nr.	Spalte	Typ	Bedeutung
35–36	`bilanz_richtig`, `bilanz_gewertet`	Zahl	Bilanz der Person in der ganzen Session (Abschnitt 6.7); in jeder Zeile der Person gleich; leer bei Gruppen
37	`antwortgruppe`	Text	Gruppe, in die die Lehrperson eine offene Antwort einsortiert hat
38	`konfidenz`	Text	`sicher`, `unsicher` oder `geraten`; leer, wenn die Frage nicht danach fragte oder nichts gewählt wurde
39	`segment`	Text	Merkmal der Person aus der Merkmalsfrage der Session (z. B. „ja“ bei „Vorlesung besucht“); leer ohne Merkmalsfrage
40 ff.	`fragensprache`, `frage_titel_en`, `antwort_en`, `thema_en`, `playlist_en`, danach je weiterer Sprache `frage_titel_<Code>`, `antwort_<Code>`, `thema_<Code>`, `playlist_<Code>`	Text	Sprache der Session und Fassungen in anderen Sprachen; Englisch ist immer vorhanden, weitere Sprachen nur, wenn eine exportierte Frage eine Fassung darin hat

4.2 Tabelle „Fragen“ (Codebuch, eine Zeile je Frage, die in den Antworten vorkommt)

`nutzer_id`, `nutzer`, `frage_id`, `fragetyp`, `titel`, `text`, `optionen`, `loesung`, `thema_id`, `thema`, `vorgesehene_dauer_s`, `schwierigkeit` (Einschätzung der Lehrperson), `tags`, `titel_en`, `text_en`, `optionen_en`, `loesung_en`, `mischen` (0/1, „Antworten je Session mischen“; leer bei Typen ohne Auswahl), `alles_oder_nichts` (0/1, nur Mehrfachauswahl), danach je Sprache `thema_<Code>`.

4.3 Tabelle „Sessions“ (eine Zeile je abgeschlossener Session)

`nutzer_id`, `nutzer`, `session_id`, `start_utc`, `ende_utc`, `kurs_id`, `kurs`, `semester`, `playlist_id`, `playlist`, `beitrittsmodus`, `teilnehmende` (Personen ohne Plenum und Gruppen), `fragen` (tatsächlich gestellte Fragen), `fragensprache`, danach je Sprache `playlist_<Code>`.

4.4 Dateiformate

	Rohdaten-Export (Datei)	Lese-Schnittstelle `?format=csv`
Trennzeichen	Strichpunkt `;`	Komma `,`
Dezimalzeichen	nach Anzeigesprache: Komma bei Deutsch, Punkt bei Englisch	immer Punkt
Kodierung	UTF-8 mit Byte-Order-Mark, Zeilenende CRLF	UTF-8
Umfang	XLSX: drei Blätter Antworten, Fragen, Sessions; CSV: nur Antworten	je Abruf eine Tabelle
Leere Werte	leere Zelle	leere Zelle (CSV) beziehungsweise `null` (JSON)

In R: `read.csv(datei, sep = ";", dec = ",", fileEncoding = "UTF-8-BOM")` für den deutschen Export; in Python: `pd.read_csv(datei, sep=";", decimal=",", encoding="utf-8-sig")`. Die Skripte dieses Kits erkennen das Dezimalzeichen selbst.

5. Die Lese-Schnittstelle `/api/v1`

Quelle: Handbuch 10.9.

Betrieb. Die Schnittstelle ist ausgeschaltet, bis die Lehrperson sie in den Einstellungen einschaltet (Expertenmodus), und antwortet nur, solange LehrPuls läuft. Sie lauscht auf `127.0.0.1` am ersten freien Port aus 8716 bis 8719; mit dem Schalter „Im Netz erreichbar“ auch auf den anderen Adressen des Rechners,

unverschlüsselt. Die IT kann sie über die Umgebungsvariable `LEHRPULS_API_SPERRE=1` ganz abschalten. Sie liest nur: Es gibt ausschließlich GET-Abrufe.

Zugang. Ein Zugangsschlüssel (`lp_…`) gilt 1, 7, 30 oder höchstens 90 Tage; LehrPuls speichert nur seinen SHA-256-Prüfwert. Der Schlüssel gehört in den Kopf jeder Anfrage: `Authorization: Bearer lp_…` . In der Adresse wird er nicht angenommen. Ohne gültigen Schlüssel antwortet die Schnittstelle mit 401, bei unbekannter Tabelle mit 404, bei ungültigen Parametern mit 400 und einem JSON-Objekt mit `code` und `message` . CORS ist freigegeben, Browser-Seiten dürfen also direkt abrufen.

Vorgaben beim Einschalten: Teilnehmende anonymisiert **ein**, Nutzer anonymisiert aus, Beispiele einschließen aus, Mindestgröße anwenden aus, Zeitangaben genau, im Netz erreichbar aus.

Endpunkte

Adresse	Inhalt
`/api/v1`	Programmname, Version, `api: 1` , Ablauf des Schlüssels (`gueltig_bis_unix`), ob anonymisiert, Liste der Tabellen und Parameter
`/api/v1/sessions`	eine Zeile je abgeschlossener Session (Spalten wie 4.3)
`/api/v1/fragen`	eine Zeile je Frage, die in den Antworten vorkommt (Spalten wie 4.2)
`/api/v1/antworten`	eine Zeile je Antwort (Spalten wie 4.1)
`/api/v1/fragen-kennzahlen`	eine Zeile je gestellter Frage mit Fragengüte (unten)
`/api/v1/themen`	eine Zeile je Thema mit Reifegrad und Abdeckung
`/api/v1/bilanz`	eine Zeile je Person und Session mit richtig von gewertet

Parameter (nach dem Fragezeichen, mit `&` verbunden): `format=json|csv` ; `session_id=<Kennung>|letzte|heute` (auch `session=` , `last` , `latest` , `today`); `von=` , `bis=` als `JJJJ-MM-TT` , `heute` oder `gestern` ; `tage=<1…3660>` (die letzten N Tage einschließlich heute); `kurs=<Name>` oder `kurs_id=` ; `thema=<Name>` (samt Unterthemen, in jeder Sprache) oder `thema_id=` . Unbekannte Parameter werden mit 400 abgewiesen; mehrdeutige Themennamen ebenfalls, mit Nennung der Kennungen. `fragen-kennzahlen` und `themen` rechnen über einen Zeitraum und nehmen kein `session_id` an.

Kennzahl-Tabellen (Spalten sind Vertrag):

- `fragen-kennzahlen` : `frage_id` , `frage_titel` , `fragetyp` , `thema_id` , `thema` , `einsaetze` (Sessions, in denen die Frage lief), `antworten` , `schwierigkeit_p` (0–1, drei Nachkommastellen), `trennschaerfe` (–1 bis 1), `tote_distraktoren` , `pi_p_runde1` , `pi_p_runde2` (leer ohne zweite Runde), `zuletzt_gestellt` .
- `themen` : `thema_id` , `thema` , `ebene` (0 = Hauptthema), `antworten` , `reifegrad_pct` , `gespielte_fragen` , `aktive_fragen` , `abdeckung_pct` . Die Zeile mit leerem Thema ist der Sammelposten „ohne Thema“.
- `bilanz` : `session_id` , `session_start_utc` , `kurs_id` , `kurs` , `teilnehmer_id` , `teilnehmer` , `richtig` , `gewertet` , `quote_pct` .

Die Schutzschalter „Fragen unter der Mindestgröße weglassen“ und „Zeitangaben runden“ gelten für `antworten` , `sessions` und `bilanz` (Abschnitt 9.3); die beiden anderen Kennzahl-Tabellen liefern ohnehin nur Summen.

Fertige Abrufbeispiele in R, Python, Power Query und als Browser-Seite liegen auf der Download-Seite von lehrpuls.com unter „Werkzeuge für eigene Auswertungen“ (`lehrpuls-kennzahlen.R` , `lehrpuls-kennzahlen.ipynb` , `lehrpuls-kennzahlen.pq` , `lehrpuls-api-demo.html`). Die Skripte dieses Kits

bauen darauf auf: Sie holen mit `schnittstelle` als erstem Argument die Antworten-Tabelle direkt ab und rechnen dieselben Kennzahlen wie auf dem Beispieldatensatz.

---

6. Kennzahlen: Definitionen und Formeln

So rechnet LehrPuls. Wenn Sie dieselben Regeln anwenden, stimmen Ihre Werte mit dem Reiter „Fragen“ der Auswertungen und mit `/api/v1/fragen-kennzahlen` überein. Grundlage ist immer: `ausgeschlossen = 0`, `gruppe = 0`, keine Plenumsantworten, nur abgeschlossene Sessions, keine Probeläufe.

Zwei Grenzen gelten durchgängig: Die **Mindestbasis** n = 10 Antworten, unter der das Cockpit Kennzahlen als „zu wenig Daten“ ausgraut und keine Hinweise gibt (Konstante `MINDESTBASIS`), und die **Mindestgröße** für Verteilungen (Vorgabe 5 Personen, einstellbar 1 bis 100), unter der die Auswertungen keine Antwortverteilung je Frage zeigen und die als Schutzschalter beim Export wirkt.

6.1 Schwierigkeit P

P = Anzahl richtiger Antworten ÷ Anzahl bewertbarer Antworten, **nur Runde 1**, **alles-oder-nichts** (`korrekt`), über alle Sessions des Zeitraums. Bewertbar ist eine Antwort, wenn `korrekt` gesetzt ist. Hinweise im Cockpit: P > 0,9 „zu leicht“, P < 0,3 „zu schwer“. Hoher Wert heißt leichte Frage (Lösungswahrscheinlichkeit, nicht Schwierigkeit im Wortsinn).

6.2 Trennschärfe (Upper-Lower-Index)

Je Session: Gesamtscore jeder Person = Anzahl richtiger Antworten in Runde 1. Die Personen werden nach Score sortiert (bei Gleichstand nach Teilnahme-Kennung); obere und untere Gruppe umfassen je g = round(0,27 · n) Personen, mindestens 1, höchstens n ÷ 2. Für jede Frage der Session: D = Anteil richtig in der oberen Gruppe – Anteil richtig in der unteren Gruppe. Die Trennschärfe der Frage ist das **Mittel von D über alle Sessions**, in denen die Frage lief. Warnsignal: D ≤ 0 bei n ≥ 10 („die Guten scheitern“: Frage missverständlich oder Lösung falsch hinterlegt).

Für Forschungszwecke ist der Upper-Lower-Index grob: Bei kurzen Sessions (8 Fragen) trägt die Frage selbst stark zum Gesamtscore bei. Die Skripte rechnen zusätzlich die **punktbiseriale Korrelation** zwischen `korrekt` und dem Rest-Score derselben Person in der Session (Gesamtscore ohne diese Frage).

6.3 Tote Distraktoren

Eine falsche Option einer Auswahlfrage gilt als tot, wenn sie in Runde 1 seltener gewählt wurde als max(1, 2 % der bewertbaren Antworten in Runde 1), und nur, wenn $n_1 \geq 10$. `tote_distraktoren` zählt diese Optionen je Frage. Stark gewählte falsche Optionen (typische Fehlvorstellungen) lesen Sie aus `antwort_roh` beziehungsweise `antwort` ab.

6.4 Peer-Instruction-Gewinn

`pi_p_runde1` = Anteil richtig über alle bewertbaren Antworten der Frage in Runde 1 im Zeitraum, `pi_p_runde2` dasselbe für Runde 2; ausgegeben nur für Fragen mit Antworten in beiden Runden. Zu beachten: Runde 1 zählt dabei auch Sessions, in denen keine zweite Runde stattfand. Für eine gepaarte Auswertung je Person verknüpfen Sie die Zeilen mit `runde = 1` und `runde = 2` derselben `session_id`, `frage_id` und `teilnehmer_id` und zählen die Wechsel falsch → richtig und richtig → falsch (McNemar-Test); so machen es die Skripte. Eine zweite Runde ist in LehrPuls eine eigene gestellte Frage (`session_frage_id`, `frage_nr`) mit `runde = 2`.

6.5 Erfüllungsgrad und Themen-Reifegrad

Der **Erfüllungsgrad** einer Antwort ist `punkte` (0 bis 1, mit Teilpunkten). Der **Reifegrad** eines Themas (`reifegrad_pct`) ist der mittlere Erfüllungsgrad aller bewertbaren Antworten auf Fragen des Themen-Teilbaums im Zeitraum, in Prozent. Stufen im Cockpit: unter 50 % „im Aufbau“, 50–75 % „in Arbeit“, ab 75 % „gefestigt“; unter n = 10 ausgegraut. **Abdeckung** (`abdeckung_pct`) = gespielte aktive Fragen ÷ aktive Fragen des Teilbaums × 100. Beide Zahlen stehen bewusst nebeneinander und werden nicht zu einem Index verrechnet.

6.6 Antwortzeiten

Drei Zeitbegriffe: **Bearbeitungszeit** = `dauer_s` je Antwort (am Gerät gemessen, von Öffnen der Frage bis Absenden; robust mit dem Median auswerten, fehlende Werte verwerfen), **Standzeit** = `closed_at` – `opened_at` der gestellten Frage (nur in der Datenbank und über die Lese-Schnittstelle, nicht im Rohdaten-Export), **Zeitpunkt** = `gesendet_am_utc`, aus dem sich die Lage der Antwort in der Einheit relativ zu `session_start_utc` ergibt. Bearbeitungszeiten sind rechtsschief; rechnen Sie auf der logarithmischen Skala.

6.7 Bilanz der Teilnehmenden

`bilanz_richtig` von `bilanz_gewertet` je Person und Session: Gewertet sind Fragen, die aufgelöst wurden und bewertbar sind oder von Hand gewertet wurden; bei Peer Instruction zählt Runde 2; Ausgeschlossenes zählt nicht. `quote_pct` = richtig ÷ gewertet × 100. Die Bilanz gilt für die ganze Session, auch wenn der Export nach Thema gefiltert ist.

6.8 Nicht bewertbare Fragetypen

Ein `korrekt` tragen Einfach- und Mehrfachauswahl, Zahl, Reihenfolge, Zuordnung, Lückentext, Formel, die Bildfrage (nur mit Zielflächen) und Mitteilungen mit Ja/Nein-Erwartung. Alle übrigen Typen — offene Frage, Umfrage, Schätzung, Wortwolke, Ideen, Skala, Rangfolge, Punktebudget, 2×2-Raster — tragen kein `korrekt` (außer bei Wertung von Hand). Für sie tritt die **Beteiligung** (Antworten ÷ Teilnehmende der Session) an die Stelle der Korrektheit; Schätzfragen werten Sie als Abweichung vom Referenzwert aus (`antwort_zahl`).

---

7. Beispielanalysen

Die Skripte `analyse.R` (nur Basis-R, keine Pakete) und `analyse.py` (pandas, matplotlib) laden den Beispieldatensatz oder einen eigenen Export, rechnen die Kennzahlen aus Abschnitt 6 nach und schreiben `kennzahlen-fragen.csv`, `laengsschnitt.csv` und eine vierteilige Grafik (`analyse-R.png`, `analyse-Python.png`). Beide Skripte wurden am 8. November 2026 auf dem Beispieldatensatz ausgeführt (R 4.6.1; Python 3.14 mit pandas über `uv run --with pandas --with matplotlib analyse.py`) und liefern dieselben Werte für P, Peer-Instruction-Gewinn und Längsschnitt; die Trennschärfe weicht leicht ab (Frage 102: −0,27 in R, −0,23 in Python), weil R und Python Gleichstände beim Sortieren der Gesamtscores verschieden auflösen. LehrPuls selbst löst Gleichstände nach der Teilnahme-Kennung auf.

Analyse	Was gerechnet wird	Ergebnis auf dem Beispieldatensatz
Item-Analyse (Gütekarte)	P und Trennschärfe je Frage, Hinweise wie im Cockpit, punktbiseriale Korrelation	40 Fragen; eine Frage mit Trennschärfe rund –0,25 (Lösung falsch hinterlegt), eine zu leicht (P = 0,91), drei zu schwer
Peer-Instruction-Gewinn	`pi_p_runde1` / `pi_p_runde2` wie LehrPuls; gepaart je Person mit Wechselzählung und McNemar-Test	16 Fragen mit zweiter Runde; mittlerer gepaarter Gewinn +0,22 (Anteil richtig)
Längsschnitt derselben Frage	P je Frage und Semester (nur Runde 1), Spannweite über die Semester	zwei Fragen werden von Semester zu Semester schwerer (Drift 0,70 → 0,43), eine leichter
Antwortzeit-Verteilungen	Median je Fragetyp; Single Choice nach Runde und Richtigkeit; Histogramm auf log-Skala	Median 17 s Single Choice, 73 s offene Frage; falsche Antworten brauchen länger (24 s gegen 17 s); Runde 2 ist schneller

Weitere Analysen, für die der Datensatz vorbereitet ist: **Kalibrierung** (Anteil richtig je `konfidenz`-Stufe `sicher`, `unsicher`, `geraten`, Brier-Score; sechs Fragen je Kurs fragen nach der Sicherheit), **Merkmalsvergleich** (`segment` „ja“/„nein“ zu „Vorlesung besucht“ im dritten Semester des ersten Kurses), **Beteiligungsverlauf** (Antworten je `frage_nr` relativ zu `teilnehmende` aus der Sessions-Tabelle), **Distraktorverteilung** aus `antwort_roh`.

Aufruf (im Ordner `skripte`): `Rscript analyse.R` beziehungsweise `python3 analyse.py`; mit Pfad zu einer eigenen CSV als Argument; oder `LEHRPULS_SCHLUESSEL=lp_… Rscript analyse.R schnittstelle "kurs=Physik 1&tage=120"` für den direkten Abruf aus einer laufenden Schnittstelle (ungetestet gegen eine laufende Instanz; die Abruflogik entspricht `lehrpuls-kennzahlen.R`).

---

8. Der Beispieldatensatz

`beispieldaten/antworten.csv` (18 870 Zeilen, 44 Spalten), `fragen.csv` (40 Fragen) und `sessions.csv` (30 Sessions) im Format des Rohdaten-Exports bei deutscher Oberfläche. Erzeugt von `erzeuge-beispieldaten.mjs` mit festem Seed (`node erzeuge-beispieldaten.mjs` liefert byte-identische Dateien).

Aufbau: zwei Lehrveranstaltungen („Wirtschaftsinformatik 1“, Beitritt `named`; „Rechnungswesen 2“, Beitritt `verified`) über drei Semester (WS 2025/26, SS 2026, WS 2026/27), je Semester ein eigener Kurs mit 62 bis 118 Teilnehmenden (533 Personen insgesamt, Pseudonyme T001 bis T533), fünf Sessions je Kurs und Semester im Abstand von zwei Wochen, acht Fragen je Session rollierend aus einem Pool von 20 Fragen je Lehrveranstaltung (14 Single Choice, 2 Mehrfachauswahl, 2 Zahlenantwort, 1 offene Frage, 1 Schätzung). Jede Frage läuft also zwei- bis dreimal je Semester und sechs- bis achtmal insgesamt.

Erzeugungsmodell: Fähigkeit θ je Person normalverteilt mit kleinem Kohortenversatz je Semester; Lösungswahrscheinlichkeit logistisch in a·(θ – b) mit Schwierigkeit b und Trennschärfe a je Frage. Acht Single-Choice-Fragen je Lehrveranstaltung sind Peer-Instruction-Kandidaten; eine zweite Runde findet statt, wenn in Runde 1 zwischen 30 und 70 % richtig lagen (Faustregel des Handbuchs). In Runde 2 bleibt eine richtige Antwort mit 92 % richtig, eine falsche wird mit einer Wahrscheinlichkeit richtig, die mit dem Anteil richtiger Erstantworten steigt. Bearbeitungszeiten sind lognormal, kürzer bei hoher Fähigkeit und richtiger Antwort, in Runde 2 kürzer. Konfidenz ist mit der Richtigkeit korreliert.

Eingebaute Effekte zum Wiederfinden: Frage 102 (Normalformen) hat eine negative Trennschärfe; 104 (SELECT mit WHERE) ist zu leicht; 107 (NULL in SQL) und 207 (Zuschlagssatz) zu schwer; 109 und 209 werden je Semester schwerer, 111 und 211 leichter (Drift); 113 und 213 haben einen toten Distraktor; 106 und 206 einen besonders attraktiven falschen Distraktor; rund 0,3 % der Zeilen sind ausgeschlossen (`ausgeschlossen = 1`); 30 % der offenen Antworten tragen eine Wertung von Hand.

Grenzen: Alle Inhalte sind erfunden (Fragetexte ohne Fachinhalt, Optionen „Option A“ bis „Option D“, Lösung immer A). Es gibt keine Gruppenantworten, Skizzen, Dateien oder Mehrsprachigkeit über Englisch hinaus. Die Pseudonyme sind mit führenden Nullen geschrieben (T001); der echte anonymisierte Export schreibt T1, T2, …

---

9. Ethik und Datenschutz

9.1 Was LehrPuls speichert und wo

Alle Daten liegen auf dem Rechner der Lehrperson (SQLite-Datei im Datenordner); LehrPuls überträgt nichts an Hersteller oder Dritte. Bei Online-Lehre läuft die Verbindung Ende-zu-Ende verschlüsselt über ein Relay, das die Inhalte nicht lesen kann. Teilnehmende brauchen kein Konto. Personenbezogen sind: Namen aus der Kursliste (`participants.display_name`, `import_ref`), Pseudonyme, die Personen selbst wählen, und alles, was Personen in offene Antworten, Rückfragen oder hochgeladene Dateien schreiben. Die Verknüpfung von Antworten zu Personen ist nur bei Beitritt `named` oder `verified` möglich.

9.2 Einwilligung

Die Teilnahme an Live-Fragen ist Lehre, keine Forschung. Sobald Antworten für eine Studie ausgewertet oder veröffentlicht werden sollen, braucht es eine informierte Einwilligung nach den Regeln Ihrer Einrichtung. Vier Empfehlungen: Einwilligung im Abhängigkeitsverhältnis Lehrperson–Studierende über eine **dritte Person** einholen; die Zuordnungsliste bis nach der **Notenvergabe versiegeln**; den **Interessenkonflikt** offenlegen, wenn Entwickler oder Anbieter von LehrPuls an der Studie beteiligt sind; den **Analyseplan vorab registrieren** (zum Beispiel OSF). LehrPuls liefert in der Kopfzeile der Kursseite einen Muster-Datenschutzhinweis zum Anpassen (Handbuch 4.5); er ist eine Vorlage, keine Rechtsberatung.

9.3 Anonymisierung und Schutz vor der Weitergabe

Bevor Daten den Rechner verlassen, bietet LehrPuls vier Stufen, die alle auch für die Schnittstelle gelten:

1. **Teilnehmende anonymisiert** (Vorgabe ein): Namen werden durch „Teilnehmer 1“, „Teilnehmer 2“ … ersetzt, Kennung `T1`, `T2` …; dieselbe Person behält ihre Nummer über alle Sessions des Exports, sofern sie über die Kursliste beigetreten ist. Die Nummern sind je Export neu vergeben, zwei Exporte lassen sich also nicht über T-Nummern verbinden.
2. **Nutzer anonymisiert**: Der Name der Lizenz (Lehrperson) bleibt leer; die zufällige Installationskennung bleibt.
3. **Fragen unter der Mindestgröße weglassen**: Antworten zu gestellten Fragen (`session_frage_id`) mit weniger Antworten als die Mindestgröße der Auswertungen (Vorgabe 5) fallen weg.
4. **Zeitangaben runden** auf die Stunde oder den Tag: Ein genauer Zeitstempel zusammen mit dem Stundenplan kann eine Person verraten. Gerundet werden `session_start_utc`, `gesendet_am_utc`, `start_utc` und `ende_utc`; die Bearbeitungsdauer bleibt.

Nicht anonymisiert werden Freitexte: Steht ein Name in einer offenen Antwort, einem Kommentar oder einer Datei, bleibt er dort. Sehen Sie offene Antworten vor der Weitergabe durch oder lassen Sie die Spalten `antwort`, `antwort_roh`, `kommentar`, `skizze_svg` und `datei` für offene Fragen weg.

9.4 Mindestgruppengröße in Veröffentlichungen

Berichten Sie keine Kennzahl, die auf weniger als der Mindestbasis beruht; LehrPuls selbst zeigt unter n = 10 keine Hinweise und graut Werte aus. Für Untergruppen (Segment, Konfidenzstufe, Kurs × Semester) gilt die Grenze je Zelle. Für die Planung von Studien gelten Richtwerte aus einer Poweranalyse (α = 0,05, Power 0,80):

gepaarter Vergleich mit d = 0,3 etwa 90 Personen, unabhängige Gruppen mit d = 0,5 etwa 64 je Gruppe, Korrelation r = 0,3 etwa 85 Personen; Mehrebenenmodelle mit Sessions oder Kursen als zweiter Ebene ab 50 Einheiten.

9.5 Löschfristen

Je Kurs lässt sich eine Löschfrist setzen (`purge_after` , typischerweise Semesterende). Ist sie erreicht, erinnert die Startseite daran; „Jetzt löschen“ entfernt die Namen aus der Kursliste, ersetzt sie in den Sessions durch Platzhalter, nimmt Rückfragen den Absender, löscht die Gruppen-Zusammenstellungen des Kurses und leert den Papierkorb für diesen Kurs. **Sessions und Antworten bleiben anonymisiert erhalten**, Längsschnitte je Frage also auch. Was Teilnehmende selbst geschrieben oder hochgeladen haben, bleibt unverändert. Planen Sie Forschungsexporte **vor** der Löschfrist, wenn Sie Personen über Sessions verknüpfen müssen; danach ist die Verknüpfung nur noch innerhalb der Platzhalternummern möglich; die Platzhalter hängen am Kursplatz und bleiben deshalb über Sessions hinweg stabil.

---

10. Offene Forschungsfragen

Zehn Forschungsfragen, die sich mit LehrPuls-Daten am direktesten bearbeiten lassen. Je Frage steht, welche Spalten sie beantworten.

1. **Peer Instruction außerhalb der Naturwissenschaften** (Replikation von Smith et al. 2009 in Wirtschaftsinformatik, Betriebswirtschaft, Recht): `runde` 1 und 2 je `teilnehmer_id` und `frage_id` , `korrekt` , dazu isomorphe Folgefragen als eigene `frage_id` .
2. **Antwortzeit als Lernindikator und die „80-%-Regel“** zum Schließen der Abstimmung: `dauer_s` je Antwort in beiden Runden, `korrekt` , `gesendet_am_utc` relativ zu `session_start_utc` .
3. **Metakognitive Kalibrierung auf Item-Ebene**: `konfidenz` (sicher, unsicher, geraten) gegen `korrekt` je Frage, Veränderung über das Semester und zwischen `runde` 1 und 2.
4. **Psychometrie formativer Live-Items**: `schwierigkeit_p` , `trennschaerfe` , `tote_distraktoren` aus `/api/v1/fragen-kennzahlen` , Distraktorwahl aus `antwort_roh` , Stabilität über `einsaetze` ; mit Einwilligung Bezug zur Prüfungsleistung über `teilnehmer_id` .
5. **KI-generierte gegen selbst geschriebene Fragen**: Spalte `source` der Fragen (Herkunftsdatei importierter Fragen) gegen P, Trennschärfe, Distraktorfunktion, `dauer_s` und Peer-Instruction-Gewinn.
6. **Verteilung vor der Diskussion zeigen oder nicht**: Wechselverhalten zwischen `runde` 1 und 2 je nach Bedingung; die Bedingung selbst muss derzeit außerhalb von LehrPuls protokolliert werden; gespeichert wird je Frage nur der Zeitpunkt der Auflösung (`revealed_at`), aus dem sich ablesen lässt, ob die Verteilung vor der zweiten Runde an der Wand stand.
7. **Aufmerksamkeitsverlauf über die Einheit**: Beteiligung (Antworten je `frage_nr` ÷ `teilnehmende`), `dauer_s` und `korrekt` in Abhängigkeit von der Lage in der Session (`gesendet_am_utc` – `session_start_utc`), über viele Sessions mit `playlist` als Kontrolle.
8. **Anonyme gegen namentliche Teilnahme**: `beitrittsmodus` (anonymous, pseudonym, named, verified) als Session-Bedingung im Crossover gegen Beteiligung, Streuung der Antworten, `dauer_s` , Länge offener Antworten.
9. **Drift derselben Frage über Kohorten**: `frage_id` × `semester` (nur Runde 1), getrennt nach Kohorteneffekt (`kurs` , `semester`) und Itemdrift (Bekanntwerden, veralteter Kontext).
0. **Überarbeitungszyklus von Fragen**: Trennschärfe und Peer-Instruction-Eignung (30–70 % richtig in Runde 1) vor und nach einer Überarbeitung; `antwortgruppe` offener Antworten als Quelle neuer Distraktoren; überarbeitete Fassung als neue `frage_id` .

11. Kooperationsangebot

Wir unterstützen Forschungsvorhaben und Abschlussarbeiten, die mit LehrPuls-Daten arbeiten: mit Testlizenzen für die Dauer der Erhebung, mit Erläuterungen zu Datenmodell und Kennzahlen, mit Rückmeldung zu Analyseplänen, und, wo ein Vorhaben eine Produkt-Ergänzung braucht (etwa das Protokollieren einer Versuchsbedingung je Frage), mit der Prüfung, ob sie in eine nächste Version passt. Daten verbleiben bei der Hochschule; wir erhalten keine Rohdaten, sofern Sie uns nicht ausdrücklich welche geben.

Kontakt: office@lehrpuls.com. Nennen Sie bitte Fragestellung, Fach, Zahl der Kurse und Teilnehmenden sowie den geplanten Zeitraum.

Anhang A: Fragetypen

`qtype` (19 Werte, Schema 1.21.1): `mc_single` (Single Choice), `mc_multi` (Mehrfachauswahl), `numeric` (Zahlenantwort), `estimate` (Schätzung), `scale` (Skala), `open` (offene Frage, auch mit Skizze oder Datei), `cloze` (Lückentext), `order` (Reihenfolge), `ranking` (Rangfolge), `matching` (Zuordnung), `hotspot` (Bildfrage), `poll` (Umfrage), `poll_multi` (Umfrage mit Mehrfachauswahl), `info` (Mitteilung), `wordcloud` (Wortwolke), `budget` (Punkte verteilen), `ideas` (Ideen sammeln), `grid` (Tabelle), `formula` (Formel).

Anhang B: Werte-Listen

Spalte	Werte
`beitrittsmodus`	`anonymous` , `pseudonym` , `named` , `verified`
`konfidenz`	`sicher` , `unsicher` , `geraten` , leer
`runde`	1, 2
`korrekt` , `gruppe` , `ausgeschlossen` , `mischen` , `alles_oder_nichts`	0, 1 (oder leer)
Zeitangaben	`JJJJ-MM-TT HH:MM:SS` in UTC; gerundet `JJJJ-MM-TT HH:00:00` oder `JJJJ-MM-TT`
Rundung in der Schnittstelle	`keine` , `stunde` , `tag`

Anhang C: Grenzen der Datenlage

- Antworten werden gegen den aktuellen Fragenstand bewertet (Abschnitt 3.3).
- `dauer_s` ist am Gerät der Person gemessen und kann fehlen; Verbindungsabbrüche verlängern sie.
- Sofortfragen ohne Thema fehlen in jeder Themenstatistik (Sammelposten „ohne Thema“).
- Plenumsantworten der Lehrperson und Gruppenantworten sind keine Personen.
- Eine Person kann in einer Session nur einmal je gestellter Frage antworten; Korrekturen überschreiben die Antwort (`updated_at` in der Datenbank).
- Die Standzeit einer Frage (`opened_at` , `closed_at`) ist nur in der Datenbank verfügbar, nicht im Export.
- Zwischen Exporten sind T-Nummern nicht stabil; innerhalb eines Exports nur bei Beitritt über die Kursliste.