

Skalierbare Kompetenzmessung für E-Assessment

Prof. Dr. Andreas Thor

Professur für Datenanalyse, Datenbanken und E-Learning
Hochschule für Technik, Wirtschaft und Kultur Leipzig
andreas.thor@htwk-leipzig.de

Gemeinsame Arbeit mit (in alphabetischer Reihenfolge)

Daniel Jenatschke, Marios Karapanos, Roy Meißner, Norbert Pengel, Heinz-Werner Wollersheim

What does research look like?

- Rahmenbedingungen für Forschung
 - im Team (Netzwerke)
 - Projektförderung (u.a. von EU, DFG, BMBF, SMWK, ...)
 - Agenda eines wissenschaftlichen Vortrags bzw. Papiers
 - Motivation und Problemstellung
 - Lösungsidee (konzeptionell)
 - Umsetzung (Implementation)
 - Evaluation
 - Zusammenfassung + Future Work
- Bezug zu
Related Work

E-Assessment-Item (Single-Choice-Beispiel)

Fall- oder Problemvignette

Gegeben seien die Relationenschemata $R(\underline{a}, b)$ und $S(\underline{c}, d)$. Attribut b ist ein Fremdschlüssel auf S mit der Löschregel *ON DELETE CASCADE*.

Aufgabenstellung

Bei welchem der folgenden Ereignisse kommt die Löschregel zur Anwendung?

Antwort-Text

Löschen eines Tupels in R.

Löschen eines Tupels in S.

Einfügen eines Tupels in R.


Einfügen eines Tupels in S.

Frage 1 von 10 - Löschregel (1 Punkt)

Gegeben seien die Relationenschema $R(\underline{a}, b)$ und $S(\underline{c}, d)$. Attribut b ist ein Fremdschlüssel auf S mit der Löschregeln *ON DELETE CASCADE*.

Bei welchem der folgenden Ereignisse kommt die Löschregel zur Anwendung?

- Löschen eines Tupels in R.
- Löschen eines Tupels in S.
- Einfügen eines Tupels in R.
- Einfügen eines Tupels in S.



Datenanalyse für E-Assessment

Item Nr.	Nutzer	
	A	B
1	Green	Green
2	Red	Green
3	Green	Red
4	Green	Green
5	Red	Red
6	Red	Green
7	Green	Red
8	Red	Red
9	Red	Green
10	Green	Red

- **Item-basierte** Analyse, u.a.
 - Schwierigkeitsgrad
 - Trennschärfe
 - ...
- **Nutzer-basierte** Analyse
 - Durchschnitts- / Gesamtpunktzahl
 - Rankings / Perzentilen
 - ...
- **Kompetenzprofile?**
- **Ausnutzen falscher Antworten?**

Annotation von Items mit Metadaten

Thema

Fall- oder Problemvignette ▲

Gegeben seien die Relationenschemata $R(a, b)$ und $S(c, d)$. Attribut b ist ein Fremdschlüssel auf S mit der Löschregel *ON DELETE CASCADE*.

Aufgabenstellung ▲

Bei welchem der folgenden Ereignisse kommt die Löschregel zur Anwendung?

Antwort-Text	Punkte	Aktionen	
Löschen eines Tupels in R.	0	+	-
Löschen eines Tupels in S.	1	+	-
Einfügen eines Tupels in R.	0	+	-
Einfügen eines Tupels in S.	0	+	-

- Funkt. Abh.
- NF1
- NF2
- NF3
- Relationenmodell
- Fremdschlüssel
- Löschregel
- Primärschlüssel

Anforderungsstufe

- 1. Erinnern
- 2. Verstehen
- 3. Anwenden
- 4. Analysieren
- 5. Evaluieren
- 6. Erschaffen

Thor, A.; Pengel, N.; Wollersheim, H.-W.: **Digitalisierte Hochschuldidaktik: Qualitätssicherung von Prüfungen mit dem E-Assessment-Literacy-Tool EAs.LiT.** Proc. of 15. E-Learning Fachtagung Informatik der Gesellschaft für Informatik (Delfi 2017)

Standing on the shoulders of giants

- Forschung baut aufeinander auf und ist häufig auch inkrementell, d.h.
 - bestehende Ansätze / Lösungen werden erweitert bzw. adaptiert
 - bestehende Ansätze werden kombiniert

Id	LastName	FirstName	Street	City	Gender
a ₁	Smith	Kristen	2 Hurley Pl	South Fork, MN	female
a ₂	Smith	Christian	Hurley St 2	S Fork MN	male
b ₁	Smith	Christoph	23 Harley St	Chicago, IL	male
b ₂	Smith	Kris L.	2 Hurley Place	South Fork, MN	female

- Beispiel: Duplikaterkennung
 - Levenshtein-Distanz zur Bestimmung ähnlicher Worte (durch Anzahl der Modifikationen)
 - Anpassung Kostenmodell auf Basis Tastaturlayout („a↔s“ < „a↔p“)
 - Verwendung / Aggregation mehrerer Ähnlichkeitsfunktionen

Kompetenzanalyse für E-Assessment mit Metadaten

Item			Nutzer	
Nr.	Thema	Anford. St.	A	B
1	RM	Erinnern	■	■
2	RM	Verstehen	■	■
3	SQL	Erinnern	■	■
4	RM	Erinnern	■	■
5	SQL	Verstehen	■	■
6	RM	Anwenden	■	■
7	SQL	Erinnern	■	■
8	SQL	Anwenden	■	■
9	RM	Verstehen	■	■
10	SQL	Erinnern	■	■

Item			Nutzer	
Nr.	Thema	Anford. St.	A	B
1	RM	Erinnern	■	■
3	SQL	Erinnern	■	■
4	RM	Erinnern	■	■
7	SQL	Erinnern	■	■
10	SQL	Erinnern	■	■
2	RM	Verstehen	■	■
5	SQL	Verstehen	■	■
9	RM	Verstehen	■	■
6	RM	Anwenden	■	■
8	SQL	Anwenden	■	■

Identifikation von Kompetenzen bei falschen Antworten

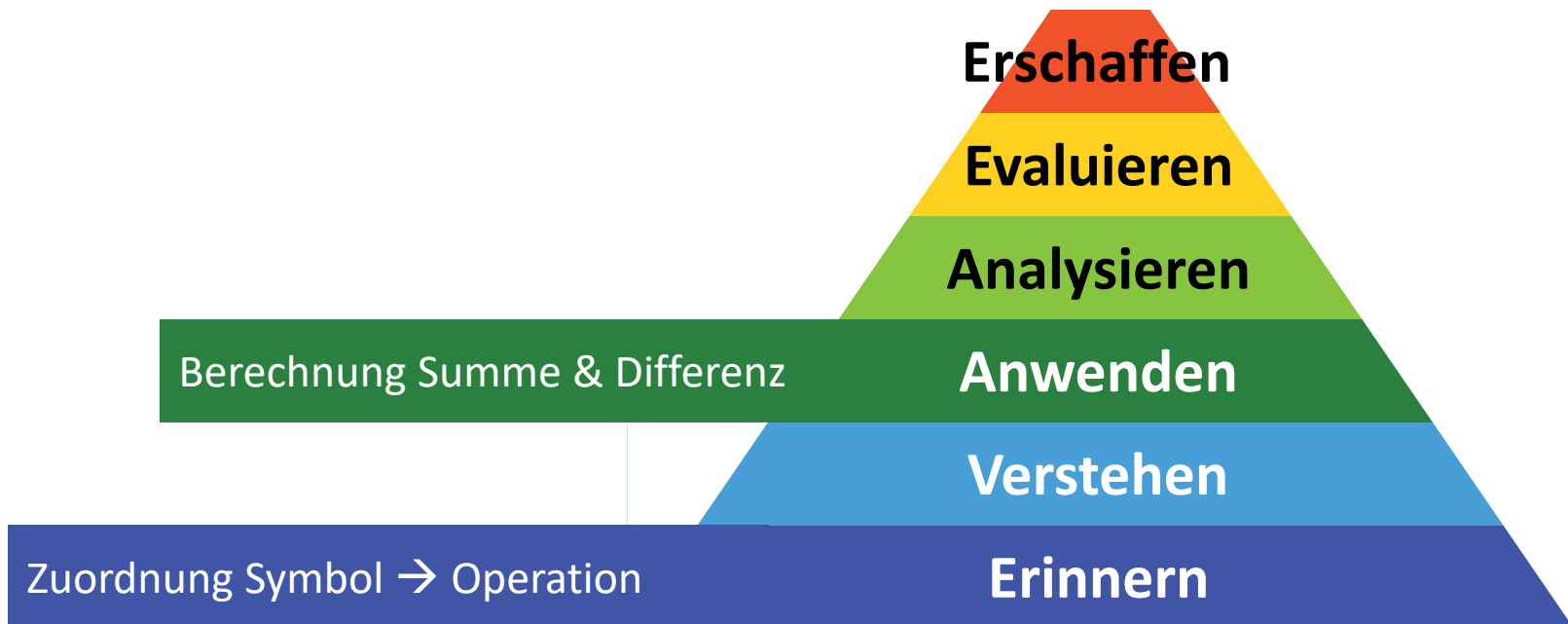
A

$$5 + 3 = 4 \quad \times$$
$$8 - 1 = 2 \quad \times$$

B

$$5 + 3 = 2 \quad \times$$
$$8 - 1 = 9 \quad \times$$

↑ ↓ ✓



Kompetenzanalyse durch E-Assessment

- **Annotation von Items** mit Metadaten (u.a. Anforderungsstufe) als Grundlage zur Identifikation von **Kompetenzprofilen** durch Datenanalyse / Data Mining
- Strukturierung von Items zur Identifikation von **Teilkompetenzen**, z.B. auf niedrigerer Anforderungsstufe, bei (teilweise) falschen Antworten

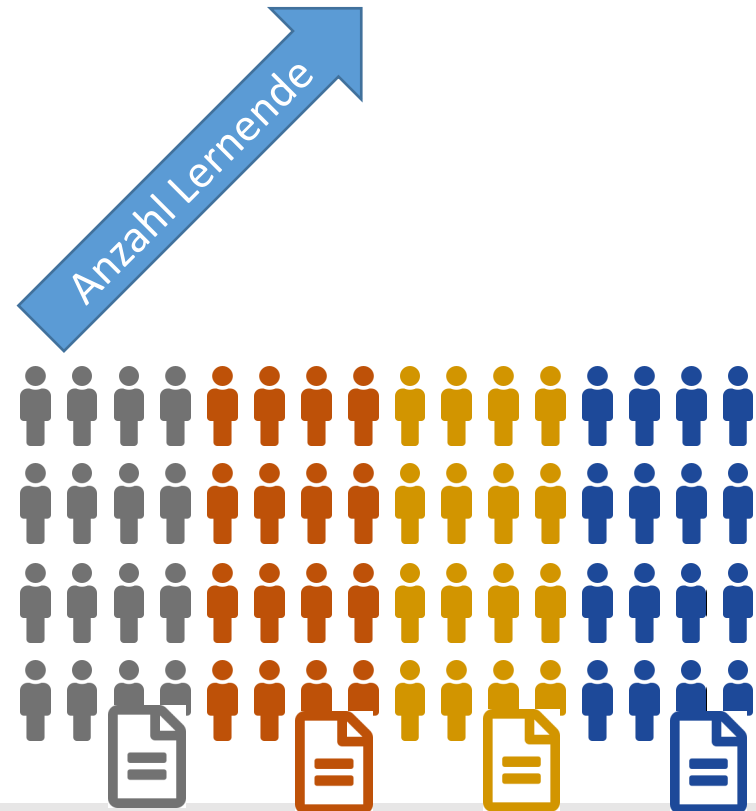
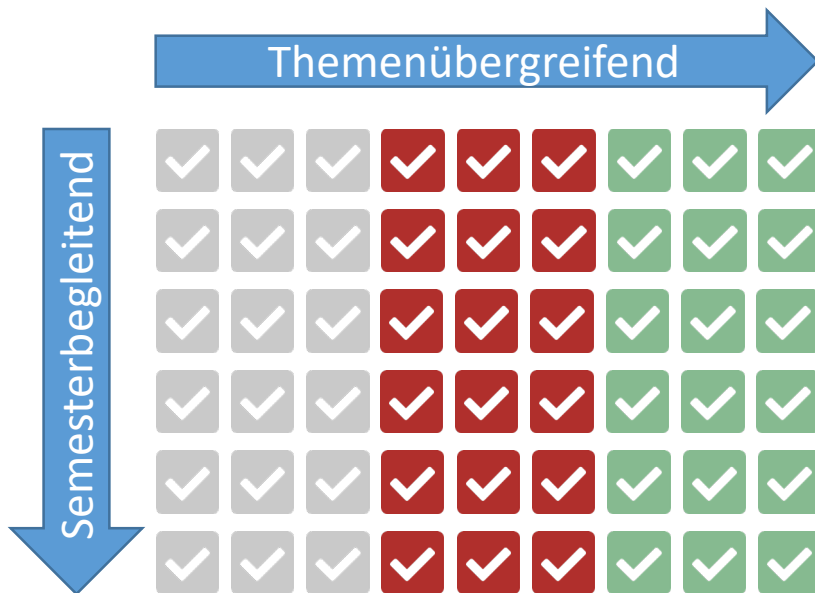
EAs.LiT

Das E-Assessment-Literacy Tool

www.easlit.de

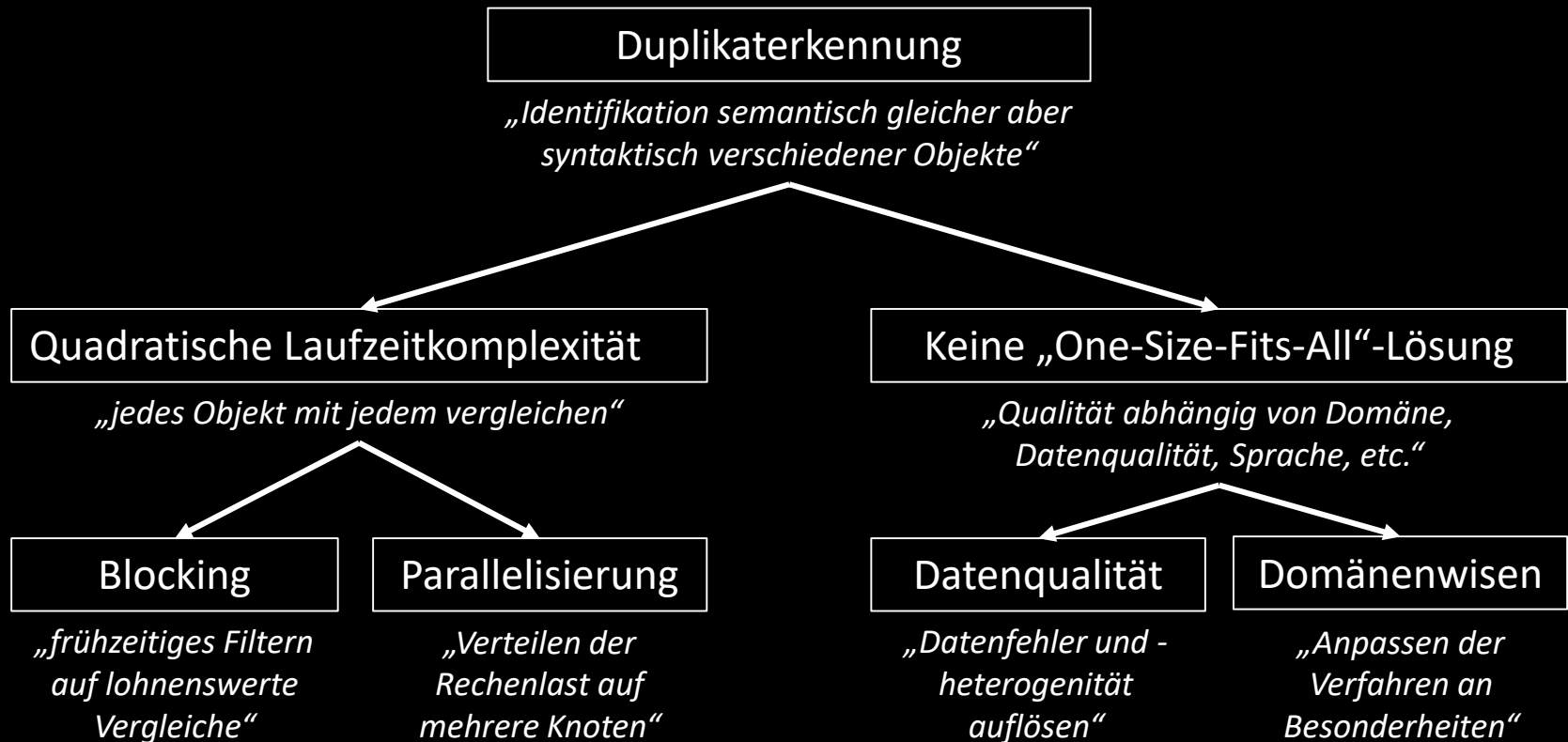
Qualität vs. Aufwand

- Qualität und Aufwand hängen maßgeblich von der Anzahl der Aufgaben, d.h. der Größe des Item Pools, ab



Research never stops

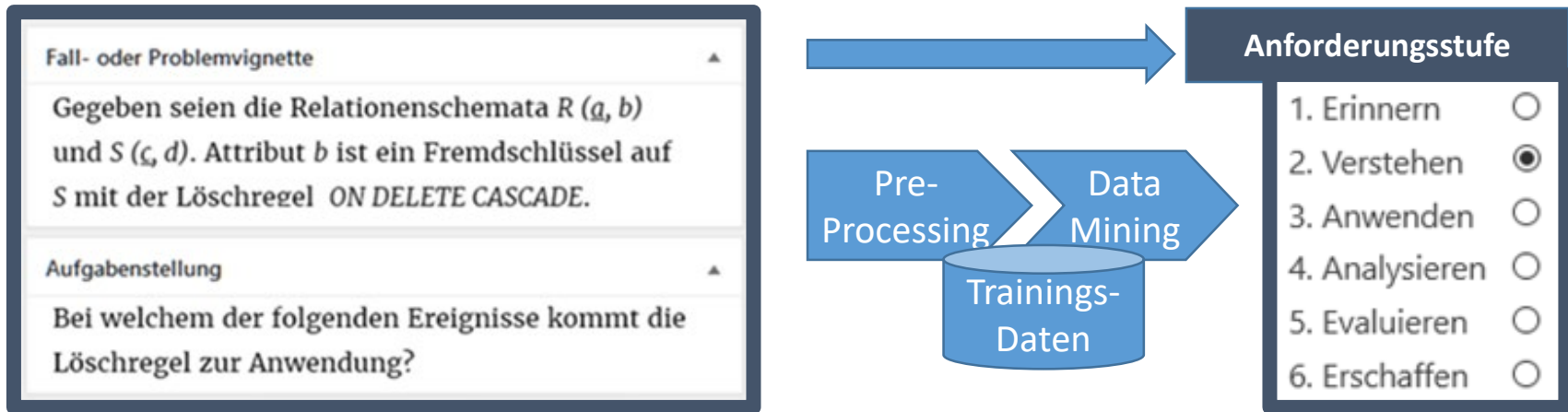
- Forschung: Lösen eines Problems führt (häufig) zum Auftreten weiterer (interessanter) Probleme und Lösungen



Skalierbare Kompetenzmessung

- Skalierbarkeit (IT): Fähigkeit eines Systems, die Leistung durch das Hinzufügen von Ressourcen (z. B. weiterer Hardware) proportional zu steigern
- „Ressource“ Lehrende nicht (beliebig) steigerbar, daher Leistungssteigerung durch
- **Automatische** Item-Verarbeitung
- Effiziente **Nutzerschnittstellen** zur Bulk-Verarbeitung von Items
- Flexible **Infrastruktur** zur Verwaltung von Items

Klassifikation der Anforderungsstufe



- Pre-Processing: Textanalyse (Natural Language Processing)
 - Identifikation von Substantiven, Verben, ...
 - Stammformen
- Data Mining: Nutzung von Klassifikations-Algorithmen
 - Naives Bayes, Entscheidungsbaum, Support Vector Machine, ...

Meissner, R.; Jenatschke, D.; Thor, A.: **Evaluation of Approaches for Automatic E-Assessment Item Annotation with Levels of Bloom's Taxonomy**. Proc. of 19th International Conference on Advances in Web-Based Learning (ICWL), 2020

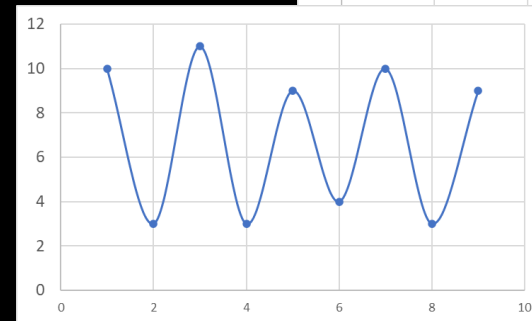
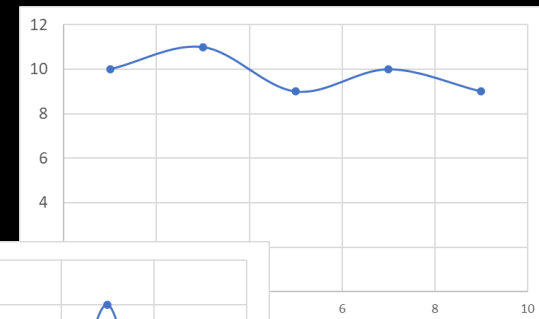
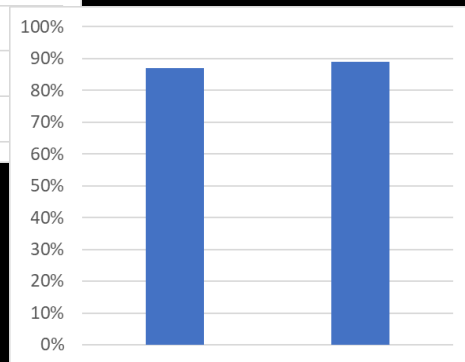
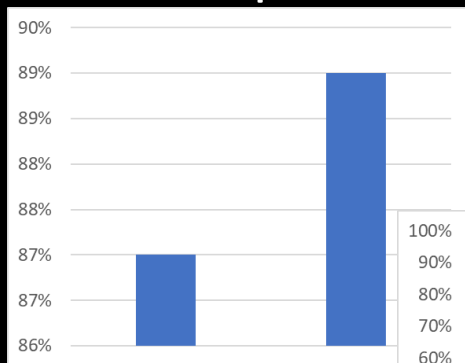
Erste Ergebnisse

- Items aus den Bereichen
 - Datenbanksysteme (DBS, n=83)
 - Bildungswissenschaften (BIWI, n=292)
- Manuelle Klassifikation für Trainings- und Testdaten

Domäne	DBS	BIWI	Mixed
Precision	89%	82%	85%
Recall	92%	87%	87%
F1 Score	90%	84%	86%

Always question research results

- Beispiel: Precision & Recall
 - Gibt es einen Bias in den Daten?
 - Gibt es Ground Truth? Wie wurden Testdaten erstellt?
 - Wie viel Aufwand wurde in Optimierung / Parametrisierung des Algorithmus gesteckt?
- Beispiel: Interpretation von Daten / Grafiken



Verarbeitung mit Excel

- Excel-Plugin als Konnektor

Items

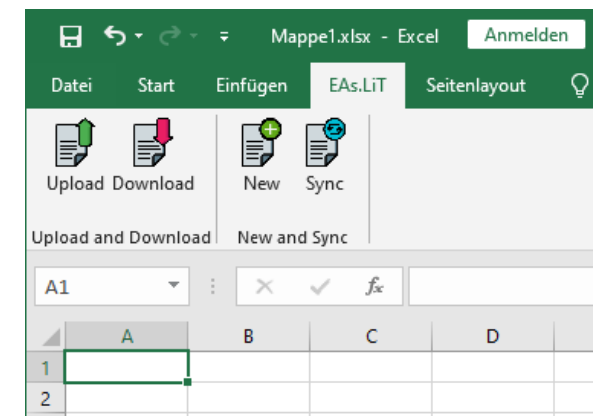
- Überprüfung / Korrektur Tippfehler
- Harmonisierung der Punktzahlen
- Erstellung ähnlicher Items

The screenshot shows the Excel Online interface with a table of test items. The table has columns for 'ealid', 'description', 'question', 'points', 'answer1', 'points1', 'answer2', 'points2', and 'answer3'. The 'EAs.Li' plugin sidebar is open on the right, showing options for 'File', 'Import format', and 'Export format'.

ealid	description	question	points	answer1	points1	answer2	points2	answer3
470	Die Länderrangfolge in PIRLS (dt. IGLU) wird maßgeblich durch die Punktzahl der mittleren Leseleistung bestimmt (siehe Grafik). Solche Rangfolgen sind jedoch nicht unproblematisch, da stets auch die Streuung der Testwerte beachtet werden muss. 	Welches Mittel wird bei PIRLS genutzt, um zu überprüfen, ob sich die Mittelwerte zweier Staaten überdurchschnittlich, d.h. signifikant, voneinander unterscheiden? Wählen Sie eine aus den vier Antwortoptionen aus.	3	Konfidenzintervalle	3	Benchmarks	0	Perzentilbänder
122	"Dem Statistischen Jahrbuch ist zu entnehmen, dass die Anzahl der Lebendgeborenen im Freistaat Sachsen von 22.734 im Jahre 1994 auf 35.091 im Jahre 2010 stetig gestiegen ist und seitdem etwa konstant ist. Aus dieser Geburtenentwicklung und der Zahl der in Ruhestand gehenden Lehrer/innen prognostiziert das Kultusministerium einen Einstellungsbedarf von ca. 1400 Lehrer/innen pro Jahr im Zeitraum 2016 - 2019."	Wenn im erziehungswissenschaftlichen Kontext über Schule gesprochen wird, können dort u. a. Funktionen, Aufgaben, Rahmenbedingungen und die bildungspolitische Bewertung von Schule thematisiert werden. Überlegen Sie, welcher dieser Bereiche im obigen Textausschnitt angesprochen wird. Wählen Sie eine aus den vier Antwortoptionen aus.	3	Es wird eine Rahmenbedingung von Schule angesprochen.	3	Es wird eine Funktion von Schule angesprochen.	0	Es wird eine Aufgabe von Schule angesprochen.
123	Schule hat vielfältige Funktionen. Wir betrachten folgenden Bereich: Soziale Verteilung von Zugangsberechtigungen, das sogenannte „Rüttelsieb“ zwischen den Generationen.	Welche der Funktionen von Schule nach Helmut Fend ist hier angesprochen? Wählen Sie eine aus den vier Antwortoptionen aus.	3	Enkulturationsfunktion	0	Qualifikationsfunktion	0	Allokationsfunktion

Karapanos, M.; Thor, A.; Wollersheim, H.-W.: **Itempool-Management mit Microsoft Excel: Eine UX-Studie**. Proc. of Gemeinschaften in Neuen Medien (GeNeMe), 2020

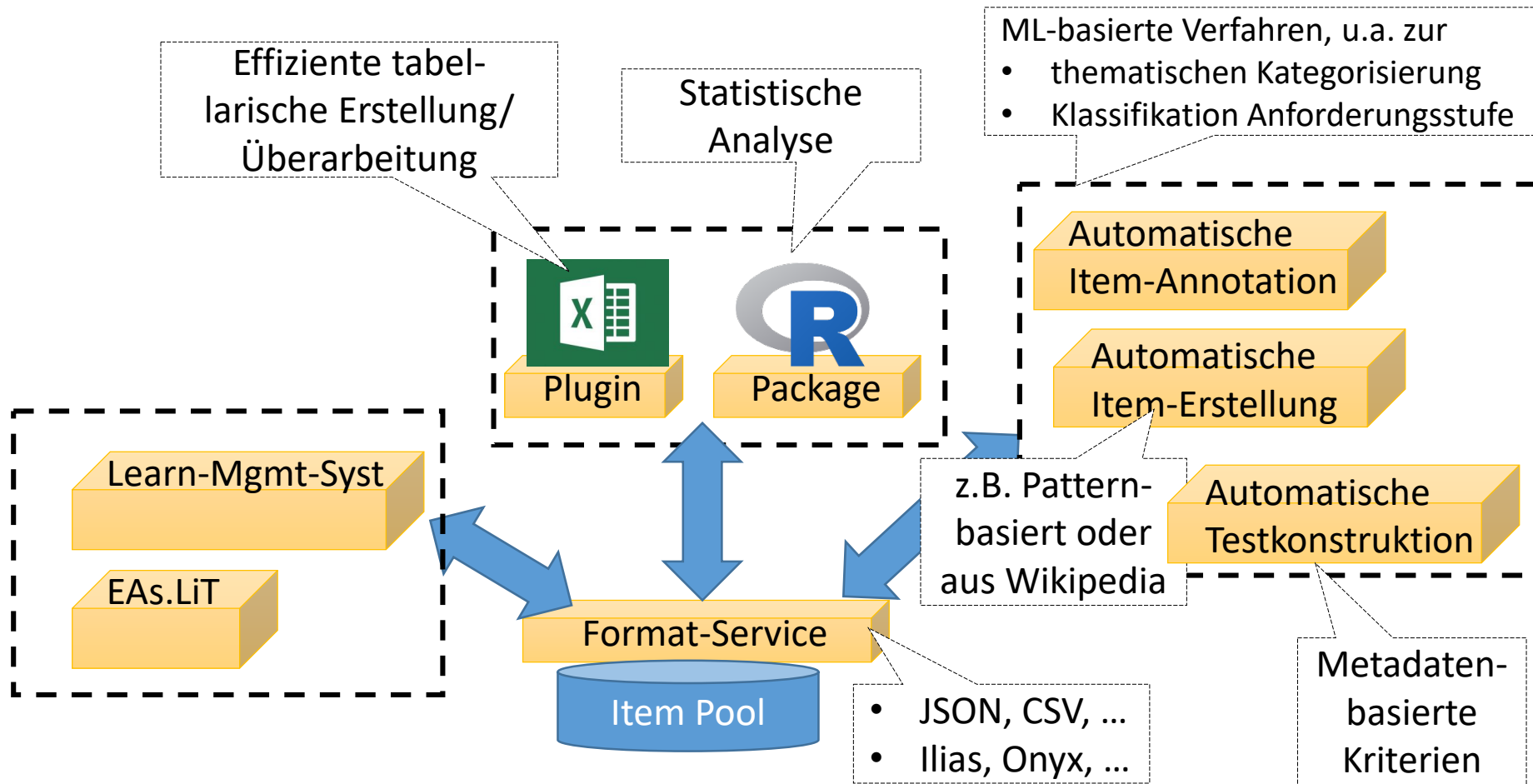
Verarbeitung mit Excel: Beispiel



Vignette	Frage	Antwort 1	Punkte 1	Antwort 2	Punkte 2	Antwort 3	Punkte 3
Sie befinden sich im Jahr 1996.	Wer ist derzeit Bundeskanzler?	Helmut Kohl	1	Roman Herzog	0	Oskar Lafontaine	0
Sie befinden sich im Jahr 1996.	Wer ist derzeit Bundespräsident?	Helmut Kohl	0	Roman Herzog	1	Oskar Lafontaine	0
Sie befinden sich im Jahr 1996.	Wer ist derzeit SPD-Vorsitzender?	Helmut Kohl	0	Roman Herzog	0	Oskar Lafontaine	1

- Effiziente Erstellung ähnlicher Items mittel Copy+Paste
- Materialisierung parametrisierter Items mittels Excel-Formeln
 - „Was ist die Quadratwurzel von X?“

Item-Ökosystem (Micro-Service-Architektur)



Meissner, R.; Thor, A.: **Flexible Educational Software Architecture: at the example of EAs.LiT 2.** Proc. International Workshop on Intelligence Support for Mentoring Processes in Higher Education (IMHE), 2020

There is not just one solution

- Das selbe Problem kann ganz unterschiedlich gelöst werden
- Dimensionen u.a.
 - Technisch vs. organisatorisch
 - Automatisch vs. semi-automatisch vs. manuell
 - ...
- Einfluss auf
 - Art der Lösung (Algorithmus, Framework, ...)
 - Art der Evaluation (Messungen, Nutzerstudie, ...)
 - ...

Skalierbare Kompetenzmessung durch E-Assessment

- **Annotation von Items** mit Metadaten (u.a. Anforderungsstufe) als Grundlage zur Identifikation von **Kompetenzprofilen** durch Datenanalyse / Data Mining
- Infrastruktur zur **effizienten Erstellung und Verwaltung** großer Item Pools

Vielen Dank!