

Möglichkeiten und Herausforderungen der Unterrichtsbeurteilung mit INSULA

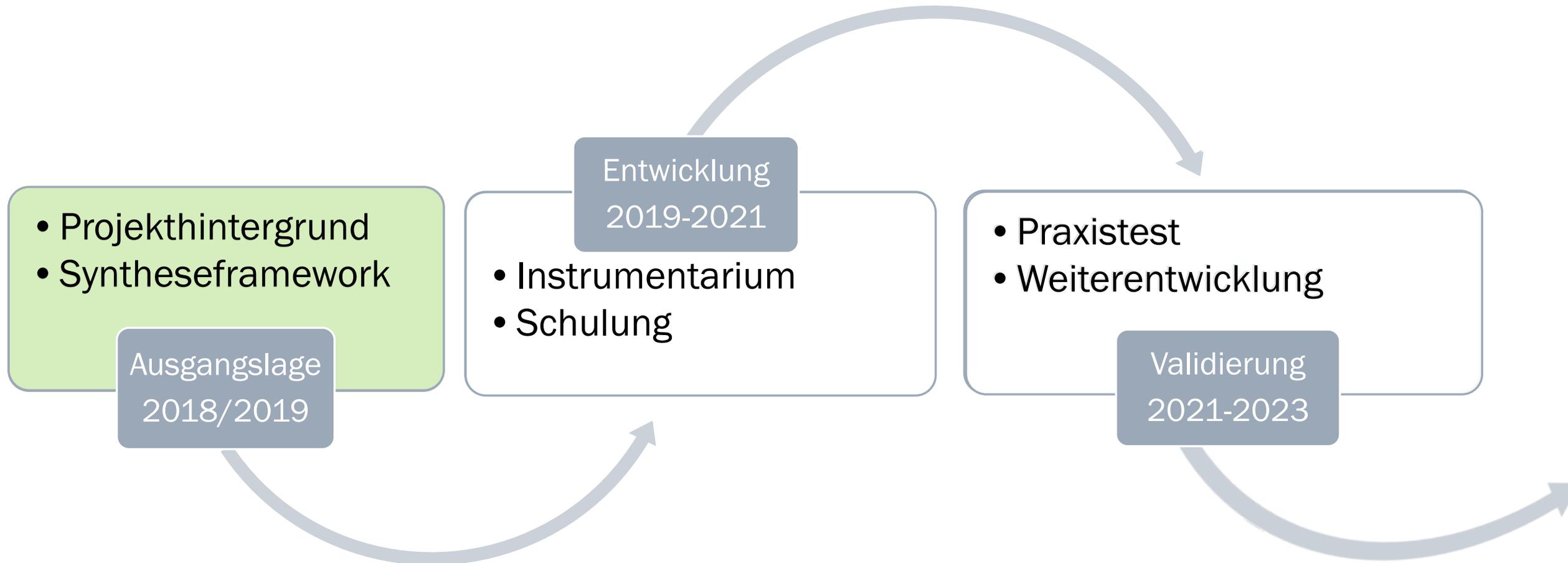
Wida Wemmer-Rogh
Universität Zürich



Team



Übersicht



Projekthintergrund

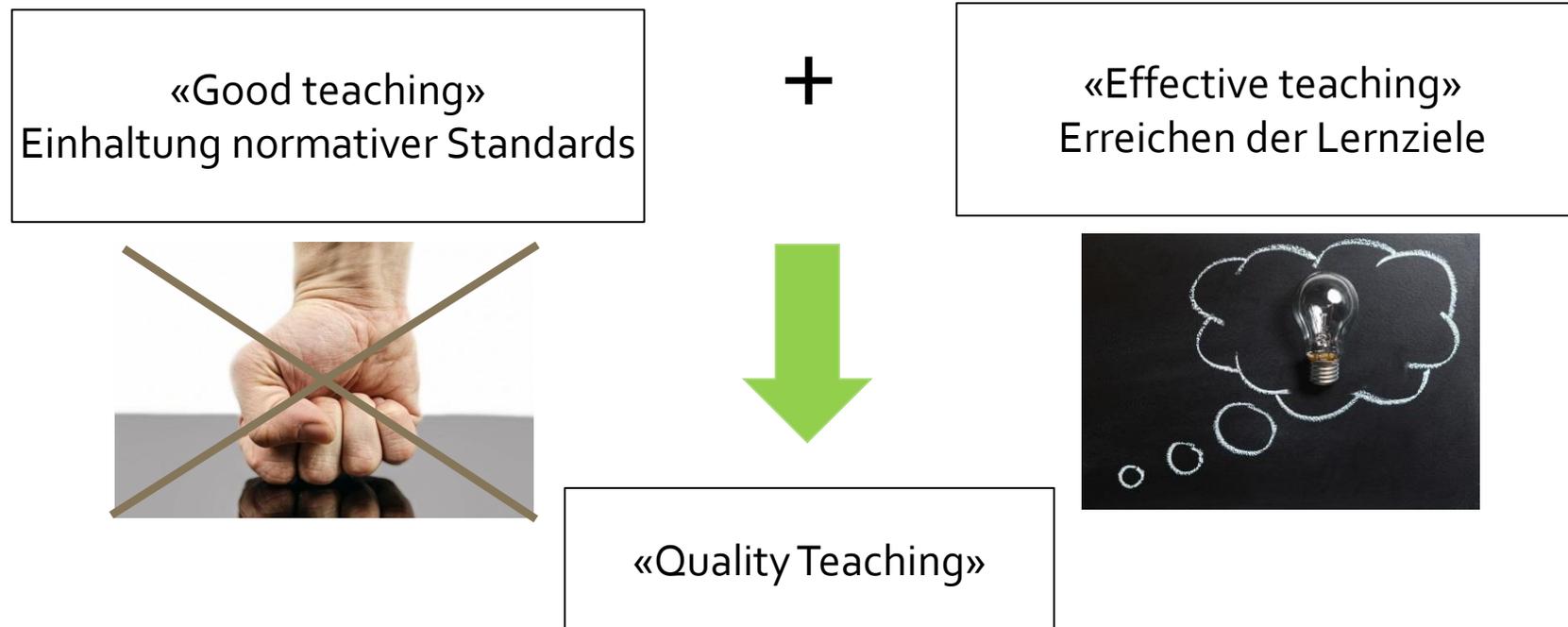
- Arbeitsgemeinschaft Externe Evaluation von Schulen (argev) lancierte das Drei-Jahresprogramm **Qualifizierung in der Unterrichtsbeurteilung** (2018-2020)
- Entwicklung eines **Instrumentariums** und eines **Qualifizierungsangebots** zur Beurteilung von Unterrichtsqualität im Rahmen der externen Schulevaluation
 - Aktueller Forschungsstand
 - Interkantonale Relevanz
 - Lehrplan 21
 - Umsetzbarkeit
- **Zusammenarbeit** mit der Universität Zürich

Projekthintergrund

argev
Interkantonale Arbeitsgemeinschaft
Externe Evaluation von Schulen

Beurteilung von Unterrichtsqualität

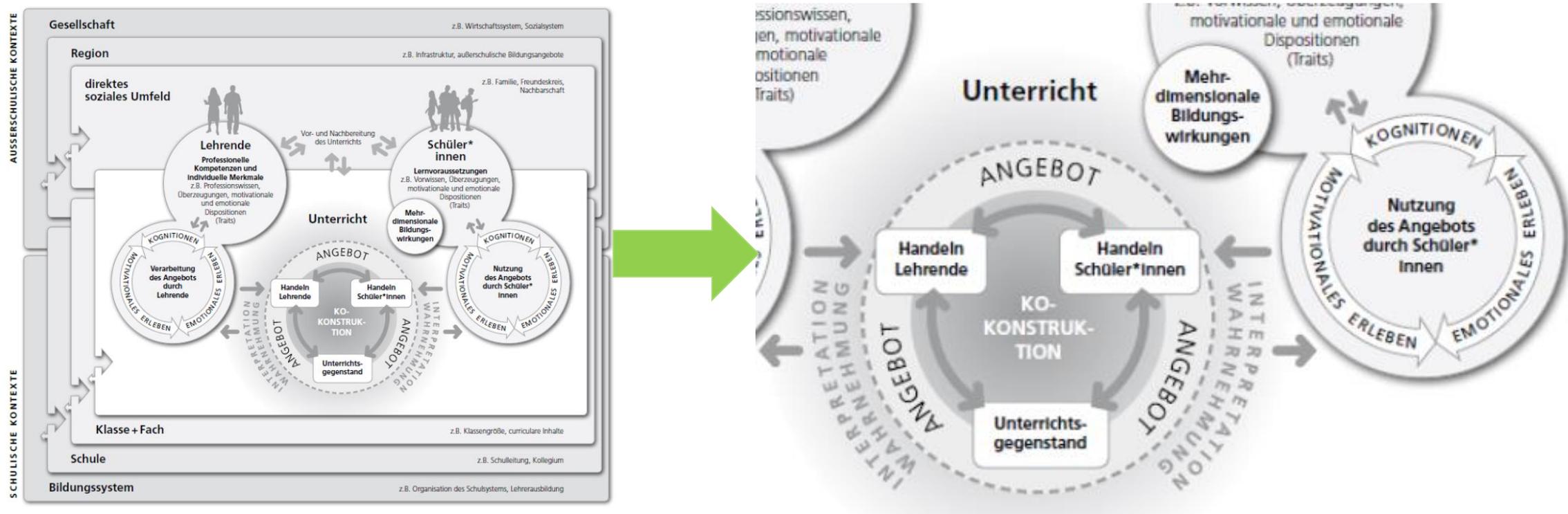
- Definition von Unterrichtsqualität



(Berliner, 2005; Fenstermacher & Richardson, 2005; Helmke, 2015; Klieme, 2019; Kunter & Trautwein, 2013)

Beurteilung von Unterrichtsqualität

- Angebots-Nutzungs-Modelle zur Unterrichtswirksamkeit



(Helmke, 2009; Lipowsky, 2007; Vieluf et al., 2020)

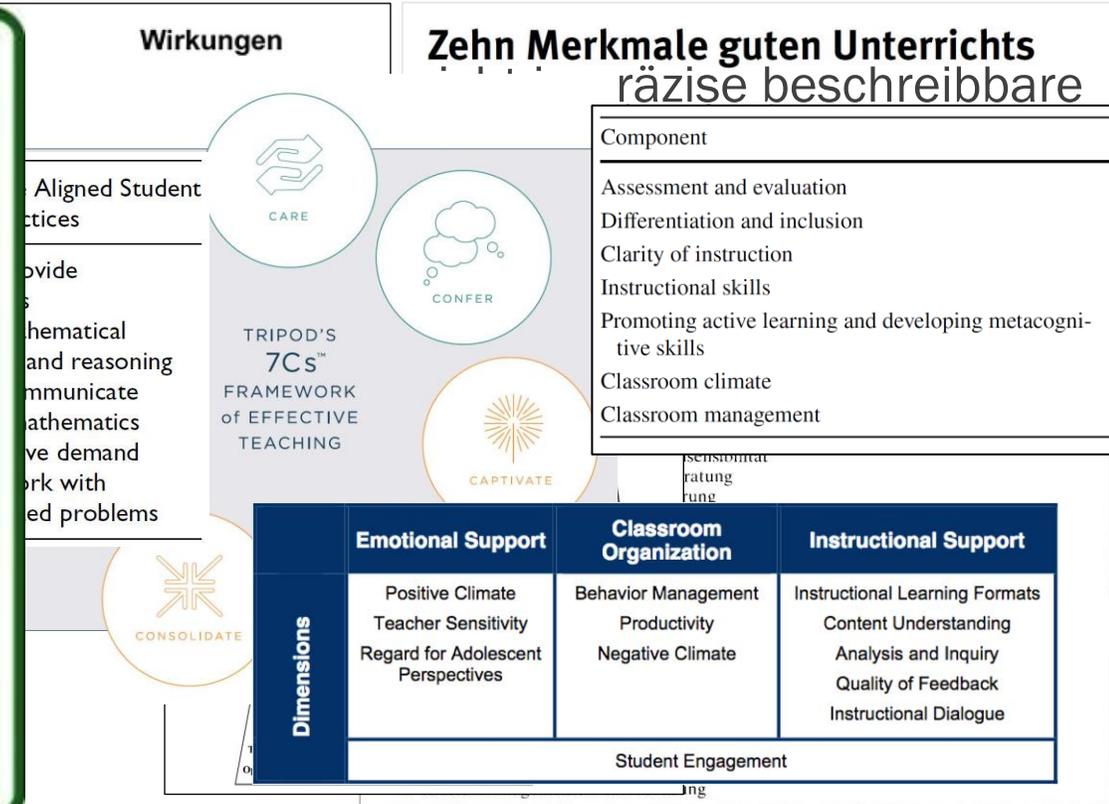
Beurteilung von Unterrichtsqualität

- Mehrperspektivität



Beobachter*innen-Perspektive
«Königsweg» zur Erfassung von Unterrichtsqualität
(Helmke, 2015, S. 292)

Beurteilung von Unterrichtsqualität



(Bell et al., 2019; Helmke, 2011; Klette & Blikstad-Balas, 2018; Klieme et al., 2001; Meyer, 2004; Praetorius & Charalambous, 2018)

Beurteilung von Unterrichtsqualität

- Herausforderungen der quantitativen empirischen Unterrichtsforschung



Heterogenität

- Ansätze/ Zugänge im Feld
- Länge
- Begrifflichkeiten
- Schwerpunktsetzungen
- Fachspezifität

Bezüge zwischen Theorie und Operationalisierung

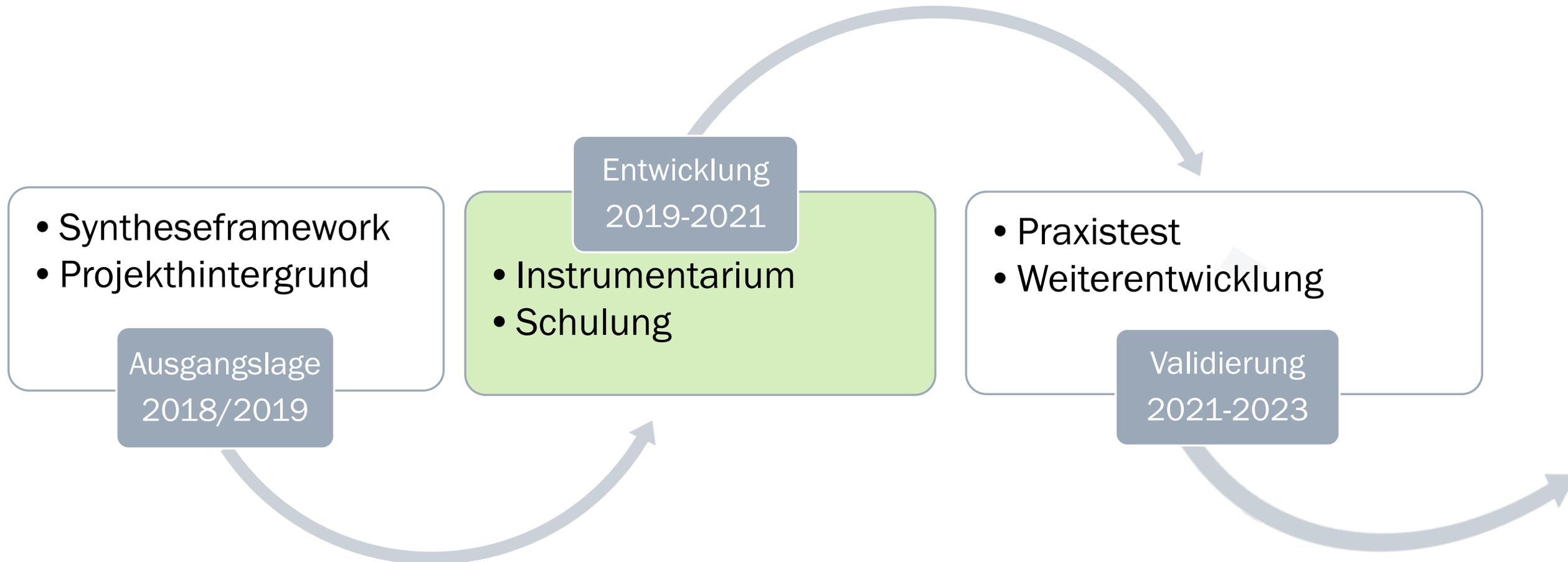
- Angebots- Nutzungsmodell
- Systematik der Merkmale
- Orchestrierung von Unterricht

Messung

- Beobachtung als «Königsweg»?
- Herausforderungen der Reliabilität und Validität
- Effektivität von Beobachtungstrainings
- Mehrperspektivität

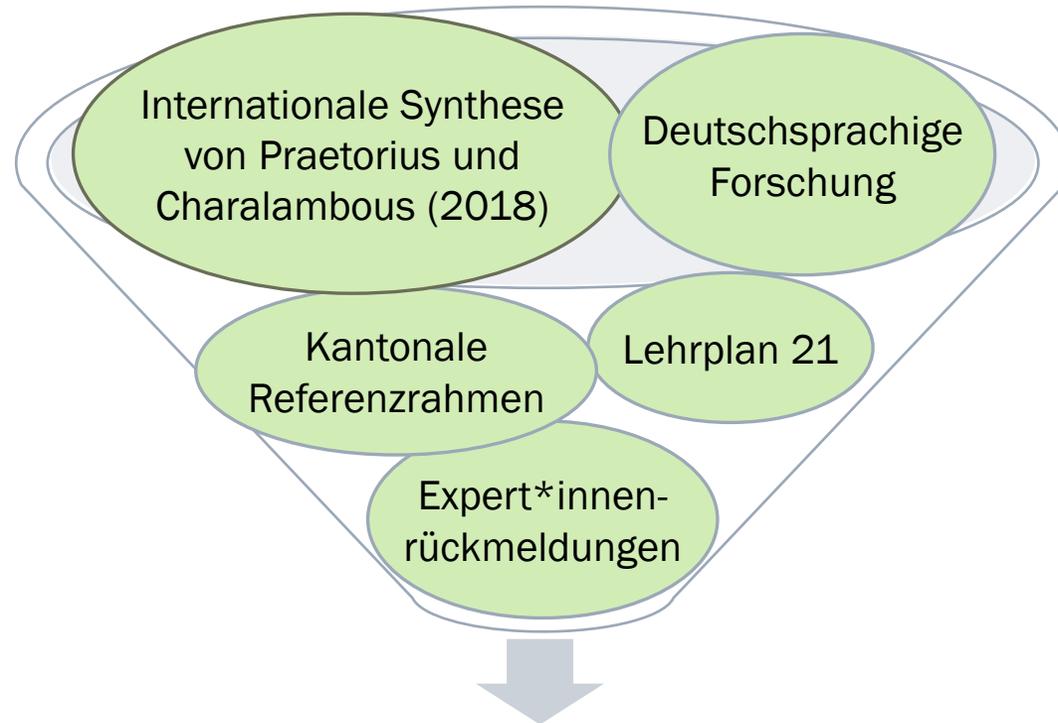
(Bell et al., 2019; Charalambous & Praetorius, 2018, 2020; Helmke, 2022; Praetorius & Charalambous, 2018)

Übersicht



Entwicklung des Instrumentariums

- Synthese als **Ausgangslage** für die Instrumentenentwicklung



Merkmalsliste für INSULA

	Generisch formulierte Merkmalslisten					Generisch und fachspezifisch formulierte Merkmalslisten		Fachspezifisch formulierte Merkmalslisten		
	Meyer	Kiel	Landwehr	& Thiel	direkt.	t al.		Steinweg	Wilhelm	Wüsten et al.
I										
II										
III										
IV										
V										
VI										
VII										

I: Classroom and time ma
V: (Formative) Assesmer

Beurteilbarkeit

Vollständigkeit

Wichtigkeit

student learning

Entwicklung des Instrumentariums

VERSION 1.0

D1: Klassenführung

D2: Motivational-emotionale Unterstützung

D3: Auswahl und Thematisierung des Inhalts

D4: Kognitive Aktivierung

D5: Unterstützung des Konsolidierens von Inhalten

D6: Beurteilung und Feedback

D7: Umgang mit Heterogenität

D1 D1S1: Ein proaktiver und effektiver Umgang mit Disziplinfällen ist gewährleistet.
D2S2: Ein hoher Anteil an Lernzeit ist gewährleistet.

D2 D2S1: Es besteht eine positive Beziehung zwischen Lehrpersonen und den Schüler*innen.
D2S2: Die Schüler*innen pflegen einen respektvollen Umgang miteinander.
D2S3: Den Schüler*innen wird in einem angemessenen Rahmen Autonomie und Verantwortung übertragen.
D2S4: Das Kompetenzerleben der Schüler*innen wird unterstützt.

D3 D3S1: Die behandelten Inhalte sind adäquat ausgewählt.
D3S2: Der Unterricht ist auf das Lernziel abgestimmt.
D3S3: Die Relevanz des Inhalts wird den Schüler*innen verdeutlicht.
D3S4: Die Inhalte werden klar und strukturiert präsentiert.
D3S5: Die Inhalte werden korrekt und präzise dargestellt.

D4 D4S1: Die Aufgaben sind gehaltvoll und auf das kognitive Niveau der Schüler*innen abgestimmt.
D4S2: Die kognitive Aktivität der Schüler*innen wird unterstützt.
D4S3: Aufgaben mit kognitiv aktivierendem Potenzial regen metakognitives Lernen der Schüler*innen an.

D5 D5S1: Übungsaufgaben sind auf die Konsolidierung von Inhalten ausgerichtet.
D5S2: Die Konsolidierung von Inhalten wird unterstützt.

D6 D6S1: Die Beurteilungen sind klar auf die zu erlernenden Kompetenzen ausgerichtet.
D6S2: Die Kompetenzen der Schüler*innen werden überprüft.
D6S3: Die Schüler*innen erhalten qualitativ hochwertiges Feedback.

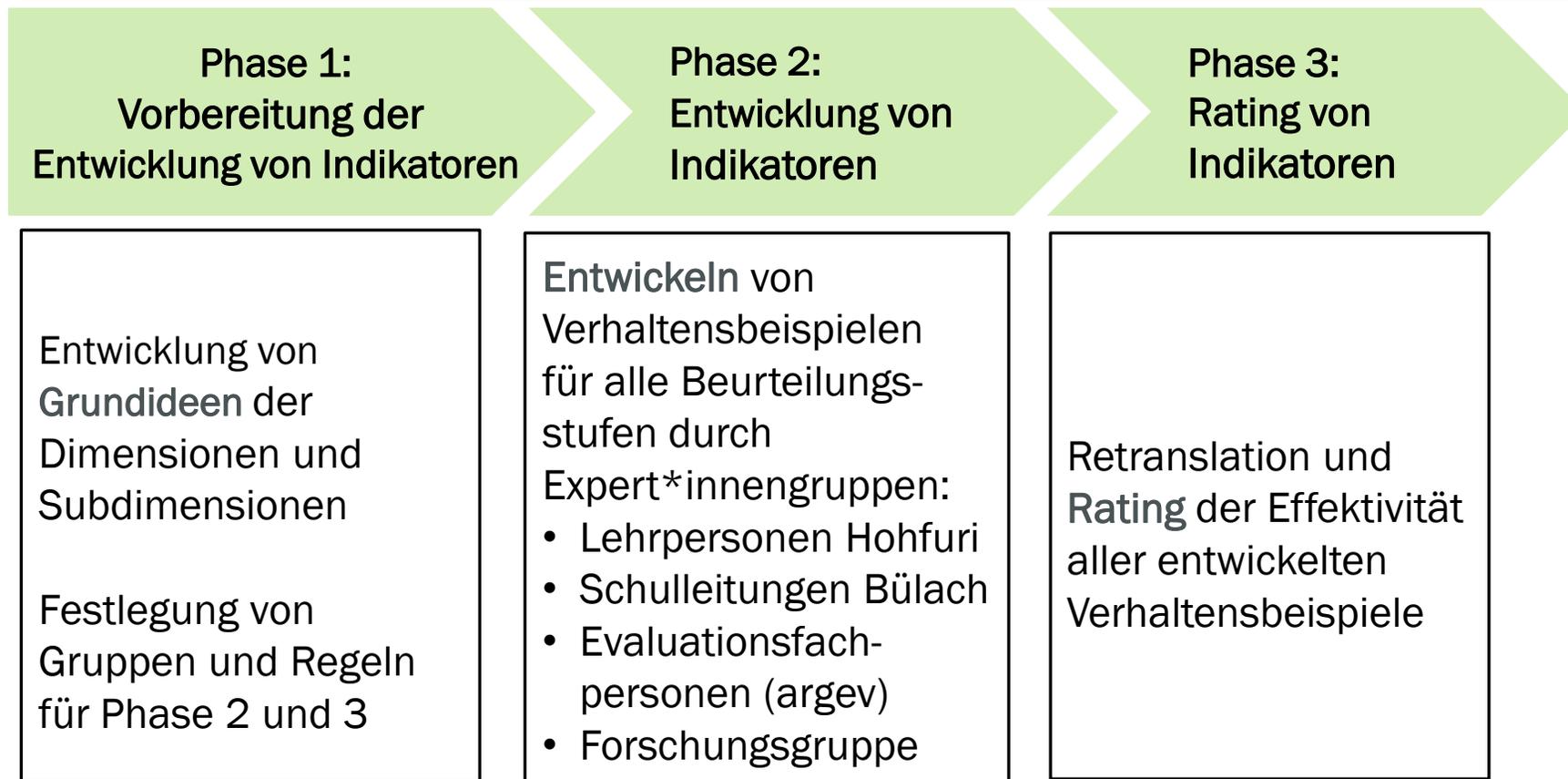
D7 D7S1: Es gibt eine angemessene und differenzierte Passung der Anforderungen an Schüler*innen.
D7S2: Es gibt eine angemessene lernprozessbezogene Begleitung der Schüler*innen.

Entwicklung des Instrumentariums

1	2	3	4
Wenig ausgeprägt	Mässig ausgeprägt	Überwiegend ausgeprägt	Umfassend ausgeprägt
Ungenügende Unterrichts- praxis	Funktionsfähige Unterrichts- praxis	Gute Unterrichts- praxis	Herausragende Unterrichts- praxis

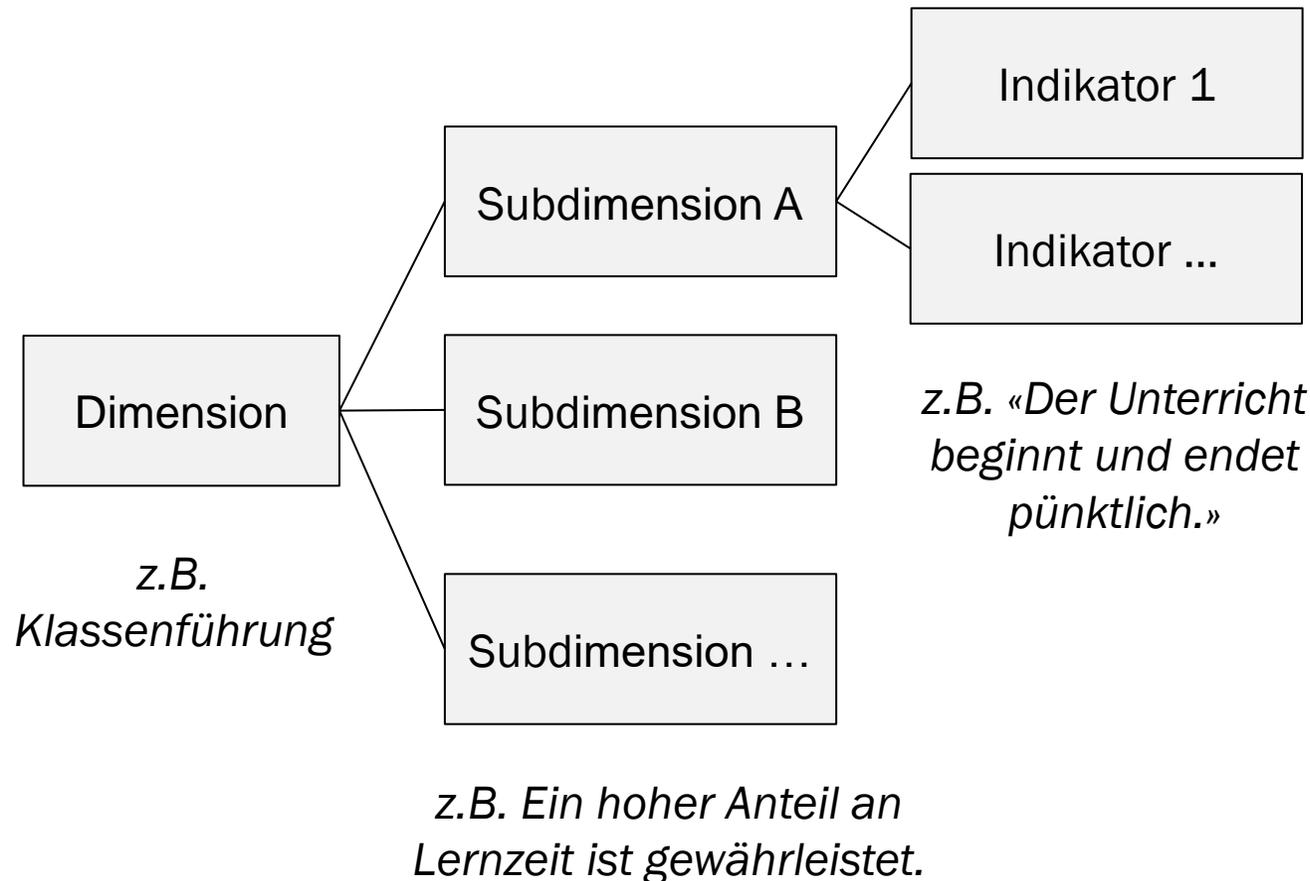
(Landwehr & Steiner, 2007)

Entwicklung des Instrumentariums



(Debnath, Lee & Tandon, 2015; Flanagan, 1954; Klieger et al., 2018; Ohland et al., 2012)

Entwicklung des Instrumentariums



Entwicklung des Instrumentariums

KF: Klassenführung

Grundidee:

Gute Klassenführung (Helmke, 2014), unter anderem durch die Bereitstellung von Raum und Ressourcen.

Achtung:

- Bei der Beurteilung von Klassenführung sind die Probleme im Unterricht zu berücksichtigen.

KF2: Ein hoher Anteil an Lernzeit ist gewährleistet.

Grundidee: Als echte Lernzeit wird die Zeit bezeichnet, die für den Unterricht zur Verfügung steht. Erfolgreiches unterrichtsbezogenes Zeitmanagement umfasst: keine unnötigen Wartezeiten oder keine Verweilzeiten, keine unnötigen Unterbrechungen, keine unnötigen Wiederholungen, keine unnötigen Wiederholungen, keine unnötigen Wiederholungen. Bezug auf zentrale Lernziele tragen ebenfalls zur Erreichung von Lernzeit im Unterrichtsverlauf seitens der Lehrpersonen.

Achtung:

- Entlastungsangebote für Schüler*innen mit besonderen Bedürfnissen sind zu berücksichtigen.

KF2: Ein hoher Anteil an Lernzeit ist gewährleistet.

Beispielindikatoren für Subdimensionen	
1	<ul style="list-style-type: none"> • Häufige oder längere Abschweifungen der Lehrpersonen von geplanten Inhalten zeugen von einer fehlenden Fokussierung und Zielorientierung. • Es gibt viele Wartezeiten für die Schüler*innen (z. B. Warteschlangen am Pult, beim Aufstrecken am Platz, fehlendes Material). • Ein grosser Anteil der Unterrichtszeit wird mit administrativen Inhalten verbracht (z. B. Organisatorisches, irrelevante Ankündigungen). • Es geht viel Zeit verloren durch vorbereitende und nicht lernwirksame Aufgaben (z. B. Schüler*innen üben eigentlich Leseverstehen, schneiden aber nur Kärtchen aus oder unterhalten sich).
2	<ul style="list-style-type: none"> • Übergänge zwischen einzelnen Aktivitäten (z. B. Wechsel von Gruppenarbeit zum Klassengespräch) nehmen relativ viel Zeit ein. • Der Unterrichtsfluss wird durch Nachfragen einiger Schüler*innen unterbrochen, weil Aufträge nicht klar oder eindeutig formuliert sind. • Der Unterricht beginnt und endet pünktlich. • Ein Zeitmanagement ist grundsätzlich erkennbar, auch wenn nicht explizit visualisiert oder kommuniziert.
3	<ul style="list-style-type: none"> • Das geplante Zeitmanagement der Lehrpersonen ist visualisiert oder kommuniziert. • Übergänge zwischen den Lernsequenzen funktionieren überwiegend reibungslos. • Wartezeiten sind weitgehend vermieden (z. B. mit Möglichkeit zur Selbstkorrektur oder zur selbstständigen Weiterarbeit nach abgeschlossenem Teilauftrag, Aufgaben, Angebote im Freispiel oder in der Auffangzeit sind organisatorisch gut vorbereitet und rechtzeitig bereitgestellt).
4	<ul style="list-style-type: none"> • Der Unterricht verläuft reibungslos. • Die Schüler*innen arbeiten konzentriert an den Inhalten; es herrscht durchweg eine konzentrierte Arbeitsatmosphäre im Klassenzimmer. • Die Schüler*innen sind in der Lage, vorhandene Zeitressourcen selbstständig inhaltsbezogen sinnvoll zu nutzen.

Beurteilungsstufen				
Wenig ausgeprägt 1	Mässig ausgeprägt 2	Überwiegend ausgeprägt 3	Umfassend ausgeprägt 4	Nicht beurteilbar
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

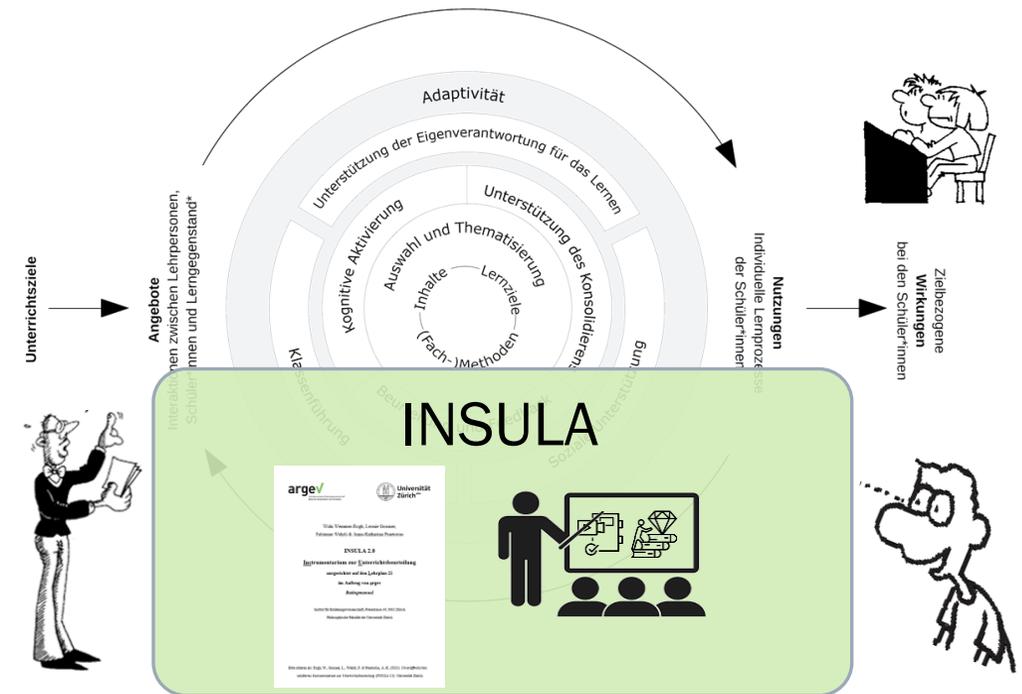
ch
er-
eit
in
eln
ng

Entwicklung des Instrumentariums

- Bezüge zwischen Theorie und Operationalisierung mithilfe des MAIN-TEACH-Modells

Begegnung der Herausforderungen Teil 2

- Angebots-Nutzungs-Modelle
 - Mehrperspektivität
 - Fragebogen zur Nutzung
- Umfassendes Verständnis von UQ
 - Orchestrierung von Unterricht
 - Systematik der Merkmale
 - Fachexpertise
 - Fragebogen zur Planung



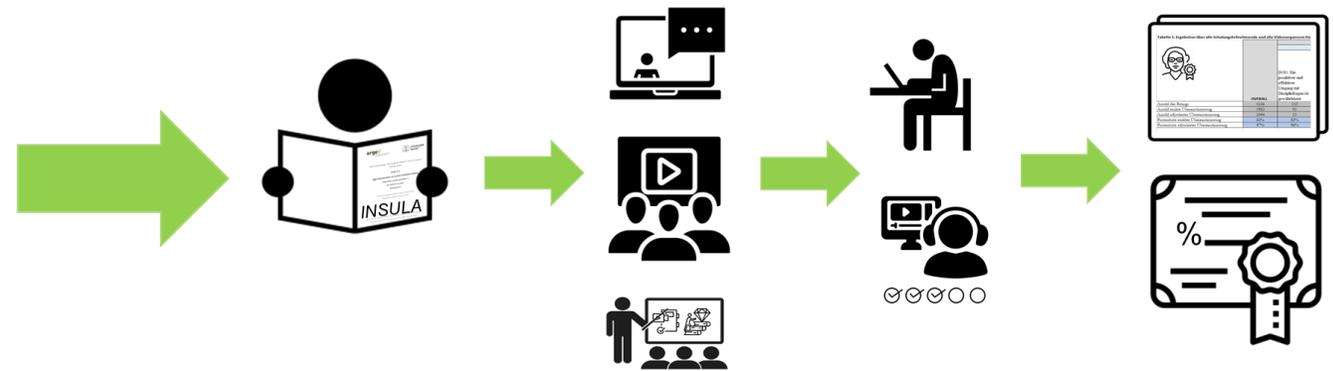
(Charalambous & Praetorius, 2020; Helmke, 2022; Hill et al., 2012; Praetorius et al., 2023)

Entwicklung der Schulung

- Gewährleistung der Reliabilität und Validität der **Messung** mithilfe von Schulungen

Begegnung der Herausforderungen Teil 3

- Beobachtung als «Königsweg»?
- Observation Systems
- Schulung mit Videomaterial
- Passung an Nutzungskontext



(Bell et al., 2019; Liu et al., 2019; OECD, 2020)

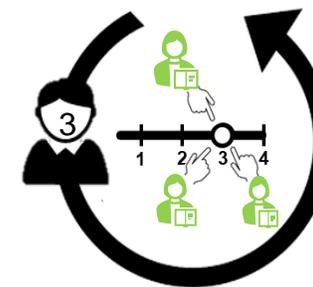
Entwicklung der Schulung

- Zertifizierung als Qualifizierungsmaßnahme

“...used to assess whether a **rater has mastered** the content of the training activities. [...] the tests often require the rater to carry out the same tasks they will be doing when scoring and **compare the ratings** the raters assign **with master raters' ratings**. Frequently, raters cannot begin their work until they pass certification.” (Liu et al., 2019, S. 66)

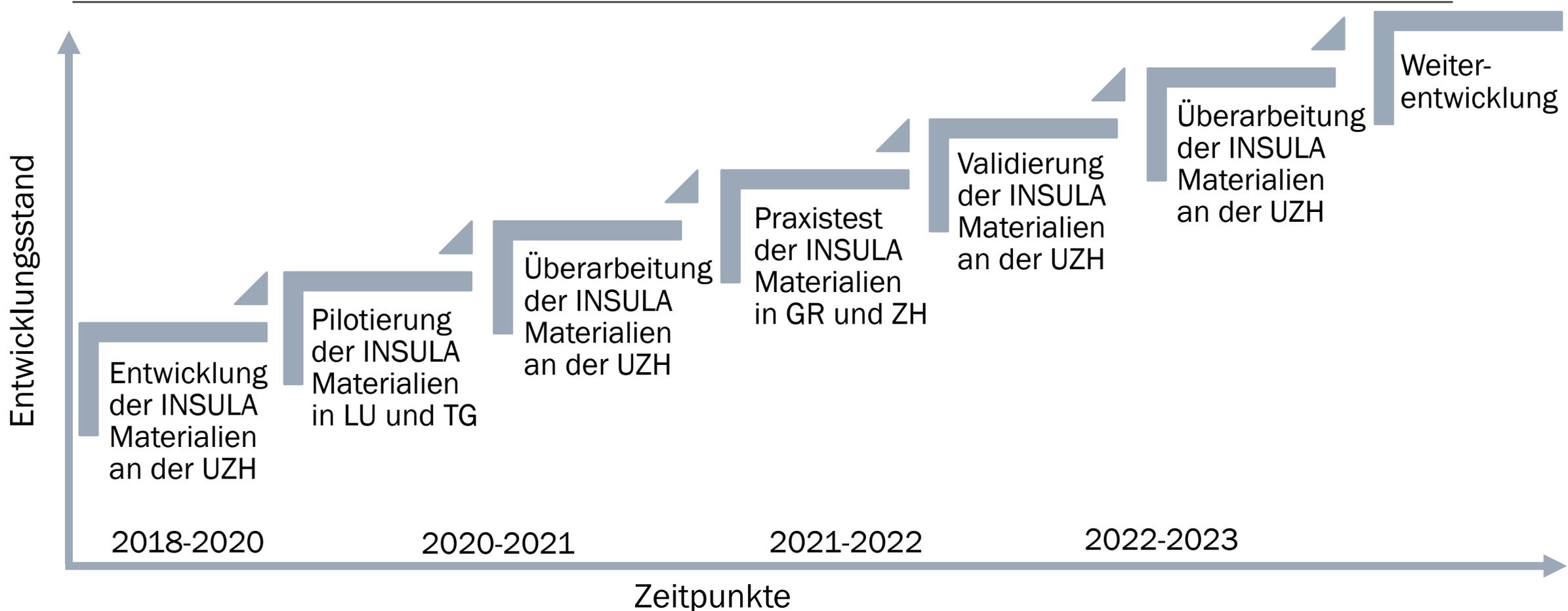


Inter-Rater-
Übereinstimmung

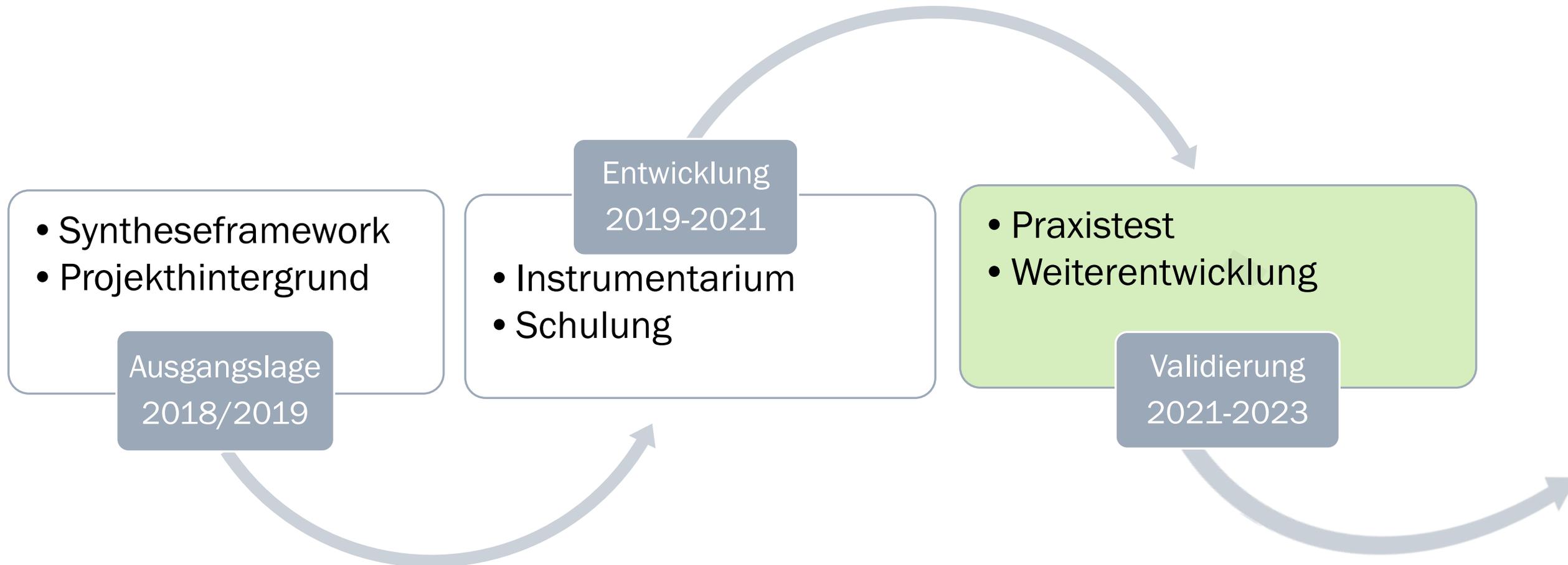


Übereinstimmung
mit Masterratings

Entwicklung von INSULA



Übersicht



Validierung von INSULA

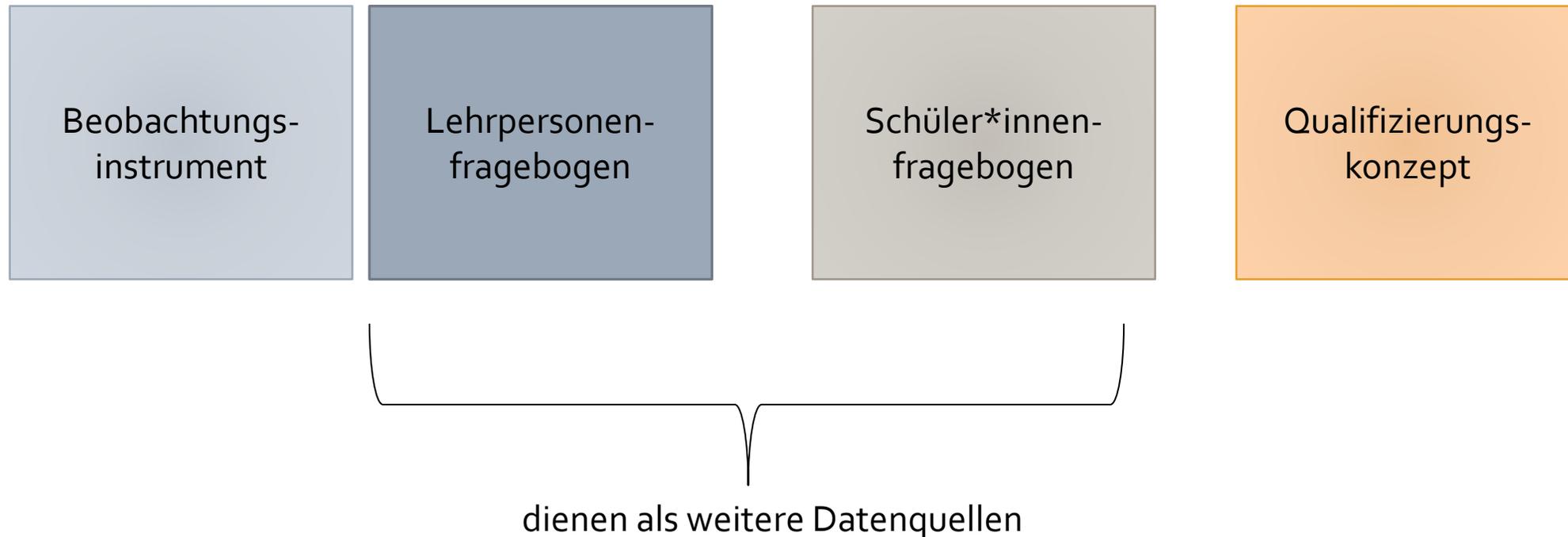
- Prozess einer **argumentationsbasierten** Validierung für den Nutzungskontext



(Hartig, Frey & Jude, 2020; Herbert & Schweig, 2021)

Validierung von INSULA

- Bestandteile



Validierung von INSULA



Datengrundlage: Beobachtungsdaten

- $N = 1939$ Unterrichtsbeobachtungen, 55 Rater*innen

Kanton	Zürich	Graubünden
Schuljahr	2021/ 22	2021/ 22
Evaluationsfachpersonen	41	14
Schulen	108	18
Unterrichtsbeobachtungen	1724	215
Doppelbeobachtungen	44	0

Datengrundlage: Qualitative Rückmeldungen

- Schuljahr 2021/2022
- alle geschulten Fachstellen
- $N = 112$ Anmerkungen
- Zweistündige Gruppendiskussion an der UZH
- $N = 9$ Personen aus Zürich und Graubünden
- Themen:
 - Eignung für den Schulevaluationskontext
 - Durchführung der Beurteilung
 - Modularisierung
 - Lehrpersonenfragebogen
 - Schüler*innenfragebogen
 - Sonstiges

Liebe Evaluatorinnen und Evaluatoren
Auf diesem Webformular können Rückmeldungen zum argev Instrumentarium INSULA eingegeben werden, sei dies zu den Dimensionen und Subdimensionen oder zum Lehrpersonen- und Schülerfragebogen. Herzlichen Dank für die Mitarbeit!

[Alle ausklappen](#)

- ▶ Rückmeldung zum Ratingmanual
- ▶ D1: Klassenführung
- ▶ D2: Motivational-emotionale Unterstützung
- ▶ D3: Auswahl und Thematisierung des Inhalts
- ▶ D4: Kognitive Aktivierung
- ▶ D5: Unterstützung des Konsolidierens von Inhalten
- ▶ D6: Beurteilung und Feedback
- ▶ D7: Umgang mit Heterogenität
- ▶ Sonstige Rückmeldungen
- ▶ Rückmeldung zum Excel-Tool, zur Modularisierung, zu den Fragebögen oder zur Schulung



Validierung von INSULA



Auswertung der Daten: Reliabilität

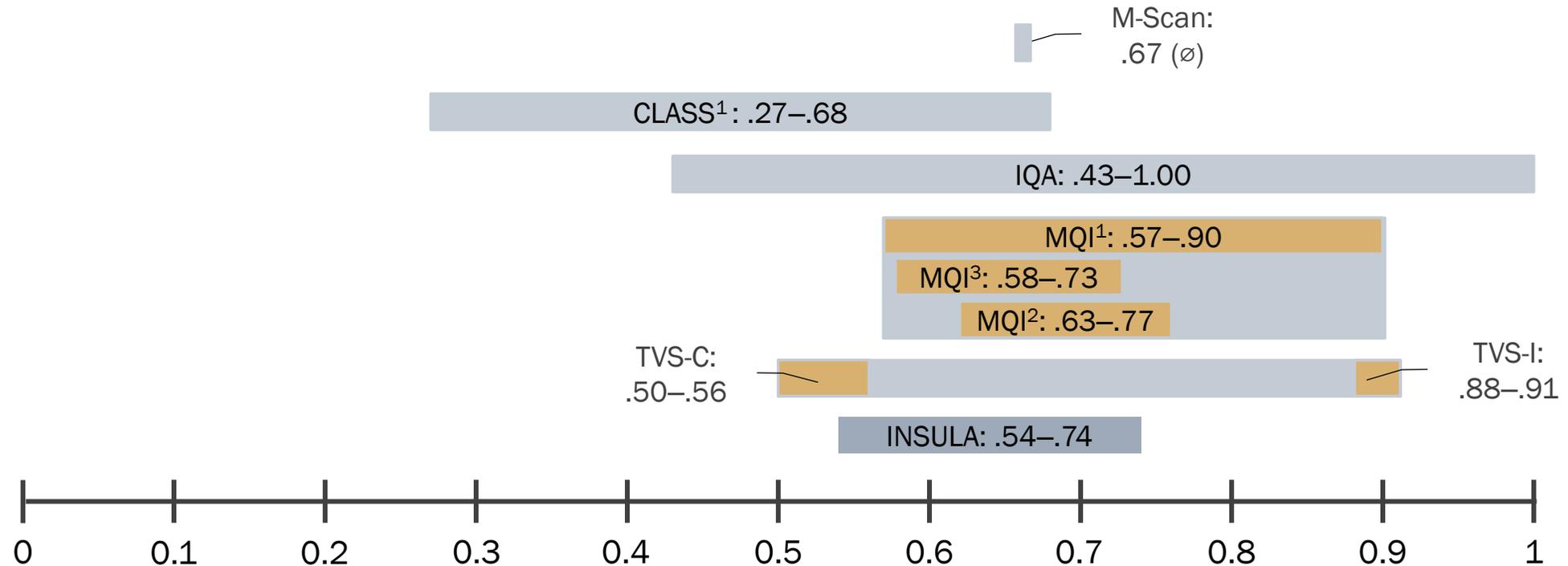
- Kombination von verschiedenen anderen Kennwerten (z. B. prozentuale Werte, ICC, rwg)
- Betrachtung von Dimensionen und Subdimensionen

Dimension	Raterübereinstimmung			
	% exakt	ICC (1.2)	r_{wg}	$r^*_{wg(j)}$
Klassenführung (D1)	61.63	0.81	0.88	0.82
Motivational-emotionale Unterstützung (D2)	64.67	0.84	0.93	0.86
Auswahl und Thematisierung des Inhalts (D3)	71.56	0.90	0.96	0.87
Kognitive Aktivierung (D4)	67.42	0.84	0.92	0.82
Unterstützung des Konsolidierens von Inhalten (D5)	74.19	0.87	0.94	0.89
Beurteilung und Feedback (D6)	53.85	0.87	0.92	0.85
Umgang mit Heterogenität (D7)	66.65	0.88	0.92	0.87

Interpretation	
.51 – .70	moderat
.71 – .90	stark
.91 – 1.00	sehr stark

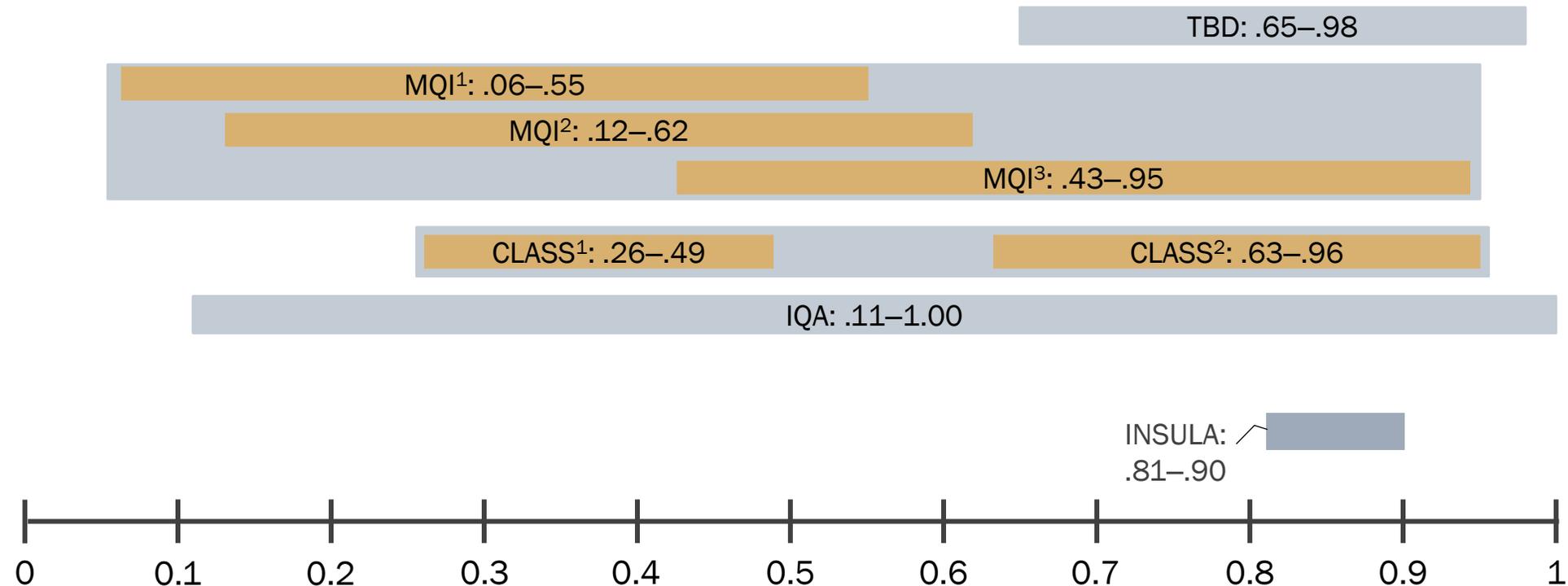
(CECR, 2012 ; LeBreton & Senter, 2008; Wirtz & Casper, 2002)

Forschungsstand: Reliabilität (% exakt)



CLASS¹: Pianta et al. (2012) mit MET-Daten; IQA: Resnick et al. (2006); M-Scan: Walkowiak et al. (2018); MQI¹: Bell et al. (2015) mit UTQ-Daten; MQI²: Bell et al. (2015) mit MET-Daten; MQI³: Mantzicopoulos et al. (2019); TVS-C: OECD (2020) für Komponenten; TVS-I: OECD (2020) für Indikatoren; TRU: Zhou et al. (2023), *adaptiert

Forschungsstand: Reliabilität (ICC 1.2)



CLASS¹: Pianta et al. (2012) mit MET-Daten; CLASS²: Pakarinen et al. (2010); IQA: Resnick et al. (2006); MQI¹: Bell et al. (2015) mit UTQ-Daten; MQI²: Bell et al. (2015) mit MET-Daten; MQI³: Mantzicopoulos et al. (2019); TBD: Praetorius et al. (2018); TEDS-Instruct: Jentsch et al. (2020)

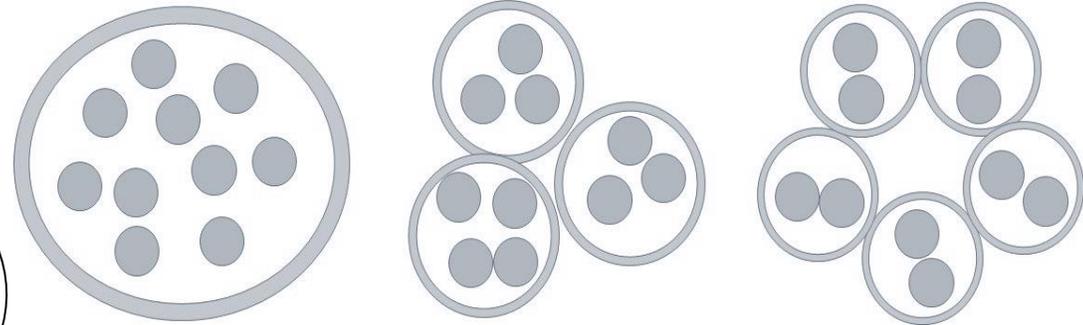
Validierung von INSULA



Auswertung der Daten: Qualitätsdimensionen

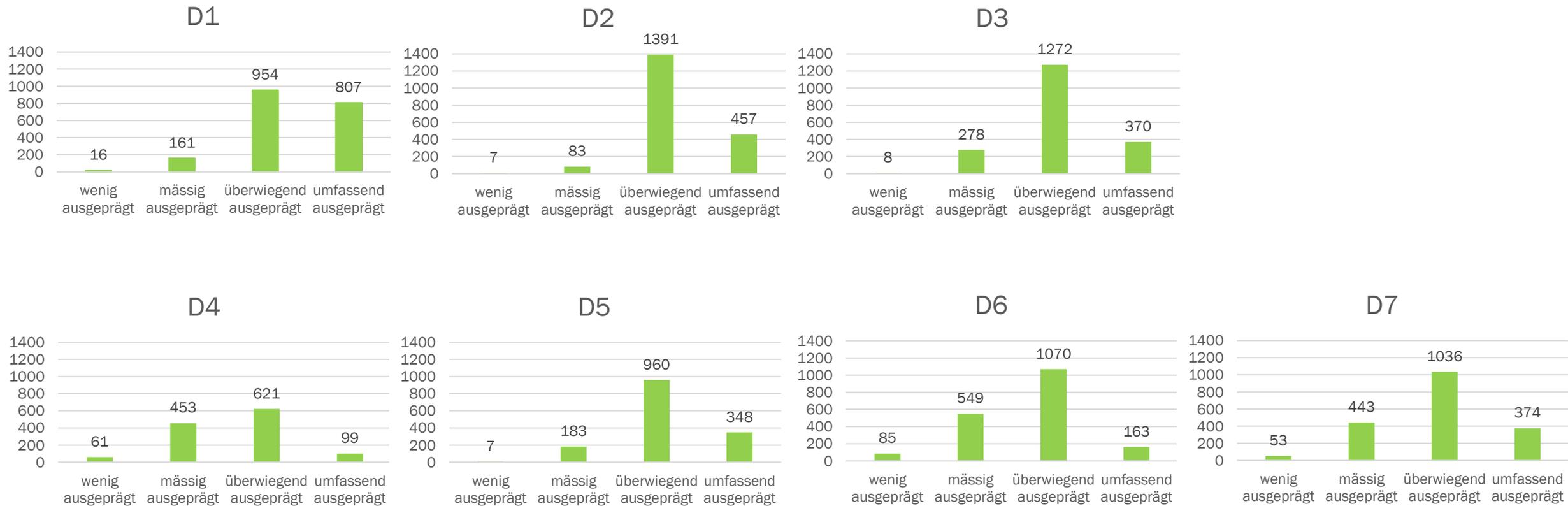
- Jede Dimensionierung von Unterrichtsqualität ist eine Setzung.

«Jede Klassifikation von Merkmalen der Unterrichtsqualität ... ist eine individuelle Konstruktion; ... Mit guter Begründung könnte man auch für eine geringere Anzahl von Qualitätsbereichen plädieren ... oder für eine höhere Anzahl.» (Helmke, 2015, S. 169)



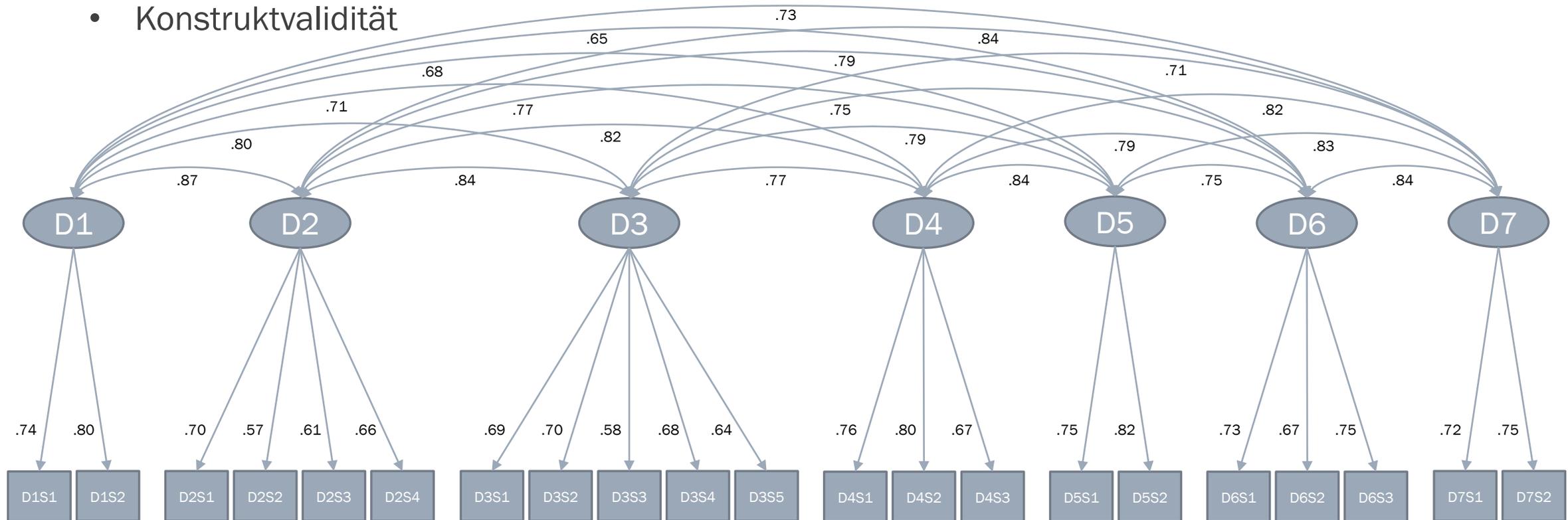
Auswertung der Daten: Qualitätsdimensionen

- Beobachtbarkeit



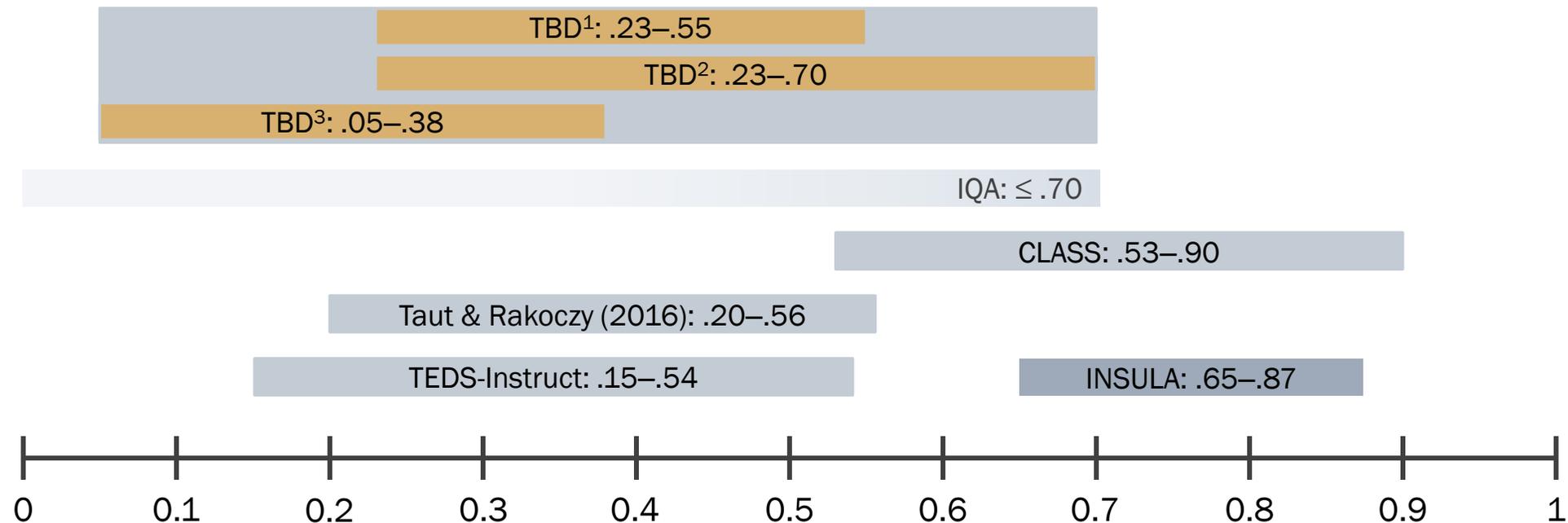
Auswertung der Daten: Qualitätsdimensionen

- Konstruktvalidität



Modellfit: $\chi^2 = 858.944$, $df = 168$, $p = .000$, $\chi^2/df = 5.113$, $CFI = 0.941$, $TLI = 0.926$, $RMSEA = 0.049$, $p(RMSEA < .05) = 0.605$, $SRMR = 0.036$, $AIC = 53303.926$, $BIC = 53759.906$, $SSABIC = 53493.049$

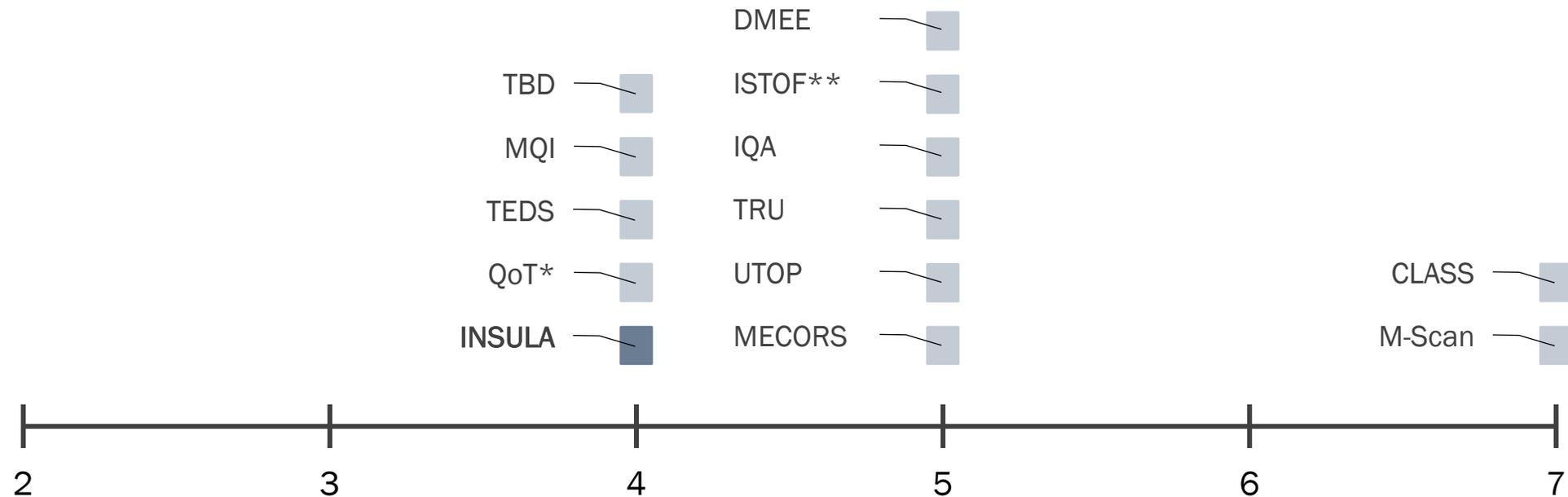
Forschungsstand: Interkorrelationen



CLASS: Pianta et al. (2012) & Li et al. (2020); IQA: Resnick et al. (2006); TBD¹: Fauth et al. (2014a); TBD²: Fauth et al. (2014b); TBD³: Künsting et al. (2016); TEDS-Instruct: Jentsch et al. (2020)

Forschungsstand: Antwortstufen

- Rückblick auf argev-Entscheidung



Praetorius & Charalambous (2018)

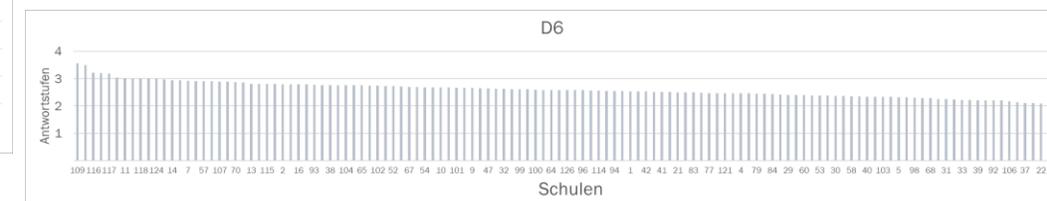
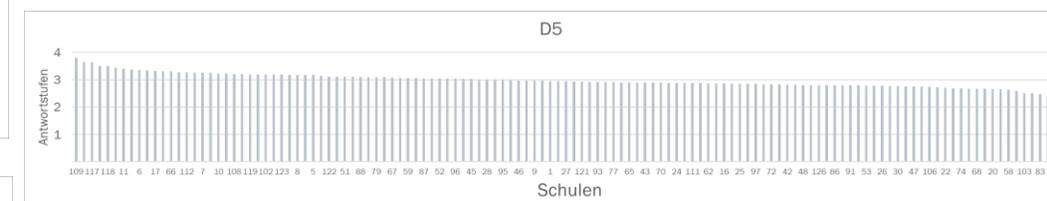
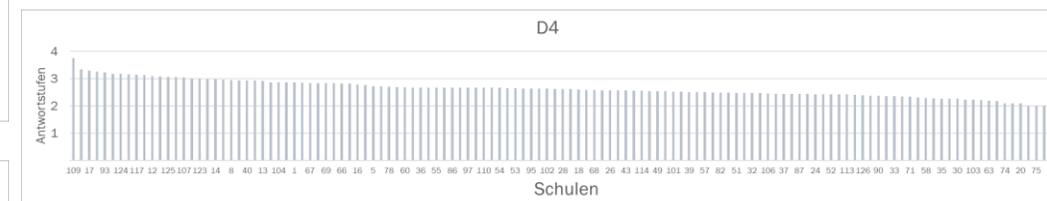
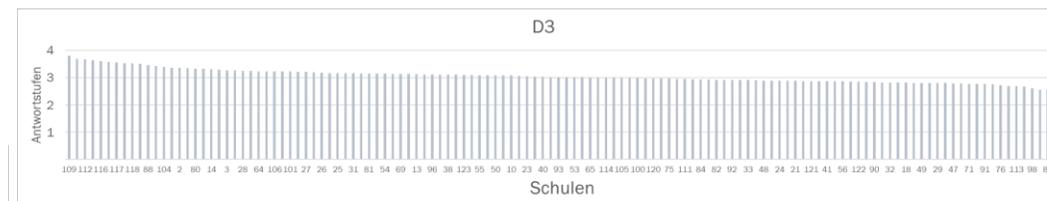
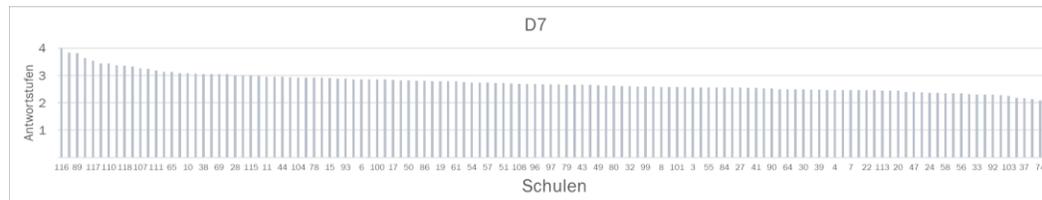
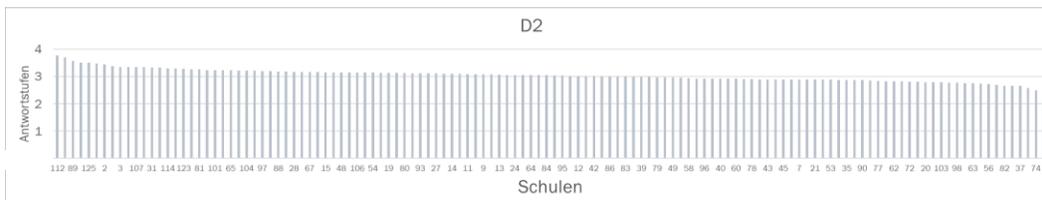
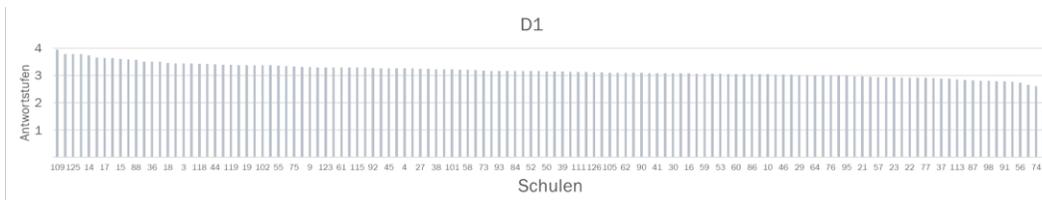
* zusätzlich: Not Present (NP) & Present (P); ** zusätzlich: Not applicable (NA) & Not observed (NO)

Validierung von INSULA



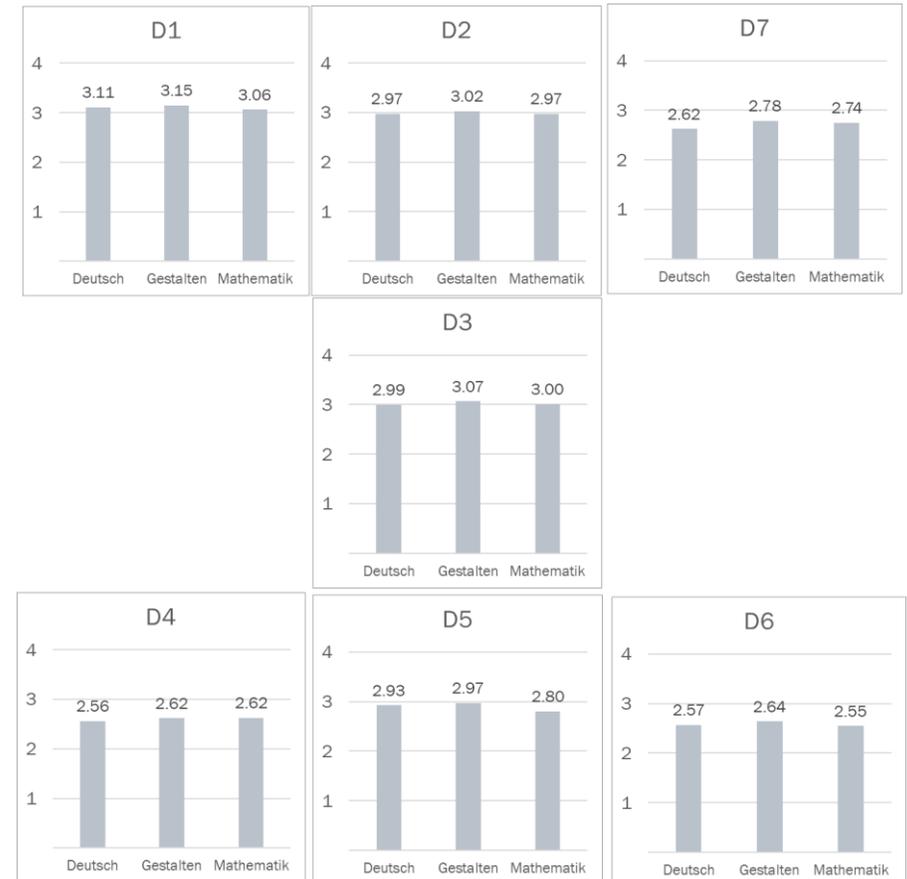
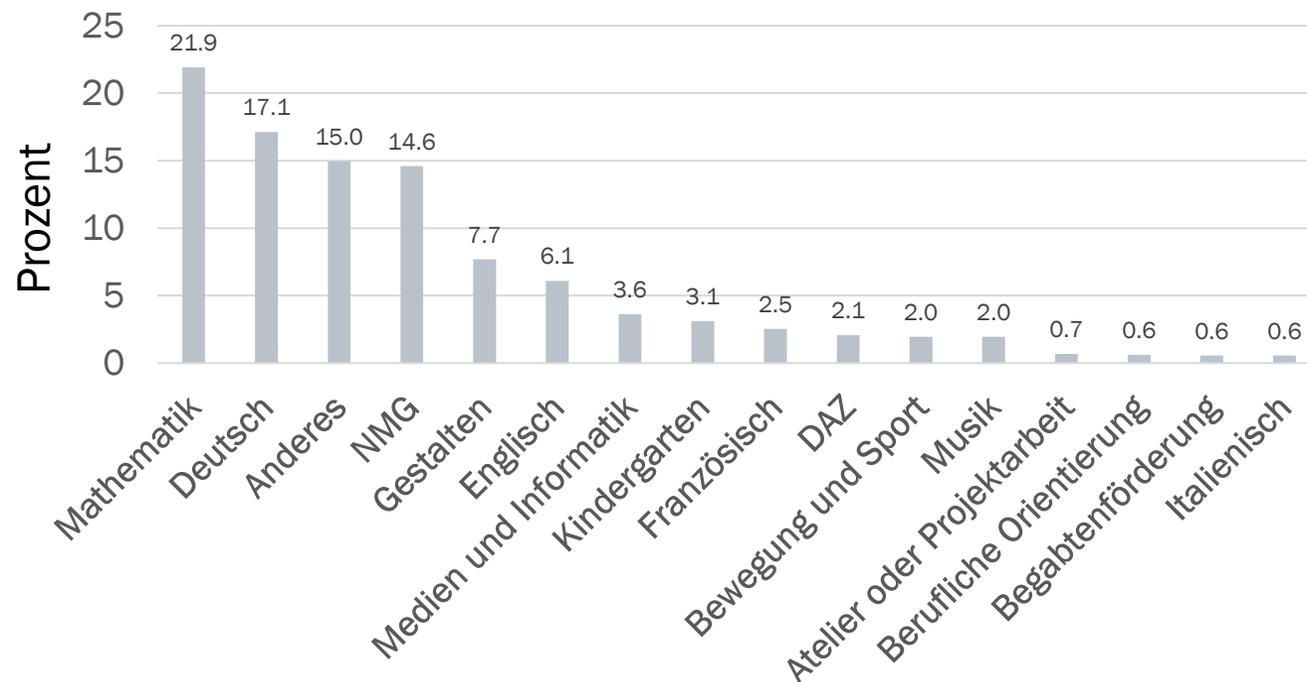
Auswertung der Daten: Variation

- Verteilung auf Schulebene



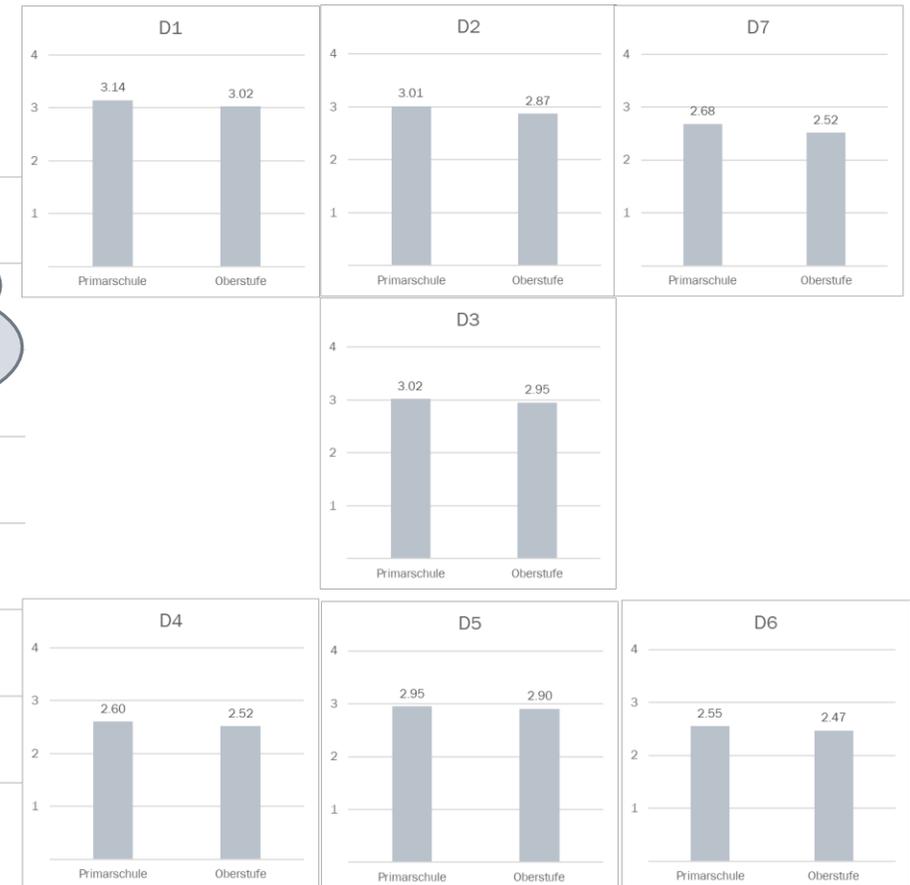
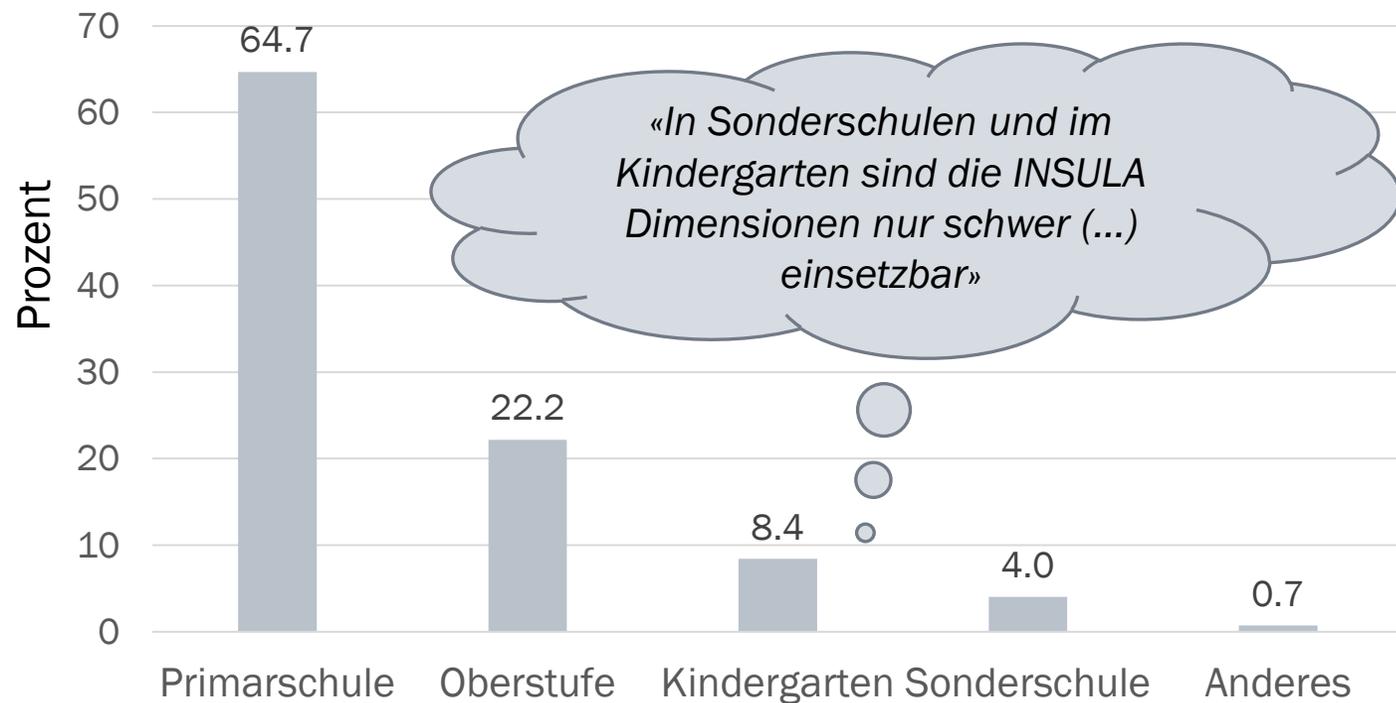
Auswertung der Daten: Variation

- Verteilung der Fächer



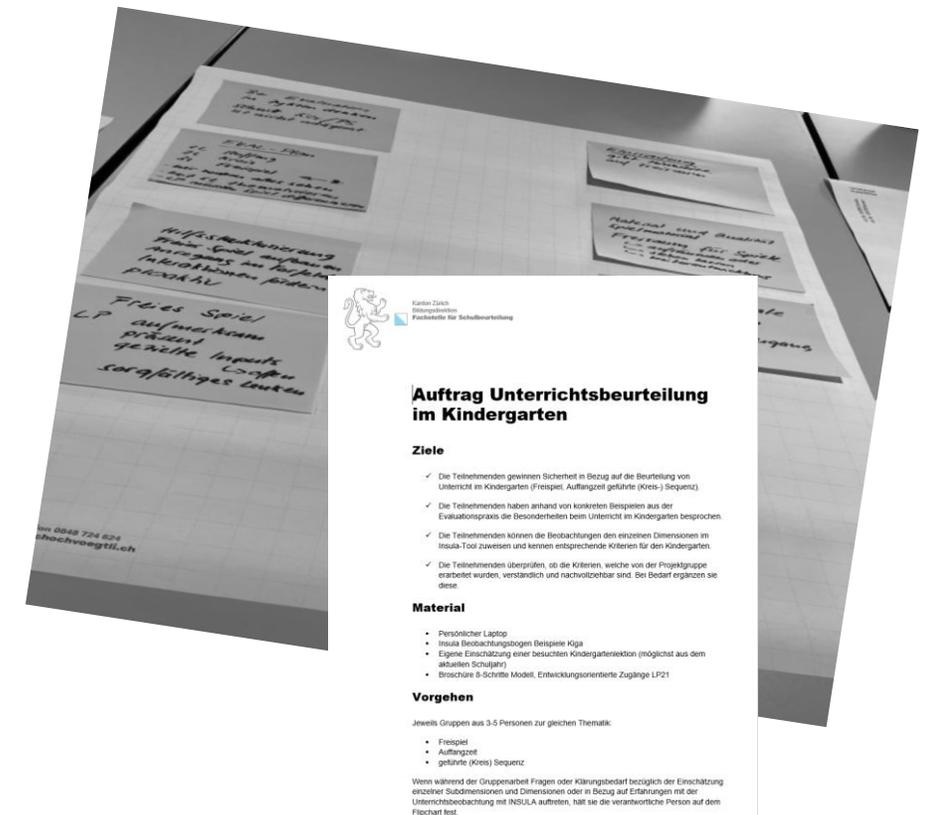
Auswertung der Daten: Variation

- Verteilung der Schulstufen



Exkurs: Fokusgruppen

- Plenumstage der FSB Zürich
- Angeleitete Beschäftigung mit Anwendung von INSULA in
 - Kindergärten
 - Inklusiven Settings
 - Modernen Settings
- Erarbeitung von Ergänzungen im Ratingmanual



Validierung von INSULA



Evaluation der Schulung: Schulungsdurchführung

- Schulungsdurchführung
 - 7 Schulungen
 - 03/2021 – 10/2022
 - 16 Fachstellen für Schulevaluation
- Schulungsteilnehmende
 - 127 Schulungsteilnehmende mit Zertifizierung
 - 59,8 % weiblich
 - Unterrichtserfahrung: $M = 15.28$ Jahre ($SD = 9.01$; $Min = 0$; $Max = 33$)
 - Evaluationserfahrung: $M = 6.38$ Jahre ($SD = 6.66$; $Min = 0$; $Max = 21$)

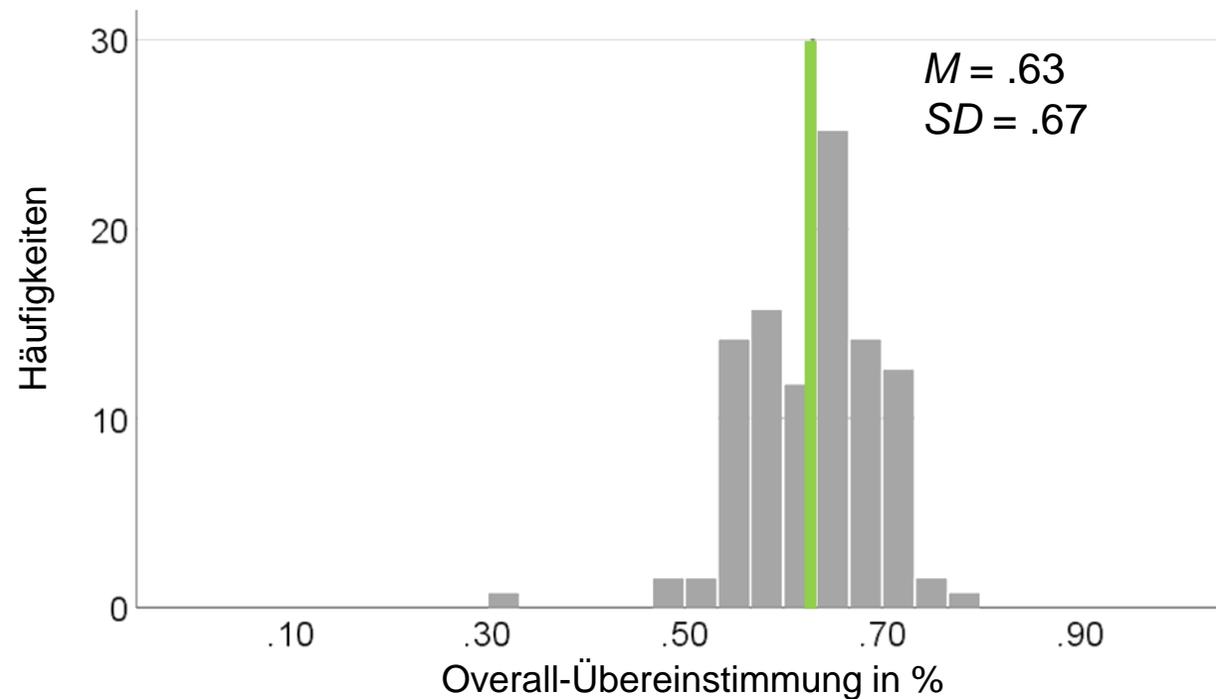
Evaluation der Schulung: Zufriedenheit und Lernertrag

- Skalen mit sechsstufigem Antwortformat (1 = «trifft gar nicht zu» bis 6 = «trifft voll und ganz zu»)

Skala	Items	Beispielitems	<i>M</i>	<i>SD</i>	α
Zufriedenheit Rezjak & Lipowsky, 2013	3	Ich habe die Schulung gerne besucht.	5.26	.80	.84
Lernertrag Benning et al., 2018	10	Ich habe viele Impulse erhalten, die mir etwas für meinen beruflichen Alltag bringen.	5.02	.80	.87
Cognitive Engagement Barlow & Brown, 2019	9	Ich habe die Übungen mithilfe der Schulungsinhalte gelöst.	4.92	.59	.72
Erforschung von Denkweisen Praetorius et al., 2018	4	Die Kursleiterin bat die Teilnehmenden, Begründungen für ihre Antworten zu liefern.	5.18	.84	.92
Diskursives Lernen Praetorius et al., 2018	4	Die Kursleiterin verknüpfte die Diskussionsbeiträge der Teilnehmenden miteinander.	5.12	.76	.82

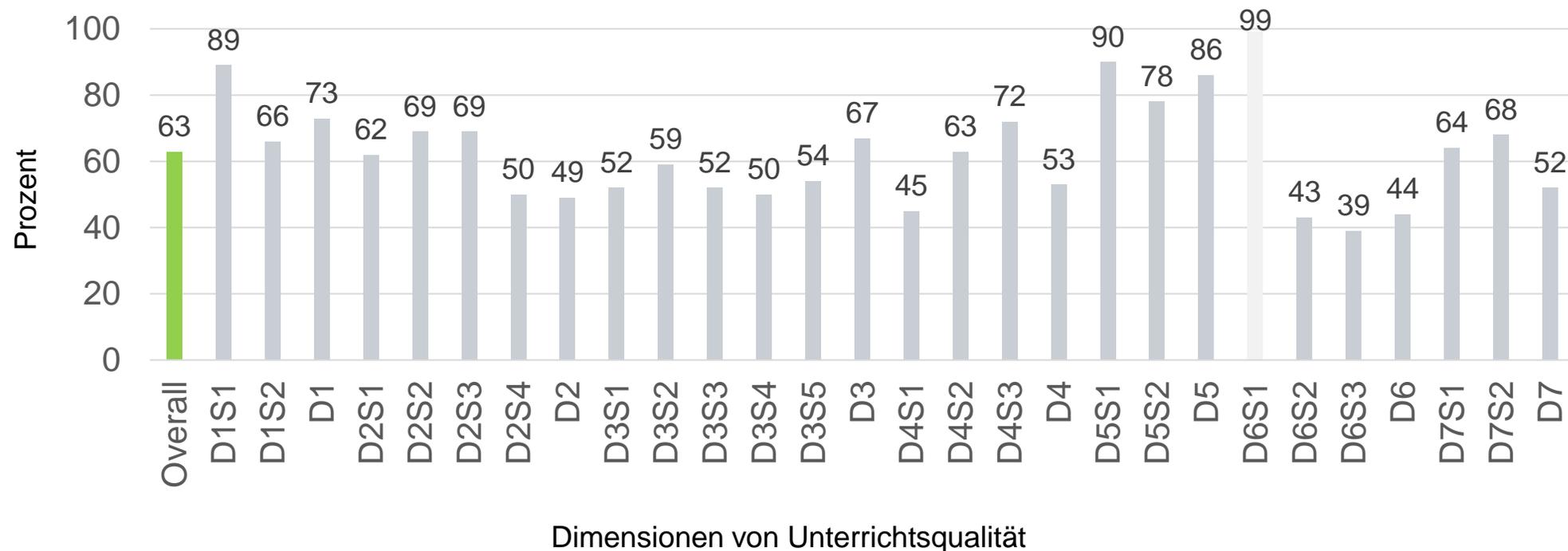
Evaluation der Schulung: Zertifizierung

- Dimensionsübergreifende exakte prozentuale Übereinstimmung mit Masterratings

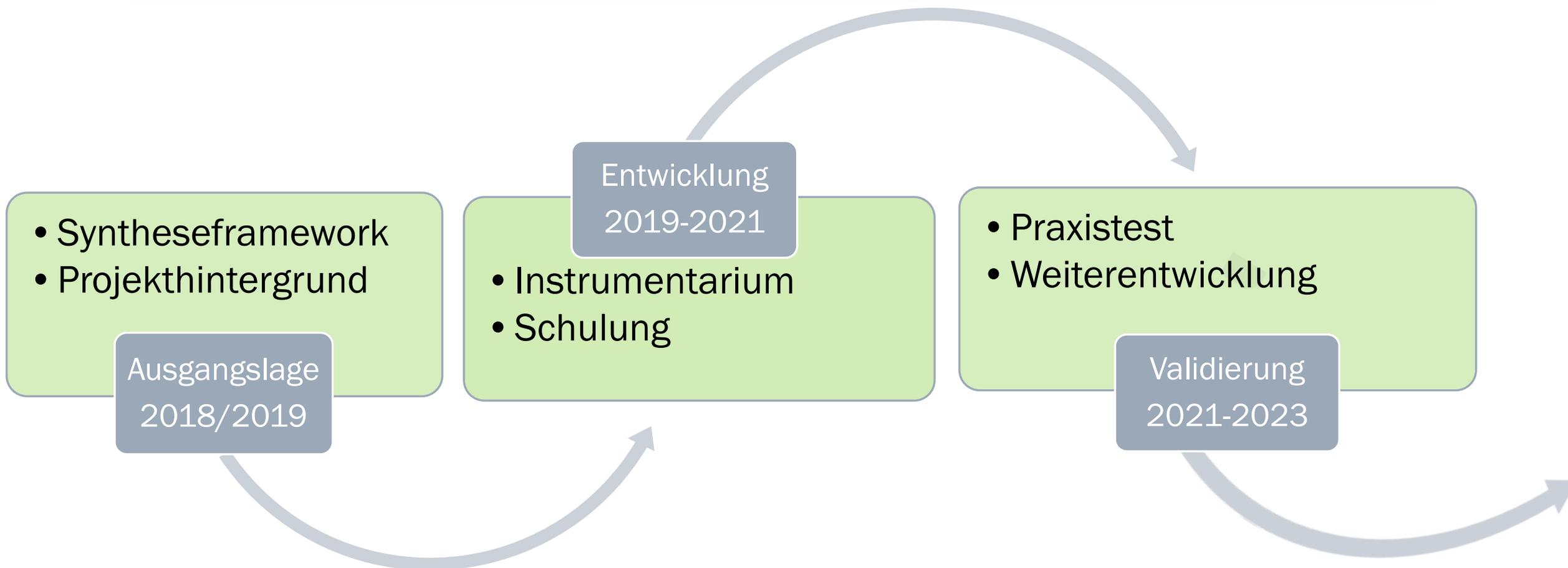


Evaluation der Schulung: Zertifizierung

- Dimensionsspezifische exakte prozentuale Übereinstimmung mit Masterratings



Fazit





INSULA 2.0

KF: Klassenführung

SE: Soziale Unterstützung

EV: Unterstützung der Eigenverantwortung für das Lernen

IN: Auswahl und Thematisierung von Lernzielen, Inhalten und (Fach-)Methoden

KA: Kognitive Aktivierung

KI: Unterstützung des Konsolidierens

BF: Beurteilung und Feedback

AD: Adaptivität

KF

KFS1: Ein proaktiver und effektiver Umgang mit Disziplinfragen ist gewährleistet.
KFS2: Ein hoher Anteil an Lernzeit ist gewährleistet.
KFS3: Der Einsatz von Raum und Ressourcen ist angemessen.

SU

SES1: Es besteht eine positive Beziehung zwischen Lehrpersonen und den Schüler*innen.
SES2: Die Schüler*innen pflegen einen respektvollen Umgang miteinander.

IN

INS1: Die behandelten Lernziele sind adäquat ausgewählt.
INS2: Die behandelten Inhalte und (Fach-)Methoden sind auf das Lernziel abgestimmt.
INS3: Die Relevanz des Inhalts wird den Schüler*innen verdeutlicht.
INS4: Die Inhalte werden klar und strukturiert präsentiert.
INS5: Die Inhalte werden korrekt und präzise dargestellt.

KA

KAS1: Die Aufgaben sind gehaltvoll und auf das kognitive Niveau der Schüler*innen abgestimmt.
KAS2: Die kognitive Aktivität der Schüler*innen wird unterstützt.

KI

KIS1: Aufgaben sind auf die Konsolidierung von Inhalten ausgerichtet.
KIS2: Die Konsolidierung von Inhalten wird unterstützt.

BF

BFS1: Die Beurteilungen sind klar auf die zu erlernenden Kompetenzen ausgerichtet.
BFS2: Die Kompetenzen der Schüler*innen werden überprüft.
BFS3: Die Schüler*innen erhalten qualitativ hochwertiges Feedback.

AD

UHS1: Es gibt eine angemessene Anpassung der Anforderungen an Schüler*innen.
UHS2: Es gibt eine angemessene Anpassung der Lernprozessbegleitung der Schüler*innen.

(Wemmer-Rogh et al., 2023)

Möglichkeiten

- Instrumentarium ermöglicht die zuverlässige und valide Erfassung eines umfassenden Verständnisses von Unterrichtsqualität entsprechend des MAIN-TEACH-Modells.



Umfassende Synthese
für Verständnis von UQ

Zufriedenstellende
Rater-Übereinstimmung

Dimensionsstruktur
statistisch abbildbar

Bezüge zum Angebot-
Nutzungs-Modell über
Operationalisierung

Fachspezifität
insbesondere mit Blick
auf Inhaltsdimension

Systematik mit
Ausrichtung auf
Lernprozesse



Übertragbarkeit auf andere (formative) Nutzungskontexte?

(Charalambous & Praetorius, 2021; Dreher & Leuders, 2021; Farell et al., 2021; Horvath, 2023; Praetorius et al, 2023)

Herausforderungen

- Weitere Forschung mit Blick auf konzeptionelle und methodische Fragen der Unterrichtsqualität ist erforderlich.



Differenzielle
Sensitivität

Kindergartenstufe und
Sonderschulen

Viele offene Baustellen
zu UQ (Nutzung, Üben)

Effektivität von
Rater*innentrainings

Fachspezifität mit Blick
auf weitere Fächer

Prognostische Validität
für SuS-Outcomes



Perspektiven der Unterrichtsqualitätsforschung?

(APA, AERA & NCME, 2014; Begrich et al., 2023; Gitzi et al., 2023; Seidel et al., 2021)

Vielen Dank!

wida.rogh@ife.uzh.ch