BzL BEITRÄGE ZUR LEHRERINNEN- UND LEHRERBILDUNG
Zeitschrift zu Theorie und Praxis der Aus- und Weiterbildung von Lehrerinnen und Lehrern

44. Jahrgang – Heft 2/2026

E-Offprint

Hinweis zum Copyright

Die «Beiträge zur Lehrerinnen- und Lehrerbildung» (BzL) sind eine Open-Access-Zeitschrift ohne Embargo-Frist für die einzelnen Artikel.

Die Autorinnen und Autoren sind frei, die in der Zeitschrift «BzL» publizierte Version («version of record», d.h. den hier vorliegenden E-Offprint) unter der Lizenz Creative Commons Attribution 4.0 International (CC BY 4.0) über weitere Kanäle (z.B. Repositorien, Plattformen, Websites) öffentlich zugänglich zu machen.



Beiträge zur Lehrerinnen- und Lehrerbildung
In Trägerschaft der Schweizerischen Gesellschaft für Lehrerinnen- und Lehrerbildung (SGL)
Erscheint dreimal jährlich

ISSN 2296-9632

Zitiervorschlag: Jeisy, E. & Nüesch, C. (2026). Digitale Leistungsüberprüfungen passend gestalten: Erfahrungen und Empfehlungen aus elf Pilotprojekten in der Lehrer:innenbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung, 44* (2), 232–247. https://doi.org/10.36950/bzl.44.2.2026.10419

www.bzl-online.ch

Editorial

Afra Sturm, Christian Brühwiler, Dorothee Brovelli, Doreen Flick-Holtsch, Bruno Leutwyler, Sandra Moroni und Markus Weil 157

Schwerpunkt

Beurteilen in der Lehrer:innenbildung

Franz Eberle Studierfähigkeit für die Studiengänge der Lehrpersonenbildung für die Volksschule 160

Res Mezger und Nadja Lindauer Formative Beurteilung diskursiver Synthesen von Studierenden mit holistischem Benchmark-Rating 176

Matthias Baumgartner und Eric Jeisy Formative Beurteilung in berufsfeld- und evidenzorientierten Lehr-Lern-Arrangements der Lehrpersonenbildung – ein Typologie- und Topologiemodell 194

Doris Moser-Frötscher, Mirjam Egli Cuenat, Wilfrid Kuster, Thomas C. Messerli, Thomas Roderer und Ingo Thonhauser Berufsspezifische Sprachkompetenz von Fremdsprachenlehrpersonen: Konzeption und Operationalisierung der PROF-L-Prüfung 212

Eric Jeisy und Charlotte Nüesch Digitale Leistungsüberprüfungen passend gestalten: Erfahrungen und Empfehlungen aus elf Pilotprojekten in der Lehrer:innenbildung 232

Forum

Maren Krempin, Simon Moses Schleimer, Anna Gronostaj und Corinna Gottmann Evidenzbasierte Schulentwicklung und Ausbildung von Change Agents: Der Doppelhelixansatz 248

Cornelia Gräsel, Stefan Keller und Anna-Katharina Praetorius Unterrichtsplanung meets Unterrichtsqualitätsforschung: Konzeption und Anwendungsmöglichkeiten des heuristischen Planungstools planTEACH 272

Thomas Rey, Eva Weingartner und Valentin Unger Adaptive Planungskompetenz im Umgang mit Heterogenität: Studentische Perspektiven auf Lerngelegenheiten im Spannungsfeld von Evidenzorientierung und Praxisbedeutsamkeit 296

Susanne Prediger, Christoph Selter, Lars Holzäpfel, Daniela Götze und Bettina Rösken-Winter Fortbildung ausrichten an Kernpraktiken: Kohärenz durch fünf zentrale unterrichtliche Anforderungssituationen im deutschlandweiten Fortbildungsprogramm QuaMath 315

Esther Brunner und Barbara Drollinger-Vetter Weiterbildungsangebote für Mathematikdidaktikerinnen und Mathematikdidaktiker: Ein Beitrag zur Professionalisierung einer Berufsgruppe 337

Marco Galle und Anja Winkler Lernzentrierte Kooperation in Modulbeschreibungen der deutschsprachigen Lehrpersonenbildung in der Schweiz 354

Rubriken

Buchbesprechungen

Plaumauer, C. & Stumpner, S. (Hrsg.). (2024). Über den Tafelrand hinaus denken! Hochschuldidaktische Impulse und Anregungen zur Fallarbeit in der Lehrer:innenbildung. Münster: Waxmann (Franziska Zellweger) 374

Caduff, C. & Weiss, S. (Hrsg.). (2025). Zwischen Theorie und Praxis. Die Qualifizierung von Dozierenden in der Berufsbildung. Bern: hep (Belinda Aeschlimann) 376

Porsch, R., Reintjes, C. & Bellenberg, G. (Hrsg.). (2025). Heterogen qualifizierte Lehrpersonen an Schulen in Deutschland, Österreich und der Schweiz. Münster: Waxmann (Sandra Matschnigg-Peer) 378

Held, O. (2024). ChatGPT im Geschichtsunterricht. Frankfurt am Main: Wochenschau (Kevin van Loon) 380

Beeli-Zimmermann, S. & Wannack, E. (Hrsg.). (2024). Schulen im Spannungsfeld analoger und digitaler Kommunikation. Bern: hep (Jurgita Lenkauskaitė) 382

Kapidzic, S., Gauly, B., Greiff, S., Martin, S., Massing, N., Zabal, A. & Rammstedt, B. (2025). Adaptives Problemlösen Erwachsener im internationalen Vergleich: Ergebnisse von PIAAC 2023. Münster: Waxmann (Franz Eberle) 384

Neuerscheinungen 386

Zeitschriftenspiegel 388

Editorial

Der Themenbereich «Beurteilen» hat für eine wirksame Lehrer:innenbildung hohe Relevanz. Entsprechend wird die Thematik seit vielen Jahren in unterschiedlichen Facetten bearbeitet, was sich auch in den BzL widerspiegelt. So befassten sich frühere Themenhefte beispielsweise mit der Förderung der diagnostischen Kompetenz von Lehrpersonen (z.B. mit der pädagogischen Diagnostik in Heft 2/2013), während sich andere Hefte mit Leistungsnachweisen in der Lehrer:innenbildung (1/2007) oder Fragen zu wissenschaftlichen Abschlussarbeiten (Doppelnummer 1/2021 und 2/2021) auseinandersetzten. Die grundlegende Aufgabe von Beurteilung im Bildungskontext liegt darin, lern- oder entwicklungsrelevante Merkmale akkurat einzuschätzen. Ein diagnostisches Urteil stellt jedoch auch in der Lehrer:innenbildung keinen Selbstzweck dar, sondern schafft eine fundierte Grundlage für die Feststellung des Lern- oder Leistungsstands der Studierenden, um daraus Massnahmen für die Förderung und die Unterstützung der professionellen Entwicklung von künftigen Lehrpersonen (formative Beurteilung) ableiten zu können oder um Selektionsentscheide (summative Beurteilung) zu fällen, zu begründen und zu legitimieren.

Das vorliegende Themenheft nimmt Fragen rund um das Beurteilen professioneller Kompetenzen in der Ausbildung von künftigen Lehrpersonen auf. Dabei werden aktuelle Themenbereiche und kontroverse Diskussionsfelder aus verschiedenen Perspektiven beleuchtet: Was wird unter Studierfähigkeit in der Lehrer:innenbildung verstanden, wie wird sie erfasst und welche Fördermassnahmen werden daraus abgeleitet? Wie lassen sich professionelle Kompetenzen von (künftigen) Lehrpersonen in unterschiedlichen fachspezifischen Kontexten beurteilen? Mit welchen (innovativen) Methoden werden relevante berufsspezifische Kompetenzen beurteilt und wie werden sie gefördert? Wie lassen sich formative Beurteilungssettings ressourcenschonend umsetzen?

Der einleitende Beitrag von **Franz Eberle** gibt einen Überblick über das System der Zulassung zu Pädagogischen Hochschulen und differenziert dabei zwischen Studierfähigkeit und der angestrebten Berufsbefähigung – zwei Aspekte, die manchmal vermischt werden. Der Beitrag führt genauer aus, welche Kompetenzbündel bei einer allgemeinen und bei einer spezifischen Studierfähigkeit zentral sind und wie sich die verschiedenen Zulassungen hinsichtlich Studienerfolg und Studienabbruch verhalten. Abschliessend zeigt der Autor auf, dass es weiterhin gute Gründe gibt, weshalb Berufsmaturand:innen eine Aufnahme- oder Ergänzungsprüfung benötigen, um zu einem Studium an einer Pädagogischen Hochschule zugelassen zu werden.

Der Beitrag von **Res Mezger und Nadja Lindauer** widmet sich dem Thema der formativen Beurteilung diskursiver Synthesen von künftigen Lehrpersonen. Sie gehen der Frage nach, wie wissenschaftliche Lese- und Schreibkompetenzen von Studienanfänger:innen systematisch erfasst und während der Ausbildung gefördert werden können. Im Beitrag wird ein formatives Assessment auf der Basis eines

holistischen Benchmark-Ratings vorgestellt und es werden darauf abgestimmte Fördermassnahmen näher beleuchtet.

Der Beitrag von **Matthias Baumgartner und Eric Jeisy** adressiert die Funktionen und Möglichkeiten, aber auch die Grenzen formativer Beurteilung in hochschuldidaktischen Lehr-Lern-Arrangements der Lehrpersonenbildung. Im Rahmen einer sportdidaktischen Interventionsstudie wird die Bedeutung von iterativen Coaching-Zyklen und formativer Beurteilung für den Performanzzuwachs nachgewiesen. Schliesslich werden Möglichkeiten für eine aufwandreduzierende Gestaltung derartiger Lehr-Lern-Arrangements (z.B. mittels automatisierter Videoproduktion) erörtert.

In ihrem Beitrag befassen sich **Doris Moser-Frötscher, Mirjam Egli Cuenat, Wilfrid Kuster, Thomas C. Messerli, Thomas Roderer und Ingo Thonhauser** mit der Förderung und der Beurteilung berufsspezifischer Sprachkompetenzen in der Ausbildung von Fremdsprachenlehrpersonen. Am Beispiel der Professionellen Sprachprüfung für Lehrpersonen (PROF-L) wird die Operationalisierung von Sprachkompetenzen in berufsrelevanten Handlungssituationen vorgestellt und das zugrunde liegende Kompetenzverständnis verortet. Zudem werden Einblicke in die institutionelle Rolle und die Nutzung von PROF-L an der HEP Vaud gewährt.

In einem Praxisbeitrag berichten **Eric Jeisy und Charlotte Nüesch** über die Entwicklung und die Erprobung von elf digitalen Leistungsüberprüfungen, die im Rahmen des PgB-Programms P-8 «Stärkung von Digital Skills in der Lehre» von swissuniversities entstanden sind. Die Evaluationsergebnisse weisen darauf hin, dass nicht das digitale Format an sich für die Qualität digitaler Leistungsüberprüfungen entscheidend ist, sondern die Abstimmung von angestrebter Kompetenz, Assessmentfunktion, Aufgabe, Medium sowie technischen und organisatorischen Bedingungen. Aus den Erfahrungen werden ein Orientierungsrahmen und konkrete Hinweise für die Gestaltung digitaler Leistungsüberprüfungen abgeleitet.

Im Forumsteil befassen sich sechs Beiträge mit unterschiedlichen Aspekten der Aus- und Weiterbildung von Lehrpersonen: Im Mittelpunkt des Beitrags von **Maren Krempin, Simon Moses Schleimer, Anna Gronostaj und Corinna Gottmann** steht ein Ansatz für eine evidenzorientierte Schulentwicklung, der sich der Analogie der Doppelhelix bedient. Dieser Ansatz zielt darauf ab, pädagogisches Schulpersonal unter Berücksichtigung aktueller Erkenntnisse der Schulqualitäts-, Schuleffektivitäts- und Schulentwicklungsforschung zu Change Agents auszubilden, die parallel zum Besuch der Fortbildung in ihren Schulen systematisch dazu beitragen, evidenzorientierte Schulentwicklung wirkungsvoll umzusetzen.

Cornelia Gräsel, Stefan Keller und Anna-Katharina Praetorius stellen in ihrem Beitrag das evidenzorientierte Unterrichtsplanungstool planTEACH vor, das zum Ziel hat, die Professionalisierung von (künftigen) Lehrpersonen bei der Unterrichtsplanung

zu unterstützen. planTEACH nimmt Befunde aus der Unterrichtsqualitätsforschung systematisch auf und bietet für jeden Planungsschritt (z.B. Lernzielbestimmung, Diagnostik der Lernvoraussetzungen, Ablauf, Aufgabenwahl) empirisch informierte, praxistaugliche Hinweise an. Zum Beitrag gehört der Online-Anhang «planTEACH: Schritt für Schritt zum guten Unterricht», der einen konkreten Leitfaden mit Fragen und Hinweisen enthält.

Thomas Rey, Eva Weingartner und Valentin Unger stellen die Konzeption und die Evaluation eines Hochschulseminars zu allgemeiner adaptiver Unterrichtsplanungskompetenz vor. Sie stützen sich dabei unter anderem auf empirische Vorarbeiten, bei denen Kriterien zur Gestaltung und Reflexion adaptiver Unterrichtsentwicklung erarbeitet und validiert wurden. Im Seminar wandten Studierende der Sekundarstufe I und II ihr im Seminar erworbenes Wissen auf reale Unterrichtsplanungen an. Die Evaluation zeigt, dass sie ihren Lerngewinn als hoch und praxisbedeutsam einschätzten.

Susanne Prediger, Christoph Selter, Lars Holzäpfel, Daniela Götze und Bettina Rösken-Winter gehen in ihrem Beitrag auf die Frage ein, was eine Ausrichtung an typischen unterrichtlichen Anforderungssituationen sowie (Kern-)Praktiken für die Gestaltung von effektiven Fortbildungen bedeutet. Sie nehmen zunächst sowohl aus generischer als auch aus fachspezifischer Perspektive eine theoretische Klärung vor und identifizieren dann fünf mathematikdidaktische Anforderungssituationen, die in das Fortbildungsprogramm QuaMath integriert wurden.

Esther Brunner und Barbara Drollinger-Vetter stellen ein Entwicklungsprojekt vor, das sich an Mathematikdidaktiker:innen richtet. Sie unterscheiden zunächst zwischen der für die Lehrpersonenbildung typischen doppelten Professionalisierung und einer kollektiven sowie individuellen Professionalisierung. Mit Blick auf Letztere führten sie eine Befragung bei amtierenden Mathematikdidaktiker:innen durch, um den Weiterbildungsbedarf genauer eruieren zu können. Abschliessend stellen sie verschiedene Weiterbildungsangebote vor.

Marco Galle und Anja Winkler führten zur Frage, inwiefern sich in den öffentlich zugänglichen Modulbeschreibungen von zwölf Pädagogischen Hochschulen der Deutschschweiz das Konzept der lernzentrierten Kooperation zeigt, eine quantitative Dokumentenanalyse durch. Diese Dokumentenanalyse ergänzten sie um eine Netzwerkkarte, um die Verhältnisse von Kooperation und Lernzentrierung besser abbilden zu können. Damit können sie zeigen, dass Kooperation in den Modulbeschreibungen primär als kommunikative Tätigkeiten beschrieben wird und dass die Beschreibungen selten gleichzeitig auch mit Bezug zu Lernzentrierung erfolgen. Daraus leiten die Autor:innen einen Handlungsbedarf für Pädagogische Hochschulen ab.

Afra Sturm, Christian Brühwiler, Dorothee Brovelli, Doreen Flick-Holtsch, Bruno Leutwyler, Sandra Moroni und Markus Weil

Digitale Leistungsüberprüfungen passend gestalten: Erfahrungen und Empfehlungen aus elf Pilotprojekten in der Lehrer:innenbildung

Eric Jeisy und Charlotte Nüesch

Zusammenfassung Der Praxisbeitrag berichtet über die Entwicklung und die Erprobung von elf digitalen Leistungsüberprüfungen in der Lehrer:innenbildung. Studierendenbefragungen und Reflexionen der beteiligten Dozierenden weisen darauf hin, dass nicht das digitale Format an sich entscheidend ist, sondern die Abstimmung von angestrebter Kompetenz, Assessmentfunktion, Aufgabe, Medium sowie technischen und organisatorischen Bedingungen. Aus den Erfahrungen werden ein Orientierungsrahmen und konkrete Hinweise für die Gestaltung digitaler Leistungsüberprüfungen abgeleitet.

Schlagwörter digitale Leistungsüberprüfung – kompetenzorientiertes Prüfen – Constructive Alignment – Lehrer:innenbildung

Designing fit-for-purpose digital assessments: Experiences and recommendations from eleven pilot projects in teacher education

Abstract This practice-oriented contribution reports on the development and piloting of eleven digital assessments in teacher education. Student surveys and reflections by participating lecturers indicate that the digital format itself is less decisive than the alignment of intended competence, assessment function, task, medium, and technical and organisational conditions. Based on these experiences, the article derives a practical framework and concrete recommendations for designing digital assessments.

Keywords digital assessment – competency-based assessment – constructive alignment – teacher education

DOI 10.36950/bzl.44.2.2026.10419 | **ISSN** 2296-9632

Dieses Werk steht unter der Lizenz Creative Commons Namensnennung 4.0 International.

1 Einleitung

Die Beurteilung studentischer Leistungen ist ein zentraler Bestandteil kompetenzorientierter Hochschullehre. Prüfungen sollen nicht nur Wissen erfassen, sondern – abhängig von den angestrebten Kompetenzen – auch dessen Anwendung in komplexeren und berufsfeldrelevanten Anforderungssituationen ermöglichen (Schaper, 2012; Zimmermann, 2024). Gerade in der Lehrer:innenbildung stellt sich damit die Frage, wie professionelle Handlungskompetenzen angemessen überprüft werden können.

Digitale Technologien erweitern die Möglichkeiten, Leistungsüberprüfungen[1] zu gestalten. So können beispielsweise Video- und Audiomaterialien, digitale Artefakte, berufstypische Anwendungen oder E-Portfolios in Leistungsnachweise einbezogen werden (Bandtel et al., 2021; Bauer-Klebl & Nüesch, 2022). Dabei ist auch zu berücksichtigen, welche Aspekte eines Prüfungsgegenstands durch ein digitales Medium zugänglich werden und welche möglicherweise weniger angemessen abgebildet werden.

Vor diesem Hintergrund wurden an der Pädagogischen Hochschule St.Gallen elf bestehende Leistungsüberprüfungen aus unterschiedlichen Studiengängen digital weiterentwickelt und erprobt.[2] Ziel war es, digitale Möglichkeiten für die kompetenzorientierte Gestaltung von Leistungsüberprüfungen zu nutzen und Erfahrungen für deren Weiterentwicklung und den hochschulinternen Transfer zu gewinnen. Der vorliegende Praxisbeitrag bündelt projektübergreifend Erfahrungen von Studierenden und Dozierenden mit den erprobten Prüfungssettings. Mit der videobasierten Beurteilung von Tanzperformanzen mittels Adaptive Comparative Judgment (ACJ) wird ergänzend ein Pilotprojekt vertieft betrachtet. Aus den Erfahrungen werden ein Orientierungsrahmen und konkrete Hinweise für die Gestaltung digitaler Leistungsüberprüfungen in der Lehrer:innenbildung abgeleitet.

2 Vom Prüfungsziel zum digitalen Setting – didaktische Bezugspunkte

2.1 Kompetenz als Ausgangspunkt der Leistungsüberprüfung

Professionelle Kompetenz umfasst mehr als verfügbares Fach- und Professionswissen. Blömeke, Gustafsson und Shavelson (2015) konzeptualisieren Kompetenz als Kontinuum, das kognitive und motivational-affektive Dispositionen über situationsspezifische Prozesse der Wahrnehmung, Interpretation und Entscheidung mit beobachtbarer Performanz verbindet. Je nach angestrebter Kompetenz können deshalb unterschiedliche Ausschnitte professionellen Handelns Gegenstand einer Leistungsüberprüfung sein.

---

[1] In Anlehnung an das Allgemeine Prüfungsreglement der Pädagogischen Hochschule St.Gallen wird «Leistungsüberprüfung» als übergeordneter Begriff für formal beurteilte Studienleistungen verwendet (Pädagogische Hochschule St.Gallen, 2023, Art. 1).

[2] Das Projekt war Teil des von swissuniversities geförderten PgB-Programms P-8 «Stärkung von Digital Skills in der Lehre» (https://www.swissuniversities.ch/themen/digitalisierung/digital-skills).

Kompetenzorientiertes Prüfen richtet Aufgaben und Beurteilung daran aus, welche Leistungen Studierende zeigen müssen, damit begründete Rückschlüsse auf die angestrebten Kompetenzen möglich sind (Schaper, 2012). Dies können neben wissensbezogenen Aufgaben beispielsweise die Analyse beruflicher Situationen, die Bearbeitung komplexer Problemstellungen, die Entwicklung von Lernprodukten oder beobachtbare Performanzen sein. Entsprechende handlungsnahe und situierte Assessmentformen haben auch in der Lehrer:innenbildung an Bedeutung gewonnen (Albu & Lindmeier, 2023). Handlungsnähe ist dabei nicht mit einer möglichst weitgehenden Nachbildung beruflicher Situationen gleichzusetzen. Entscheidend ist, welche Aspekte professioneller Kompetenz beurteilt werden sollen und welche Prüfungsanforderung diese angemessen zugänglich macht. Ausgangspunkt der Prüfungsgestaltung sollte deshalb die angestrebte Kompetenz sein.

2.2 Funktion und Passung der Leistungsüberprüfung

Einen Bezugsrahmen für die Abstimmung von Kompetenzzielen, Lehre und Leistungsüberprüfung bietet das Constructive Alignment. Intendierte Lernergebnisse, Lehr-Lern-Aktivitäten und Assessment werden dabei so aufeinander bezogen, dass die angestrebten Leistungen im Lernprozess ermöglicht und in der Leistungsüberprüfung tatsächlich eingefordert werden (Biggs, 1996; Biggs & Tang, 2011). Die Wahl eines Prüfungsformats sollte entsprechend nicht von den verfügbaren technischen Möglichkeiten ausgehen, sondern von den Kompetenzen, die Studierende zeigen sollen. Prüfungsevaluationen können dazu beitragen, die Abstimmung dieser Elemente zu reflektieren und Prüfungssettings weiterzuentwickeln (Bauer-Klebl & Nüesch, 2022).

Zu berücksichtigen ist zudem die Funktion der Leistungsbeurteilung. Diagnostische Beurteilungen erfassen Lernvoraussetzungen, formative Beurteilungen unterstützen laufende Lernprozesse und summative Beurteilungen bewerten erreichte Leistungen und sind häufig mit Zertifizierungs- oder Berechtigungsentscheidungen verbunden (Metzger & Nüesch, 2004; Zimmermann, 2024). Dasselbe Verfahren kann deshalb je nach Funktion unterschiedlich geeignet sein. Für die Prüfungsgestaltung sind damit zunächst zwei Fragen zu klären: Welche Kompetenz soll sichtbar werden und zu welchem Zweck wird sie beurteilt? Erst darauf aufbauend lässt sich bestimmen, welche Aufgaben- und Beurteilungsform geeignet ist und welche Rolle digitale Technologien dabei übernehmen können.

2.3 Digitale Prüfungssettings als Gestaltungsraum

Digitale Leistungsüberprüfung wird im vorliegenden Beitrag in Anlehnung an Bandtel et al. (2021) als Oberbegriff für Prüfungsszenarien verwendet, bei denen die Bearbeitung oder die Abgabe zumindest teilweise digital erfolgt. Darunter fallen unterschiedliche Settings wie digitale Vor-Ort-Prüfungen, ausserhalb der Hochschule bearbeitete Take-Home-Prüfungen, E-Portfolios oder mündliche Online-Prüfungen, wobei verschiedene Elemente miteinander kombiniert werden können (Bandtel et al., 2021; Reinmann, 2021).

Digitale Technologien können den Gestaltungsraum von Leistungsüberprüfungen erweitern, beispielsweise indem multimediale Situationen, digitale Artefakte oder zeitlich verteilte Bearbeitungs- und Feedbackprozesse einbezogen werden. Mit der digitalen Umsetzung verändern sich jedoch zugleich die Bedingungen, unter denen Leistungen erbracht, sichtbar und beurteilt werden. Digitale Medien bilden einen Prüfungsgegenstand dabei nicht lediglich ab, sondern strukturieren auch, welche Merkmale einer Leistung zugänglich werden. So können etwa Ausschnitt, Perspektive, Wiederholbarkeit oder zeitliche Entkopplung bestimmte Aspekte hervorheben, während andere weniger unmittelbar erfahrbar sind. Kulgemeyer et al. (2023) zeigen für die Lehrer:innenbildung, dass die Digitalisierung von Assessmentverfahren auch deren Authentizität und Aussagekraft beeinflussen kann. Hinzu kommen technische und organisatorische Anforderungen, etwa an Infrastruktur, Support und zuverlässige Prüfungsbedingungen (Budde, Tobor & Beyermann, 2023).

Der Forschungsstand zu digitalem Assessment zeigt entsprechend eine grosse Vielfalt an Verfahren und Ergebnissen (Hu, Liu & Daniel, 2025). Daraus lässt sich nicht ableiten, dass bestimmte digitale Formate grundsätzlich geeigneter sind als andere. Ihre Eignung ist vielmehr im jeweiligen Prüfungskontext zu beurteilen. Ausgangspunkt bilden die angestrebte Kompetenz und die Funktion der Leistungsbeurteilung. Darauf bezogen sind Aufgabe und Beurteilungsform sowie die mediale, technische und organisatorische Umsetzung zu bestimmen. Diese Abstimmung bildet den Bezugspunkt für die im Folgenden dargestellten Pilotierungen.

3 Entwicklungsprojekt mit elf Pilotierungen digitaler Leistungsüberprüfungen

3.1 Projekt und Erfahrungsgrundlage

Das Entwicklungsprojekt zielte darauf ab, Dozierende der Pädagogischen Hochschule St.Gallen bei der Entwicklung digitaler Leistungsüberprüfungen zur Beurteilung berufsfeldbezogener Handlungskompetenzen zu unterstützen und Erfahrungen für den hochschulinternen Transfer zu gewinnen. Aus 17 Projekteingaben wurden elf Pilotprojekte ausgewählt. Kriterien waren die inhaltliche Passung zum Förderprojekt, die Umsetzbarkeit, die Transferfähigkeit sowie der Innovationsgrad der geplanten digitalen Umsetzung. Dieser bezog sich darauf, inwiefern digitale Technologien bestehende Prüfungssettings funktional erweitern oder verändern können. Die Auswahl zielte damit bewusst auf ein digitales Entwicklungspotenzial. Daraus wurde jedoch nicht abgeleitet, dass eine stärkere Digitalisierung didaktisch grundsätzlich vorzuziehen ist. Zudem wurde eine gewisse Breite in Bezug auf Fachbereiche und Studiengänge angestrebt. Insgesamt waren 16 Dozierende aus dem Bachelorstudiengang «Kindergarten- und Primarschulstufe», dem Masterstudiengang «Sekundarstufe I» sowie der Weiterbildung beteiligt.

Die Entwicklung der Prüfungssettings wurde durch eine interne Weiterbildung begleitet. Die Dozierenden brachten ihr jeweiliges Modul ein, entwickelten ein auf die angestrebten Kompetenzen und Lehr-Lern-Aktivitäten abgestimmtes digitales Prüfungssetting und erprobten dieses anschliessend im eigenen Modul. Die Weiterbildung folgte einem handlungsorientierten Blended-Learning-Ansatz und bezog Elemente der kooperativen Praxisbewältigung in Gruppen ein («KOPING»; Wahl, 2013). Die elf Pilotprojekte unterschieden sich hinsichtlich Prüfungsanforderung, Assessmentfunktion und digitaler Umsetzung. Tabelle 1 gibt einen Überblick über die erprobten Settings und fasst jeweils eine zentrale Erfahrung aus der Umsetzung zusammen.

Tabelle 1: Übersicht über die elf Pilotprojekte und zentrale Erfahrungen aus der Umsetzung

Pilotprojekt	**Digitales Prüfungssetting**	**Zentrale Prüfungsanforderung**	**Funktion**	**Zentrale Erfahrungen aus der Umsetzung**
BNE-Didaktik	LMS-Lernpfad[a]/ Take-Home	BNE-orientierte Unterrichtsplanung schrittweise entwickeln und auf ein selbstgewähltes Thema im NMG-Unterricht anwenden	formativ + summativ	Strukturierter Lernpfad unterstützt selbstständige Bearbeitung; systematische Einführung bleibt wichtig
Sozialwissenschaftliche Methoden	LMS-Lernpfad[a]/ Take-Home	Fachinhalte selbstorganisiert erarbeiten, Lernstand überprüfen, Meilensteine bearbeiten und Peer-Feedback geben	formativ + summativ	Zeitliche Flexibilität wird geschätzt, erfordert aber klare Kursstruktur und Freischaltungen
Kommunikation	LMS-Lernpfad[a] + E-Portfolio	Eigenes Kommunikationshandeln videobasiert analysieren und reflektieren (Videografierung durch Studierende); Peer-Feedback	formativ + summativ	Praxisbezug wird geschätzt; Peer-Feedback und Selbsteinschätzung werden unterschiedlich beurteilt
Grundlagen Medienmentor:in	E-Portfolio	Berufspraktische Erfahrungen dokumentieren, reflektieren und weiterentwickeln	formativ + summativ	Technische Offenheit kann zusätzlichen Unterstützungsbedarf erzeugen
Kompetenzportfolio Erwachsenenbildung	E-Portfolio	Kompetenzentwicklung dokumentieren und anhand von Artefakten nachweisen	formativ + summativ	Technische Anforderungen dürfen den eigentlichen Kompetenznachweis nicht überlagern
Grundzubereitungsarten (WAH)	E-Portfolio + praktische Prüfung	Zubereitungsverfahren dokumentieren/ reflektieren und praktisch anwenden	formativ + summativ	Aufwand einzelner Portfolioaufträge wurde reduziert; Werkzeugwahl wurde flexibilisiert

Kernmodul «Medien & Informatik»	E-Portfolio + Fachgespräch	Kompetenzentwicklung anhand von Artefakten dokumentieren, Peer-Feedback und vertiefende Diskussion	formativ + summativ + diagnostisch	Austausch und gegenseitige Unterstützung wurden positiv erlebt, erfordern aber eine verbindliche organisatorische Rahmung
Lernen/Klassenführung und soziales Lernen	Digitale Vor-Ort-Prüfung	Unterrichtssituationen anhand von bestehenden Videovignetten theoriegestützt analysieren	summativ	Video unterstützt den Situationsbezug
Fachdidaktik Deutsch	Digitale Vor-Ort-Prüfung mit Safe Exam Browser[b]	Unterrichtssituationen, Fälle und sprachdidaktische Aufgaben analysieren/bearbeiten	summativ	Gute Vorbereitung und Funktionstests unterstützen die Durchführung; Bedienung muss mitgedacht werden
Grundlagen des Unterrichtens	Digitale Vor-Ort-Prüfung + mündliche Online-Prüfung	Lehrgespräch analysieren sowie eigene Praxiserfahrungen theoriegestützt reflektieren	summativ	Nicht jedes digitalisierbare Prüfungselement bietet einen Mehrwert; mündliche Online-Prüfung wurde wieder aufgegeben
Bewegung und Sport (Tanz)	Videobasiertes Adaptive Comparative Judgment	Tänzerische Performanz demonstrieren	summativ erprobt	Stark übereinstimmende Rangordnungen, aber Vorbehalte gegenüber der Beurteilungslogik für die summative Prüfung

Anmerkungen: [a] LMS-Lernpfade bezeichnen innerhalb eines Learning Management Systems sequenziell strukturierte Abfolgen von Lern- und Prüfungselementen. [b] Der Safe Exam Browser ist ein für Prüfungssituationen konfigurierter Browser, der den Zugriff auf weitere Anwendungen und Ressourcen auf dem Prüfungsgerät einschränkt. BNE = Bildung für nachhaltige Entwicklung; LMS = Learning Management System; NMG = Natur, Mensch, Gesellschaft; WAH = Wirtschaft, Arbeit, Haushalt. Alle Pilotprojekte umfassten einen *summativen* Kompetenznachweis. Als *formativ* werden zusätzliche zeitraumbezogene Assessmentelemente wie Kurztests oder Peer-Feedback bezeichnet, die den Lernprozess unterstützen sollen. Die letzte Spalte fasst jeweils eine ausgewählte Erfahrung aus den Studierendenbefragungen bzw. den Reflexionen der beteiligten Dozierenden zusammen. Die projektübergreifende Einordnung erfolgt in Abschnitt 3.2.

Zehn der elf Pilotierungen wurden mittels Online-Befragungen aus der Perspektive der Studierenden begleitet. Die Fragebogen enthielten einen projektübergreifenden Kern sowie projektspezifische geschlossene und offene Fragen. Von 908 eingeladenen Studierenden nahmen 495 teil (55%); die projektbezogenen Rücklaufquoten lagen zwischen 25% und 86%. Aufgrund unterschiedlicher Stichprobengrössen (N = 6 bis N = 142) und teilweise projektspezifischer Instrumente werden keine quantitativen Vergleiche zwischen den Pilotprojekten vorgenommen.

Ergänzend reflektierten die beteiligten Dozierenden und Mitarbeitenden des technischen Supports ihre Erfahrungen mit der Entwicklung und Durchführung der Prüfungssettings in Gruppengesprächen und projektbezogenen Erfahrungsberichten. Für die folgende Darstellung wurden wiederkehrende Erfahrungen aus beiden Perspektiven projektübergreifend gebündelt. Das Pilotprojekt zum ACJ wurde mit einem eigenständigen Untersuchungsdesign durchgeführt und wird in Abschnitt 4 separat dargestellt.

3.2 Projektübergreifende Erfahrungen zur Gestaltung digitaler Leistungsüberprüfungen

Über die unterschiedlichen Pilotprojekte hinweg zeigten sich ähnliche Erfahrungen hinsichtlich Praxis- und Anwendungsbezugs, Flexibilität und Struktur, Aufwand, medialtechnischer Umsetzung sowie Peer-Feedback und Selbstbeurteilung. Nicht alle diese Erfahrungen sind spezifisch für digitale Leistungsüberprüfungen. Teilweise betreffen sie grundlegende Fragen der Prüfungsgestaltung. Im Folgenden wird deshalb jeweils unterschieden, ob sich ein Befund auf die Prüfungsanforderung insgesamt oder spezifisch auf deren digitale Umsetzung bezieht.

3.2.1 Praxis- und Anwendungsbezug

Studierende beurteilten Prüfungssettings insbesondere dann positiv, wenn sie die Anwendung von Wissen und einen Bezug zur künftigen Berufspraxis erkannten. In der Zwischenprüfung zu Lernen, Klassenführung und sozialem Lernen wurde beispielsweise die Anwendungsorientierung gegenüber dem Auswendiglernen positiv hervorgehoben. Auch bei Portfolioformaten und Fachgesprächen wurden die Verbindung von Theorie mit eigenen Produkten oder Praxiserfahrungen und deren mögliche Verwendung im späteren Unterricht geschätzt.

Kritischer fielen Rückmeldungen aus, wenn der Anwendungs- oder Berufsfeldbezug wenig erkennbar war. Im Kommunikationsprojekt wurden stärker berufsfeldbezogene Aufgaben gewünscht. Im Modul «Sozialwissenschaftliche Methoden» wurden Inhalte teilweise als abstrakt und wenig anschaulich wahrgenommen; aus den Rückmeldungen liess sich jedoch nicht eindeutig ableiten, ob sich diese Einschätzung auf die Inhalte selbst oder auf deren Aufbereitung im digitalen Setting bezog. Der wahrgenommene Praxisbezug erwies sich damit zunächst als Frage der Prüfungsanforderung und ihrer erkennbaren beruflichen Relevanz, nicht der digitalen Umsetzung an sich.

3.2.2 Flexibilität und Struktur

Bei zeitraumbezogenen Settings wurden zeitliche und räumliche Flexibilität sowie die Möglichkeit zur Bearbeitung im eigenen Tempo wiederholt positiv beurteilt. Im Kommunikationsprojekt hoben Studierende insbesondere die Übersichtlichkeit des Lernpfads und die über das Semester verteilten Abgabetermine hervor. Flexibilität erwies sich jedoch nicht unabhängig von der Gestaltung des Settings als Vorteil. Im Modul «Sozialwissenschaftliche Methoden» nannten 13 Rückmeldungen das asynchrone und ortsunabhängige Lernen positiv, während sich 16 kritische Rückmeldungen auf

organisatorische Schwierigkeiten, Freischaltungen und die Kursstruktur bezogen. Zeitliche und räumliche Freiheitsgrade wurden somit insbesondere dort positiv erlebt, wo Arbeitsaufträge, Erwartungen und Abläufe nachvollziehbar strukturiert waren.

3.2.3 Aufwand und Ressourcen

Bei zeitraumbezogenen Leistungsüberprüfungen wurde wiederholt das Verhältnis von Arbeitsaufwand und wahrgenommenem Nutzen für die eigene Kompetenzentwicklung oder die spätere Berufstätigkeit thematisiert. Beim E-Portfolio im Bereich «Wirtschaft, Arbeit, Haushalt» bezogen sich beispielsweise 13 offene Rückmeldungen von Studierenden auf einen hohen zeitlichen Aufwand und einen teilweise als gering eingeschätzten zusätzlichen Ertrag einzelner Dokumentations- und Reflexionsaufträge. In der Folge wurde der Umfang des Portfolios reduziert.

Auch aus der Sicht der Dozierenden war die erstmalige Entwicklung digitaler Prüfungssettings teilweise zeit- und planungsintensiv. Dies betraf insbesondere komplexere Lernpfade und Portfolioformate sowie videobasierte Elemente, bei denen je nach Setting die technische Einbindung, Aufzeichnung, Bereitstellung und Speicherung zusätzlichen Aufwand erzeugten. Als hilfreich wurden frühzeitige Planung, Testphasen und technischer Support beschrieben. Nach der Erstentwicklung konnten sich zugleich Effizienzgewinne ergeben, etwa durch die Wiederverwendung bestehender Lernpfade oder automatisiert auswertbare Aufgaben. Digitale Leistungsüberprüfungen erwiesen sich damit nicht grundsätzlich als ressourcensparend. Der Aufwand verlagerte sich teilweise von Durchführung und Korrektur hin zu Entwicklung, technischer Vorbereitung und Begleitung.

3.2.4 Prüfungsgegenstand, Medium und technische Umsetzung

Technische Stabilität sowie die darstellungsbezogene Gestaltung beeinflussten die Nutzbarkeit der erprobten Settings. Bei der digitalen Fachdidaktik-Deutsch-Prüfung mit dem Safe Exam Browser wurden vorgängige Informationen, Funktionstest und Unterstützung positiv bewertet; gleichzeitig erwiesen sich einzelne Aufgaben aufgrund ihrer Bildschirmdarstellung als unübersichtlich. Bei einer Prüfung mit Videovignetten unterstützten die Videos zwar den Anwendungsbezug, 16 Studierende berichteten jedoch von Problemen beim Abspielen oder mit der Netzwerkverbindung.

Über diese technischen und darstellungsbezogenen Aspekte hinaus zeigte sich, dass die Wahl des Mediums beeinflussen kann, welche Merkmale eines Prüfungsgegenstands zugänglich werden. Im Modul «Grundlagen des Unterrichtens» musste eine geplante videobasierte Analyse aufgrund technischer Kapazitätsgrenzen durch die Arbeit mit einem schriftlichen Transkript ersetzt werden. Die Analyse blieb möglich, der ursprünglich beabsichtigte realitätsnähere Kontext konnte jedoch nur eingeschränkt abgebildet werden. Eine zusätzlich erprobte mündliche Online-Prüfung wurde nicht weitergeführt, da gegenüber der Durchführung vor Ort kein hinreichender Mehrwert gesehen wurde. Technische Funktionsfähigkeit allein erwies sich damit nicht als hinreichendes

Kriterium für die Eignung eines digitalen Prüfungssettings. Ebenso relevant ist, wie das gewählte Medium den Prüfungsgegenstand repräsentiert, welche Merkmale dadurch zugänglich werden und welche zusätzlichen Anforderungen mit seiner Nutzung verbunden sind.

3.2.5 Peer-Feedback und Selbstbeurteilung

Die Erfahrungen mit peerbasierten Elementen fielen unterschiedlich aus. Im Modul «Sozialwissenschaftliche Methoden» wurde Peer-Feedback in vier offenen Rückmeldungen bei insgesamt 31 Befragungsteilnehmenden ausdrücklich als wenig hilfreich oder verzichtbar beschrieben. Auch im Kommunikationsprojekt wurden Peer-Feedback und Selbsteinschätzung teilweise kritisch beurteilt; einzelne offene Rückmeldungen bezeichneten das Peer-Feedback als wenig gewinnbringend und die Selbsteinschätzung als wenig relevant. Im Kernmodul «Medien und Informatik» wurden Austausch und gegenseitige Unterstützung in den Learning Circles dagegen mehrfach positiv hervorgehoben. Die Möglichkeit, Gruppenaufträge ortsunabhängig zu bearbeiten, erleichterte nach einzelnen Rückmeldungen die Zusammenarbeit. Gleichzeitig wurden Schwierigkeiten bei der Terminfindung, unterschiedlich starkes Engagement der Gruppenmitglieder und teilweise unklare Erwartungen an die Gruppenarbeit genannt.

Die Unterschiede weisen darauf hin, dass der wahrgenommene Nutzen von Peer-Feedback und kooperativen Elementen weniger von ihrer digitalen Umsetzung an sich als von ihrer Funktion im Lern- und Prüfungsprozess, ihrer Verbindlichkeit und der organisatorischen Rahmung abhängt. Klare Erwartungen an Beteiligung und Zielsetzung der Zusammenarbeit erscheinen dabei besonders relevant.

3.3 Erfahrungen für die Weiterentwicklung nutzen

Die erste Umsetzung war in mehreren Pilotprojekten Ausgangspunkt weiterer Anpassungen. Verändert wurden unter anderem Umfang und Struktur von Aufgaben, die technische Bereitstellung von Materialien sowie die eingesetzten digitalen Werkzeuge. Einzelne digitale Elemente wurden vereinfacht oder – wie die mündliche Online-Prüfung – nicht weitergeführt. Die Weiterentwicklung bestand somit nicht zwangsläufig in einer stärkeren Digitalisierung, sondern in einer erneuten Klärung der Frage, welche Leistung sichtbar werden soll und welches Setting sie angemessen zugänglich macht. Das folgende Fallbeispiel vertieft, wie Medium, Beurteilungslogik und Assessmentfunktion zusammenspielen.

4 Fallbeispiel: Videobasierte Beurteilung tänzerischer Performanzen mit Adaptive Comparative Judgment

Ein Pilotprojekt aus dem Bereich «Bewegung und Sport» verdeutlicht, dass für die Weiterentwicklung einer Leistungsüberprüfung neben der technischen Umsetzbarkeit auch die mediale Repräsentation des Prüfungsgegenstands, die Beurteilungslogik und

die Assessmentfunktion berücksichtigt werden müssen. Erprobt wurde, ob eine bestehende praktische Tanzprüfung durch eine videobasierte Beurteilung mit ACJ weiterentwickelt werden könnte. Die Untersuchungsanlage und die Ergebnisse wurden bereits ausführlich publiziert (Jeisy, 2025). Hier werden die für die Gestaltung digitaler Leistungsüberprüfungen zentralen Befunde zusammengefasst.

4.1 Ausgangslage und Erprobung

Beim bestehenden Kompetenznachweis «Tanz» präsentieren Studierende eine selbstständig zusammengestellte Choreografie in einer Sporthalle, die von jeweils zwei Dozierenden anhand eines Kriterienrasters analytisch beurteilt wird. Kriterienbasierte Beurteilungen sind in der sportbezogenen Lehrpersonenbildung etabliert (Dyson & Williams, 2012). Als alternative Beurteilungslogik wurde ACJ erprobt. Dabei werden jeweils zwei Performanzen holistisch miteinander verglichen, worauf die jeweils bessere ausgewählt wird. Aus wiederholten Paarvergleichen wird algorithmisch eine Rangordnung geschätzt. In der Übersicht von Bartholomew und Jones (2022) finden sich keine Studien zur Beurteilung mittels ACJ in der sportbezogenen Lehrpersonenbildung.

In vier Prüfungsgelegenheiten wurden die Tanzperformanzen zunächst regulär vor Ort beurteilt. 55 Studierende erklärten sich schriftlich damit einverstanden, dass ihre Performanz zusätzlich videografiert und für Forschungszwecke ausgewertet wurde. Rund zwei Wochen später beurteilten dieselben fünf Dozierenden die videografierten Performanzen mit dem ACJ-Tool «Comproved». Anschliessend reflektierten sie in einer Fokusgruppe die Handhabung des Verfahrens und seine Eignung als Ersatz für den bestehenden summativen Kompetenznachweis. Ergänzend wurden die Reliabilität beider Beurteilungsverfahren und die Übereinstimmung der resultierenden Rangordnungen untersucht (Jeisy, 2025).

4.2 Ergebnisse und Konsequenzen für die Prüfungsgestaltung

Für beide Verfahren ergaben sich hohe methodenspezifische Reliabilitätskennwerte. Beim kriterienbasierten Beurteilen lagen die sitzungsspezifischen Intraklassenkorrelationen (ICC) für den Mittelwert der jeweils zwei Dozierendenurteile zwischen .928 und .983. Für das ACJ-Verfahren wurde eine Scale Separation Reliability (SSR) von .83 ermittelt. Da beide Kennwerte auf unterschiedlichen Messmodellen beruhen, sind sie nicht unmittelbar vergleichbar. Die mit den beiden Verfahren erzeugten Rangordnungen korrelierten stark (Spearman's $\rho = .818$, $p < .001$, 95%-KI [.706, .890]; Jeisy, 2025). Beide Verfahren ermöglichten damit eine reliable Differenzierung der Performanzen und führten zu stark übereinstimmenden Rangordnungen. Der Konvergenzbefund ist jedoch vorsichtig zu interpretieren. Dieselben fünf Dozierenden beurteilten dieselben Performanzen in beiden Verfahren im Abstand von rund zwei Wochen. Erinnerungs- und Reihenfolgeeffekte sowie die gemeinsame Beurteiler:innenvarianz können daher nicht ausgeschlossen werden und könnten zur Übereinstimmung der Rangordnungen beigetragen haben.

Für die Weiterentwicklung des Kompetenznachweises waren diese Befunde jedoch nicht allein ausschlaggebend. Die Dozierenden beschrieben das ACJ nach kurzer Eingewöhnung als technisch gut handhabbar, beurteilten jedoch die Videorepräsentation der Tanzperformanzen kritisch. Insbesondere die dreidimensionale Raumnutzung sowie das «Energetische» und die unmittelbare Wirkung einer Performanz seien auf Video nur eingeschränkt erfassbar. Vorbehalte bestanden zudem gegenüber der vergleichenden Beurteilungslogik. Für den summativen Kompetenznachweis erschien mehreren Dozierenden die Beurteilung, ob definierte Anforderungen erfüllt werden, geeigneter als der unmittelbare Vergleich zweier Performanzen. Sie entschieden sich deshalb gegen einen Ersatz der bestehenden Vor-Ort-Prüfung durch videobasiertes ACJ. Potenzial sahen sie dagegen für formative Anwendungen, etwa für vergleichendes Beurteilen im Lernprozess oder videobasiertes Feedback.

Bei der Einordnung der Ergebnisse ist zudem zu berücksichtigen, dass in der Erprobung gleichzeitig das Medium – Livebeobachtung versus Video – und die Beurteilungslogik – analytisch-kriterienorientiert versus holistisch-vergleichend – verändert wurden. Die wahrgenommenen Unterschiede lassen sich deshalb nicht eindeutig einem der beiden Merkmale zuordnen. Auch die beschriebenen Einschränkungen der Videodarstellung sind als Erfahrungen der Beurteilenden zu verstehen und nicht als Nachweis einer geringeren Beurteilungsqualität.

Das Fallbeispiel zeigt, dass technische Praktikabilität und gute Messkennwerte für den Einsatz eines digitalen Assessmentverfahrens allein nicht ausreichen. Im vorliegenden Fall sprachen die wahrgenommenen Grenzen der medialen Repräsentation und die Vorbehalte gegenüber der vergleichenden Beurteilungslogik beim ACJ gegen einen Ersatz der bestehenden Prüfung.

5 Folgerungen für die Gestaltung digitaler Leistungsüberprüfungen

Die Pilotierungen zeigen, dass sich die Eignung digitaler Leistungsüberprüfungen nicht am Prüfungsformat allein festmachen lässt. Relevant ist vielmehr, ob die gewählte Lösung den fachlichen und didaktischen Anforderungen der Leistungsüberprüfung entspricht und unter den gegebenen medialen, technischen und organisatorischen Bedingungen tragfähig ist. Die folgenden Abschnitte bündeln diese Erfahrungen zu Gestaltungsprinzipien, einem Orientierungsrahmen und konkreten Hinweisen für die Umsetzung.

5.1 Passung als Leitprinzip

Ausgangspunkt einer Leistungsüberprüfung sollten die angestrebten Kompetenzen sein. Im Sinne des Constructive Alignment ist zu bestimmen, welche Leistungen Studierende zeigen müssen und wie diese mit den Lehr-Lern-Aktivitäten und dem Assessment abgestimmt werden (Biggs, 1996; Biggs & Tang, 2011). Zugleich ist zu

berücksichtigen, ob eine Leistungsbeurteilung diagnostischen, formativen oder summativen Zwecken dient (Zimmermann, 2024). Erst darauf bezogen stellt sich die Frage nach der geeigneten medialen und digitalen Umsetzung. Digitalisierung erweitert zwar die Gestaltungsmöglichkeiten, verändert aber zugleich die Bedingungen, unter denen Leistungen sichtbar und beurteilt werden. Dies betrifft sowohl die mediale Repräsentation des Prüfungsgegenstands (Kulgemeyer et al., 2023) als auch die konkrete Ausgestaltung einzelner Formate. So hängt der Nutzen von Portfolios wesentlich von Aufgabenstellung, Strukturierung und Begleitung (Feder & Cramer, 2025) und derjenige von Peer-Assessment von seiner didaktischen und organisatorischen Einbettung ab (Fleckney, Thompson & Vaz-Serra, 2025). Auch Infrastruktur, technische Zuverlässigkeit und Support sind Teil der Gestaltung digitaler Prüfungssettings (Budde et al., 2023).

5.2 Orientierungsrahmen für die Entwicklung digitaler Leistungsüberprüfungen

Aus der theoretischen Einordnung und den Erfahrungen der Pilotierungen lassen sich fünf aufeinander bezogene Fragen ableiten:

1. Welche Kompetenz soll sichtbar werden?
Zu klären ist, welche Leistung Studierende zeigen sollen und welche Prüfungsanforderung dafür geeignet ist. Handlungsnähe bedeutet dabei nicht möglichst grosse Realitätsähnlichkeit, sondern eine angemessene Repräsentation der angestrebten Kompetenz.

2. Welche Funktion hat die Leistungsbeurteilung?
Ein Verfahren kann für formative Zwecke geeignet sein, ohne deshalb auch für eine summative Prüfung zu passen. Dies zeigte sich beispielsweise im ACJ-Pilotprojekt.

3. Welche Aufgabe und welche Beurteilungsform sind geeignet?
Die Aufgabe sollte die angestrebte Leistung zugänglich machen und deren berufliche Relevanz erkennen lassen. Gleichzeitig ist zu bestimmen, nach welcher Logik die Leistung beurteilt wird – beispielsweise kriterienorientiert oder vergleichend – und wer die Beurteilung vornimmt, etwa Dozierende, Peers oder die Studierenden selbst.

4. Welchen Beitrag leisten Medium und digitale Umsetzung?
Zu prüfen ist, welche zusätzlichen Möglichkeiten eine digitale Umsetzung eröffnet, welche Merkmale der Leistung dadurch besonders zugänglich werden und welche Merkmale möglicherweise weniger gut erfasst werden können. Ebenso ist zu berücksichtigen, welche zusätzlichen Anforderungen mit der medialen und technischen Umsetzung verbunden sind.

5. Unter welchen Bedingungen lässt sich das Setting verlässlich umsetzen?
Infrastruktur, Usability, Workload, technische Voraussetzungen und Support sind bei der Planung mitzuberücksichtigen. Die Pilotierungen zeigen zudem, dass sich die Tragfähigkeit eines Settings häufig erst bei der Durchführung vollständig beurteilen lässt.

Abbildung 1 bündelt die fünf Fragen zu einem visuellen Orientierungsrahmen. Dieser ist nicht als Entscheidungsbaum zu verstehen; Erprobung und Weiterentwicklung können Anpassungen auf mehreren Ebenen erforderlich machen.

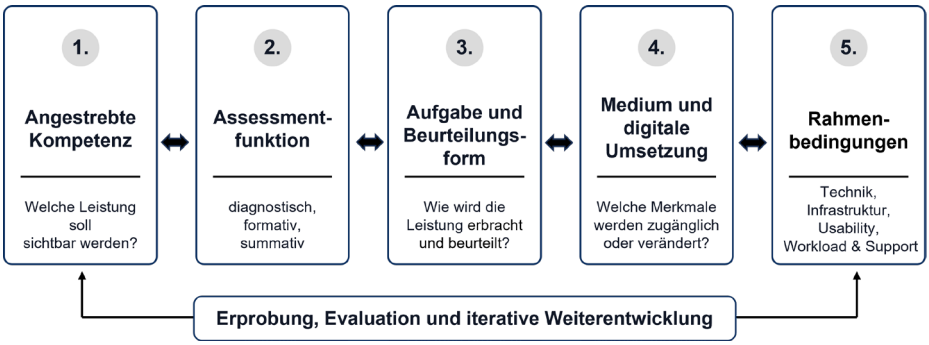


Abbildung 1: Orientierungsrahmen für die Gestaltung digitaler Leistungsüberprüfungen.

5.3 Hinweise für die konkrete Umsetzung

Aus den Pilotierungen ergeben sich zusätzlich einige konkrete Hinweise. Bei digitalen Vor-Ort-Prüfungen sollten nicht nur die technische Funktionsfähigkeit, sondern auch Darstellung und Bedienbarkeit der Aufgaben aus Studierendenperspektive getestet werden. Beim Einsatz eines Safe Exam Browsers erwiesen sich frühzeitige Informationen, Funktionstests und technischer Support als hilfreich. Bei Bring-your-own-device-Modellen sollten Mindestanforderungen an die verwendeten Geräte transparent sein. Bei digitalen Prüfungssettings sollte zudem transparent gemacht werden, welche personenbezogenen Daten erhoben oder gespeichert werden, zu welchem Zweck sie verwendet werden und wie mit ihnen nach Abschluss der Leistungsüberprüfung umgegangen wird.

Bei videobasierten Aufgaben sind sowohl die inhaltliche Eignung des Mediums als auch dessen technische Bereitstellung zu prüfen. Probleme mit Dateigrössen oder Netzwerkbelastung können die Bearbeitung beeinträchtigen. Zugleich ist zu berücksichtigen, welche Merkmale des Prüfungsgegenstands in der medialen Darstellung erhalten bleiben. Zeitraumbezogene Settings wie Lernpfade verbinden Flexibilität mit einem erhöhten Bedarf an Orientierung. Klare Arbeitsaufträge, transparente Erwartungen, übersichtliche Abläufe und verlässlich funktionierende Freischaltungen erleichtern

die Bearbeitung. Bei Portfolioformaten sollten Umfang und Sinnhaftigkeit einzelner Dokumentations- und Reflexionsaufträge sowie ein realistischer Workload geprüft werden. Gestaltungsspielräume bei der Wahl technischer Werkzeuge sind mit den Voraussetzungen der jeweiligen Zielgruppe abzustimmen.

Schliesslich benötigt die erstmalige Entwicklung digitaler Prüfungssettings häufig zusätzliche zeitliche und technische Ressourcen. Beratungs- und Supportangebote sowie ausreichende Testphasen können diesen Prozess unterstützen. Zugleich lassen sich einmal entwickelte Strukturen wiederverwenden und anpassen. Digitalisierung führt daher nicht zwangsläufig zu einer Reduktion des Aufwands, sondern kann diesen von Durchführung und Korrektur hin zu Entwicklung, Vorbereitung und Begleitung verschieben.

5.4 Grenzen und Ausblick

Die berichteten Erfahrungen sind vor dem Entstehungskontext des Projekts einzuordnen. Die Initiative war als Hochschulentwicklungsprojekt und nicht als Forschungsprojekt im engeren Sinne angelegt. Die elf Pilotierungen unterschieden sich in Bezug auf Studiengang, Fach, Assessmentfunktion und konkrete Umsetzung. Diese Breite ermöglichte Einblicke in unterschiedliche Prüfungskontexte, begrenzt zugleich aber die Spezifität projektübergreifender Schlussfolgerungen. Auch die Studierendenbefragungen enthielten neben einem gemeinsamen Kern projektspezifische Fragen. Die Ergebnisse erlauben daher keine Aussagen darüber, welches digitale Prüfungssetting generell geeigneter oder wirksamer ist.

Die Evaluation erfasste überwiegend Wahrnehmungen und Erfahrungen von Studierenden und Dozierenden. Auswirkungen der unterschiedlichen Settings auf den Kompetenzerwerb oder die Qualität von Kompetenzurteilen wurden – mit Ausnahme der spezifischen Analysen im ACJ-Pilotprojekt – nicht untersucht. Hinzu kommen Einschränkungen durch den institutionellen Kontext, die Selbstselektion der beteiligten Dozierenden und die unterschiedlichen Rücklaufquoten der Studierendenbefragungen. Die hier formulierten Folgerungen sind deshalb als Orientierungs- und Reflexionsangebote für die Entwicklung digitaler Leistungsüberprüfungen zu verstehen, nicht als empirisch abgesicherte Wirksamkeitsaussagen.

Für die weitere Entwicklung erscheint insbesondere die Frage relevant, wie unterschiedliche Assessmentdesigns dazu beitragen, die angestrebten Kompetenzen von Studierenden angemessen sichtbar und beurteilbar zu machen. Durch generative KI gewinnt diese Frage zusätzlich an Bedeutung, da sich die Bedingungen, unter denen Leistungen erbracht und zugerechnet werden können, weiter verändern (Lodge, Howard, Bearman, Dawson & Associates, 2023; Budde, Eichhorn & Tobor, 2024). Unabhängig von einzelnen Technologien bleibt damit die zentrale Aufgabe, Leistungsüberprüfungen so zu gestalten, dass die angestrebten Leistungen angemessen sichtbar und beurteilbar werden.

Literatur

Albu, C. & Lindmeier, A. (2023). Performance assessment in teacher education research – A scoping review of characteristics of assessment instruments in the DACH region. *Zeitschrift für Erziehungswissenschaft, 26* (3), 751–778.
Bandtel, M., Baume, M., Brinkmann, E., Bedenlier, S., Budde, J., Eugster, B., Ghoneim, A., Halbherr, T., Persike, M., Rampelt, F., Reinmann, G., Sari, Z. & Schulz, A. (2021). *Digitale Prüfungen in der Hochschule. Whitepaper einer Community Working Group aus Deutschland, Österreich und der Schweiz.* Berlin: Hochschulforum Digitalisierung.
Bartholomew, S. R. & Jones, M. D. (2022). A systematized review of research with adaptive comparative judgment (ACJ) in higher education. *International Journal of Technology and Design Education, 32* (2), 1159–1190.
Bauer-Klebl, A. & Nüesch, C. (2022). Prüfungsevaluationen und damit verbundene Feedbackgespräche: Wie sie zur Hochschulentwicklung beitragen können. *Journal für LehrerInnenbildung, 22* (1), 66–76.
Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32* (3), 347–364.
Biggs, J. & Tang, C. (2011). *Teaching for quality learning at university* (4. Auflage). Maidenhead: Open University Press.
Blömeke, S., Gustafsson, J.-E. & Shavelson, R. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie, 223* (1), 3–13.
Budde, J., Eichhorn, J. & Tobor, J. (2024). *Vision einer neuen Prüfungskultur* (Diskussionspapier Nr. 28). Berlin: Hochschulforum Digitalisierung.
Budde, J., Tobor, J. & Beyermann, J. (2023). *Blickpunkt Digitale Prüfungen.* Berlin: Hochschulforum Digitalisierung.
Dyson, B. P. & Williams, L. H. (2012). The role of PE Metrics in physical education teacher education. *Journal of Physical Education, Recreation & Dance, 83* (5), 29–32.
Feder, L. & Cramer, C. (2025). Research on portfolios in teacher education: A systematic review. *European Journal of Teacher Education, 48* (3), 508–525.
Fleckney, P., Thompson, J. & Vaz-Serra, P. (2025). Designing effective peer assessment processes in higher education: A systematic review. *Higher Education Research & Development, 44* (2), 386–401.
Hu, A., Liu, Q. & Daniel, B. (2025). Digital technologies in authentic assessment in higher education: A systematic literature review and narrative synthesis. *Sage Open, 15* (3), 1–27.
Jeisy, E. (2025). Wer ist besser? Videobasiertes Adaptive Comparative Judgment in der Ausbildung Sport unterrichtender Lehrpersonen. In K. Bäuerlein, S. Bietenhard, C. Conk, M. Longhitano & S. Mahler (Hrsg.), *Unterrichtsvideos in der Schweizer Lehrpersonenbildung* (S. 61–73). Bern: hep.
Kulgemeyer, C., Riese, J., Vogelsang, C., Buschhüter, D., Borowski, A., Weißbach, A., Jordans, M., Reinhold, P. & Schecker, H. (2023). How authenticity impacts validity: Developing a model of teacher education assessment and exploring the effects of the digitisation of assessment methods. *Zeitschrift für Erziehungswissenschaft, 26* (3), 601–625.
Lodge, J. M., Howard, S., Bearman, M., Dawson, P. & Associates. (2023). *Assessment reform for the age of artificial intelligence.* Melbourne: Tertiary Education Quality and Standards Agency.
Metzger, C. & Nüesch, C. (2004). *Fair prüfen. Ein Qualitätsleitfaden für Prüfende an Hochschulen.* St.Gallen: Institut für Wirtschaftspädagogik der Universität St.Gallen.
Pädagogische Hochschule St. Gallen. (2023). *Allgemeines Prüfungsreglement.* St.Gallen: Pädagogische Hochschule St.Gallen.
Reinmann, G. (2021). Prüfungstypen, -formate, -formen oder -szenarien? *Impact Free, 36,* 1–6.
Schaper, N. (2012). *Fachgutachten zur Kompetenzorientierung in Studium und Lehre.* Bonn: Hochschulrektorenkonferenz.
Wahl, D. (2013). *Lernumgebungen erfolgreich gestalten* (3. Auflage). Bad Heilbrunn: Klinkhardt.
Zimmermann, T. (2024). *Leistungsbeurteilungen an Hochschulen lernförderlich gestalten. Prüfen, Beurteilen und Rückmelden von Lernleistungen.* Opladen: Budrich.

Autor und Autorin

Eric Jeisy, Dr., Pädagogische Hochschule St.Gallen, Institut Bewegung, Sport und Gesundheit, eric.jeisy@phsg.ch
Charlotte Nüesch, Prof. Dr., Pädagogische Hochschule St.Gallen, Institut Digitale und Informatische Bildung, charlotte.nueesch@phsg.ch