The University of Tehran Press

Journal of Foreign Language Research

Online ISSN: 2588-7521

Journal Homepage: https://jflr.ut.ac.ir/

Exploring the Effectiveness of Online Assessment Systems in Higher Education in Iran and Germany: A Qualitative Study

Sonya Anvar[1]✉ Mohammadreza Dousti Zadeh [2]. Stefan Piasecki [3]

1. Department of German Language and Literature, Faculty of Literature and Humanities Shahid Beheshti University, Tehran, Iran, Email: s_anvar@sbu.ac.ir

2- Department of German Language and Literature, Faculty of Foreign Languages, University of Tehran, Tehran, Iran. Email: dostizadeh@ut.ac.ir

3-Professor at the North Rhine-Westphalia Police and Public Administration University (HSPV NRW). Email: stefan.piasecki@hspv.nrw.de

Article Info

Article type:
Research Article

Article history:
Received: 25 August 2025
Received in revised form: 14 November 2025
Accepted: 15 November 2025
Available online : 22 November 2026

Keywords:
online assessment, higher education, qualitative research, lived experience, Iran and Germany, validity, e-learning

Extended Abstract

Introduction

The growing integration of digital technologies has transformed online assessment from an alternative practice into an important component of higher education. Despite its advantages, including flexibility, accessibility, and cost reduction, online assessment raises concerns about validity, academic integrity, educational equity, privacy, and the quality of learning experiences. Previous research has examined technological and security dimensions extensively, while comparatively less attention has been devoted to instructors' lived experiences and cross-cultural perspectives (Butler-Henderson & Crawford, 2020; Dashtestani & Karami, 2019). The theoretical foundation of this study draws on Messick's (1995) unified approach to validity, Kane's (2013) framework for validating interpretations and uses of test scores, and Biggs and Tang's (2011) concept of constructive alignment. These perspectives emphasize that assessment effectiveness depends not merely on technological functionality but also on the validity of score interpretations, alignment with educational objectives, academic integrity, and equitable participation. Previous studies have also highlighted the importance of infrastructure, digital readiness, and feedback in online assessment (Davis et al., 2023; Heil & Ifenthaler, 2023). Despite developments in Iran and Germany, comparative qualitative evidence concerning university instructors' experiences remains limited. This study addresses this gap by examining the perceived effectiveness, challenges, and opportunities of online assessment in both countries and identifying practical strategies for improving its design and implementation.

Methodology

The study adopted a qualitative phenomenological approach to investigate the lived experiences of university instructors involved in online assessment. Participants comprised 13 university faculty members, including seven from Iran (53.85%) and six from Germany (46.15%), with experience in both online and face-to-face teaching and involvement in assessment design or implementation. Data collection took place between 2024 and 2025. In Iran, semi-structured face-to-face interviews were conducted at the University of Tehran, Allameh Tabataba'i University, and Shahid Beheshti University. Interviews were recorded with participants' consent and transcribed verbatim. In Germany, semi-structured questions were distributed to instructors, who submitted written responses and, in some cases, audio recordings through email or WhatsApp. The collected textual and audio data were analysed using MAXQDA 2022 and thematic analysis based on Braun and Clarke (2006). The analytical process involved repeated data familiarisation, initial coding, categorisation, identification of subthemes, and

development of overarching themes. Trustworthiness was supported through participant review, parallel coding by two independent researchers, contextual description, documentation of research procedures, and reflexive notes, following Lincoln and Guba (1985).

Findings

The findings identified three overarching themes: challenges, opportunities, and strategies for improving online assessment. The first theme concerned challenges. Examination security and academic dishonesty were prominent concerns, particularly regarding the authenticity of test-takers, unauthorised access to examination materials, and insufficient remote monitoring. Technical difficulties, including internet interruptions, platform instability, and inadequate equipment, were more pronounced in Iran. In both countries, instructors reported reduced opportunities for individualised feedback and weaker pedagogical interaction compared with face-to-face assessment. Concerns about unequal access, technological anxiety, and institutional trust were also identified. The second theme concerned opportunities. Participants emphasised temporal and spatial flexibility, reduced institutional and individual costs, large-scale assessment, and the continuity of educational activities during crises. Other benefits included broader accessibility, inclusion of students with physical limitations, digital literacy development, data storage for subsequent analysis, and opportunities for future blended learning. The third theme involved strategies for improving effectiveness. Participants recommended open-ended and project-based examinations, continuous assessment, hybrid assessment models, and multi-layered evaluation involving both instructors and software. They also emphasised investment in secure local platforms, technological infrastructure, faculty training, student digital literacy, academic integrity guidelines, institutional coordination, and the localisation of international practices. Comparative analysis indicated that security and technical challenges were shared by both countries but were more pronounced in Iran because of weaker technological infrastructure and lower digital readiness. Conversely, flexibility, cost efficiency, and opportunities for broader assessment were common advantages.

Discussion and Conclusion

The findings demonstrate that the effectiveness of online assessment cannot be evaluated solely through technological functionality. Rather, it depends on the interaction between assessment design, educational objectives, technological infrastructure, academic integrity, equity, and feedback. The identified security challenges reinforce the importance of innovative assessment approaches, particularly open-ended and project-based examinations, as discussed by Crisp and Ward (2022) and Joughin (2010). The comparative findings also highlight the importance of contextual conditions. While technological infrastructure and unequal access intensified challenges in Iran, both countries experienced similar benefits concerning flexibility, cost reduction, and assessment scalability. These findings emphasise the relevance of consequential validity (Messick, 1995; Kane, 2013), particularly regarding the relationship between assessment practices, institutional conditions, and educational opportunities. Three levels of improvement emerged. At the policy level, investment in secure technological infrastructure, local platforms, and institutional coordination is necessary. At the institutional level, stronger security procedures, multi-layered evaluation, and cooperation among instructors can improve reliability and trust. At the educational level, faculty development, student digital literacy, continuous assessment, and innovative examination design are essential. Overall, the study concludes that sustainable online assessment requires simultaneous improvements in technological infrastructure, institutional policies, and educational practices. By integrating validity, equity, academic integrity, and effective feedback, higher education institutions can move beyond emergency-

based online assessment towards a more reliable and sustainable assessment system.

References

Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (4th ed.). Open University Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. Computers & Education, 159, 104024. https://doi.org/10.1016/j.compedu.2020.104024

Crisp, G., & Ward, C. (2022). Re-imagining assessment in higher education: Digital, authentic, and inclusive practices. Assessment & Evaluation in Higher Education, 47(4), 565–580. https://doi.org/10.1080/02602938.2021.1887812

Dashtestani, R., & Karami, H. (2019). Validity and reliability issues in online assessment: A review of literature. Asian EFL Journal, 21(3), 85–110. Google Scholar

Davis, V. W., Smith, L. T., & Chan, M. (2023). Privacy, equity, and validity in online testing: A framework for higher education. Assessment in Education: Principles, Policy & Practice, 30(1), 118–134. https://doi.org/10.1080/0969594X.2022.2094567

Heil, C., & Ifenthaler, D. (2023). Quality and acceptance of online assessment in German universities: Policies and practices. British Journal of Educational Technology, 54(1), 13–29. https://doi.org/10.1111/bjet.13224

Joughin, G. (2010). The hidden curriculum revisited: A critical review of research into the influence of summative assessment on learning. Assessment & Evaluation in Higher Education, 35(3), 335–345. https://doi.org/10.1080/02602930903221493

Kane, M. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1–73. https://doi.org/10.1111/jedm.12000

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. SAGE Publications.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Cite this article: Anwar, S. Dousti Zadeh , M, & Piasecki, S. (2026). Exploring the Effectiveness of Online Assessment Systems in Higher Education in Iran and Germany: A Qualitative Study. *Journal of Foreign Language Research*, *16*(4), 459-476. https://doi.org/ 10.22059/jflr.2025.401256.1240

© Author(s) retain the copyright. **Publisher**: University of Tehran Press.
DOI: https://doi.org/ 10.22059/jflr.2025.401256.1240

1. Einleitung

Die rapide Entwicklung digitaler Lehr- und Lerntechnologien hat Online-Prüfungssysteme in den vergangenen Jahren von einer ergänzenden Option zu einem zentralen Element hochschulischer Bewertung transformiert. Obwohl dieses Format Vorteile wie erhöhte Flexibilität, breiteren Zugang und geringere Kosten bietet, bleiben grundlegende Fragen hinsichtlich Validität, Authentizität von Antworten, Bildungsgerechtigkeit, Datenschutz und Nutzererfahrung bestehen (Butler-Henderson & Crawford, 2020; Heil & Ifenthaler, 2023). Neuere qualitative Studien zeigen zudem, dass selbst im Bereich des Fremdsprachenlernens die

Online-Durchführung mit erhöhter kognitiver Belastung sowie spezifischen Anforderungen an Validität und Lernqualität verbunden sein kann (Mones Gharamalki, Ahangari & Hadidi Tamjid, 2024). Solche Befunde verdeutlichen, dass die bloße Übertragung traditioneller Prüfungsformen in digitale Umgebungen unzureichend ist und dass die Gestaltung sowie Validierung onlinebasierter Prüfungsformate unter Berücksichtigung der gelebten Erfahrungen von Lehrenden und Studierenden neu gedacht werden müssen. Die Bedeutung digitaler Prüfungen ergibt sich nicht nur aus globalen Krisen wie Pandemien oder geografischen Restriktionen, sondern ebenso aus Kontexten wie Kriegen, politischen Spannungen oder finanziellen Engpässen, in denen Online-Formate häufig die einzige Möglichkeit zur Aufrechterhaltung von Lehr-Lern-Prozessen darstellen. Gleichwohl konzentriert sich ein Großteil der bisherigen Forschung primär auf technologische oder sicherheitsbezogene Aspekte, während qualitative, erfahrungsorientierte Perspektiven bisher deutlich seltener berücksichtigt wurden (Dashtestani & Karimi, 2019). Diese Erkenntnisse verdeutlichen, zusammen mit internationalen Untersuchungen, eine signifikante Forschungslücke hinsichtlich der Erfahrungen von Hochschullehrenden mit Online-Assessments. Aus theoretischer Sicht ist Validität keine Eigenschaft des Prüfungsinstruments an sich, sondern der auf Basis der erzielten Ergebnisse vorgenommenen Interpretationen und Entscheidungen. Sie erfordert eine kohärente Kette argumentativer Begründungen und empirischer Evidenzen (Messick, 1995: 741; Kane, 2013: 2). Im Kontext digitaler Prüfungen muss diese Argumentationskette nicht nur die Angleichung an Lernziele und das constructive alignment berücksichtigen (Biggs & Tang, 2011: 45–47), sondern auch Aspekte wie akademische Integrität, Fairness und Datenschutz (Davies et al., 2023: 119; Mutimucuho et al., 2020: 64–65). Dennoch existieren bislang nur wenige Studien, die transnationale Erfahrungen von Hochschullehrenden vergleichen und ihre Wahrnehmungen zur Wirksamkeit digitaler Prüfungssysteme in unterschiedlichen kulturellen Kontexten analysieren (Heil & Ifenthaler, 2023: 14–15). Vor diesem Hintergrund untersucht die vorliegende Studie die Wirksamkeit, Herausforderungen und Potenziale onlinebasierter Prüfungen aus Sicht von Hochschullehrenden in zwei kulturell und institutionell unterschiedlichen Kontexten: Iran und Deutschland. Das Ziel besteht darin, zentrale Dimensionen der wahrgenommenen Wirksamkeit – insbesondere Lernorientierung, Feedback, Fairness und constructive alignment – zu identifizieren und einzuordnen. Zugleich sollen länderspezifische Unterschiede und Gemeinsamkeiten herausgearbeitet werden. Ein weiteres Ziel ist die Ableitung nutzerzentrierter Strategien zur Verbesserung des Designs und der Durchführung digitaler Prüfungsformate, insbesondere hinsichtlich Faktoren, die Vertrauen, Fairness und Akzeptanz seitens der Studierenden erhöhen können. Daraus ergibt sich die leitende Frage der Untersuchung: Wie konzeptualisieren iranische und deutsche

Hochschullehrende die Wirksamkeit digitaler Prüfungen, welche Herausforderungen und Schlüsselfaktoren dominieren jeweils, und welche Gemeinsamkeiten bzw. Unterschiede treten in ihren Erfahrungen hervor? Die Studie basiert auf der Annahme, dass Transparenz der Bewertungskriterien und Authentizität der Aufgaben zur Erhöhung der Validität digitaler Prüfungen beitragen und dass Datenschutz sowie Fairness grundlegende Voraussetzungen für die Akzeptanz proktorierter Online-Assessments darstellen. Im Folgenden werden zunächst die theoretischen Grundlagen und die einschlägige Literatur diskutiert. Anschließend werden Methodik, Datenerhebung und Auswertungsansatz beschrieben, bevor die Ergebnisse entlang der identifizierten Hauptthemen dargestellt und anschließend diskutiert werden. Den Abschluss bilden Schlussfolgerungen, praktische und theoretische Implikationen sowie Limitationen und Empfehlungen für zukünftige Forschung.

2. Forschungsstand

Die Diskussion über Online-Assessment in der Hochschulbildung hat sich im vergangenen Jahrzehnt umfangreich in der internationalen Literatur entfaltet. Die theoretischen Grundlagen in diesem Bereich betonen insbesondere drei zentrale Dimensionen: erstens technologische Infrastruktur und Anfangsinvestitionen als grundlegende Voraussetzung für Funktionalität; zweitens Validität und Reliabilität als Basis für die Interpretation von Ergebnissen (Messick, 1995: 741; Kane, 2013: 2); und drittens die gelebten Erfahrungen und Einstellungen der relevanten Akteure – Studierende und Lehrende –, die maßgeblich die Akzeptanz und Wirksamkeit digitaler Prüfungsformate beeinflussen (Heil & Ifenthaler, 2023: 14–15). Ein Blick auf Studien aus Iran, Deutschland und anderen Ländern verdeutlicht, dass trotz gemeinsamer Bezugspunkte jeder Kontext mit eigenen Chancen und Herausforderungen konfrontiert ist. Internationale Untersuchungen zwischen 2016 und 2025 heben drei Schlüsselfaktoren für wirksame Online-Assessments hervor: eine leistungsfähige technische Infrastruktur, die Schulung und Vorbereitung der Nutzerinnen und Nutzer sowie transparente Kriterien für das Assessment-Design. Gamag et al. (Gmaj et al., 2022: 14–15) zeigten in einer systematischen Review zu Moodle-Anwendungen, dass Online-Assessments bei ausreichenden technischen Kapazitäten und Supportstrukturen die Beteiligung und Leistungsfähigkeit der Studierenden verbessern können – obwohl Entwicklungsaufwand und technische Schwierigkeiten weiterhin Hemmnisse darstellen. Muzaffar et al. (2020: 7–8) identifizierten Sicherheits- und Integritätsprobleme als globale Herausforderungen, die in vielen Ländern auftreten. Weitere Studien wie jene von Loureiro und Gomes (2022: 233–234) sowie Craig und Kay (2025: 49) haben sich auf Lernerfahrungen und Interaktionen in Online-Umgebungen konzentriert und die Bedeutung von Peer-Matching und Anonymität für die Qualität von Feedback und Motivation herausgearbeitet. Zugleich zeigen Untersuchungen wie Choume

(2025: 88–89) oder Taher et al. (2025: 65–66), dass Online-Prüfungen an der Schnittstelle zwischen Effizienz und Gefährdung akademischer Integrität liegen und bestehende digitale Ungleichheiten verstärken können. In Iran zeichnen Studien ein ergänzendes Bild der Herausforderungen. Azimi et al. (2016: 122–123) stellten in einem Vergleich zwischen Iran und Russland fest, dass Plagiatserkennung und Vortests in Iran weniger berücksichtigt werden, während sie in Russland als grundlegende Elemente für Bildungsgerechtigkeit gelten. Dashtestani (2020: 85) berichtete, dass Studierende zwar positive Einstellungen zum Online-Lernen haben, Lehrende und Hochschulverantwortliche jedoch der Validität solcher Prüfungen skeptisch gegenüberstehen. Azizi (2022: 41–42) betonte die Bedeutung distributiver, prozeduraler und interaktiver Gerechtigkeit als Kernelemente des Online-Assessment-Designs in Iran und identifizierte schwache Internetinfrastruktur sowie geringe digitale Kompetenz als wesentliche Hindernisse. Rahimi et al. (2019: 56) verwiesen zudem auf technische Störungen und wiederkehrende Internetausfälle, die die Validität digitaler Prüfungen beeinträchtigen. In Deutschland zeigen politiknahe Analysen wie jene von Heil und Ifenthaler (2023: 16–18), dass Investitionen in Infrastruktur und die Entwicklung transparenter Bewertungskriterien zentrale Voraussetzungen zur Sicherung der Qualitätsstandards darstellen. Empirisch fundierte Hochschulstudien bleiben jedoch begrenzt. Feld- und technologieorientierte Untersuchungen haben sich daher auf spezifische Instrumente des digitalen Assessments konzentriert. So zeigten Reiss et al. (2009: 3) sowie Größinger und Schnabel (2024: 5), dass Online-Selbsteinschätzungen zur Identifikation studienrelevanter Fähigkeiten und zur Förderung der Selbstreflexion beitragen können. Petri und Kersting (2024: 8) hoben dieselben Instrumente als unterstützende Tools für die Studienorientierung hervor. Studien wie jene von Lehmann et al. (2015: 12) und Bauer et al. (2009: 77) unterstreichen den Nutzen von Peer-Assessment für die Qualität von Feedback und Studienerfolg. Im Bereich automatisierter Prüfungen zeigten Lenhard et al. (2007: 155) mittels Latent-Semantic-Analysis, dass computergestützte Bewertungssysteme eine hohe Übereinstimmung mit menschlicher Bewertung aufweisen, auch wenn Herausforderungen wie geringe Akzeptanz und begrenzte semantische Tiefe bestehen bleiben. Dohmen et al. (2021: 92) sowie Jeremias und Selmer (2019: 101) untersuchten elektronische Prüfungsformate mit ONYX-, OPAL- und MAXIMA-Systemen und bestätigten deren technische Effizienz, verwiesen jedoch zugleich auf Probleme wie Ressourcenverteilung, veränderte Arbeitsbelastung und ethische Bedenken. Darüber hinaus betonten Kohler et al. (2020: 48) und Strickroth et al. (2011: 63) das Potenzial adaptiver und diagnostischer Testverfahren zur präziseren Erfassung von Kompetenzen und zur Unterstützung individueller Lernpfade. Insgesamt zeigen die Studienlandschaften Deutschlands, dass Online-Assessment nicht nur technisch umsetzbar ist, sondern zur Förderung von Selbstreflexion, interaktivem

Lernen und Studierendenzufriedenheit beitragen kann. Zugleich bleiben geringe Akzeptanz bestimmter Technologien, institutionelle Restriktionen und Fragen der Bildungsgerechtigkeit weiterhin kritische Themen (Reiss et al., 2009: 3; Lehmann et al., 2015: 12; Lenhard et al., 2007: 155; Dohmen et al., 2021: 92). Damit bleibt in beiden Ländern – Iran wie Deutschland – die Frage nach Feedbackqualität und Interaktion als zentraler, gemeinsamer Herausforderung bestehen. Die Analyse des Forschungsstands verdeutlicht: (1) Validität und Fairness bilden zwei grundlegende Dimensionen digitaler Prüfungsprozesse; (2) institutionelle und kulturelle Unterschiede prägen die Erfahrungen von Lehrenden in Iran und Deutschland maßgeblich; und (3) es besteht eine deutliche Forschungslücke hinsichtlich vergleichender Studien zwischen beiden Ländern. Die vorliegende qualitative Untersuchung, gestützt auf persönliche Interviews in Iran und offene Fragebögen in Deutschland, soll diese Lücke schließen und ein präziseres Bild der Gemeinsamkeiten und Unterschiede in den gelebten Erfahrungen der Hochschullehrenden beider Länder zeichnen.

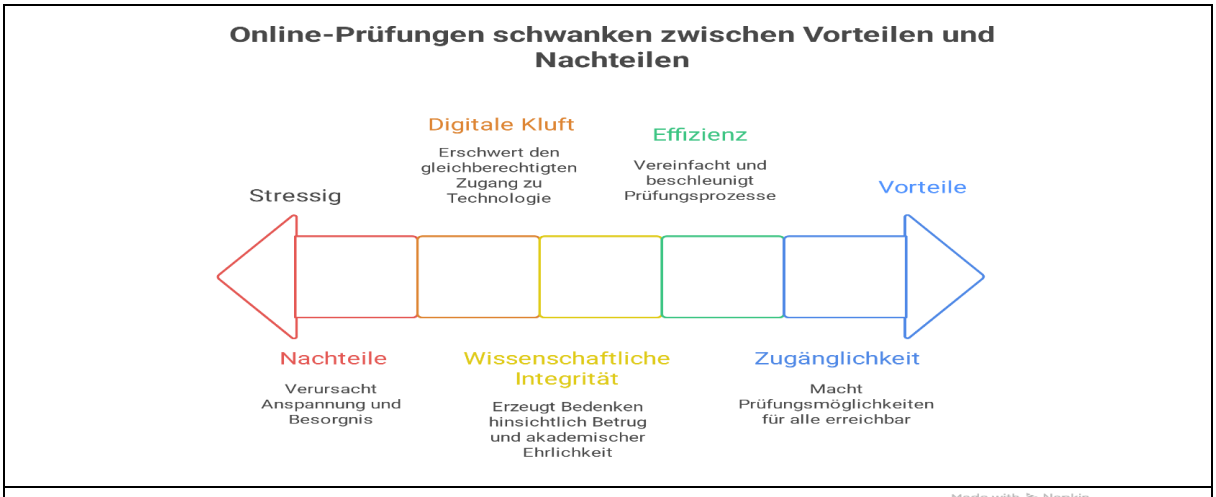


Abbildung 1. Widerstreitende Pole im Online-Assessment der Hochschulbildung: einerseits Zugänglichkeit und Effizienz, andererseits Prüfungsstress, akademische Integrität und digitale Kluft.[1]

3. Methodik

Die vorliegende Untersuchung wurde im Rahmen eines qualitativen Forschungsansatzes konzipiert und durchgeführt, mit dem Ziel, die Wirksamkeit digitaler Prüfungssysteme in der Hochschulbildung Irans und Deutschlands zu analysieren. Die Entscheidung für einen qualitativen Zugang gründet auf der Annahme, dass das Forschungsthema ein tiefes Verständnis der gelebten Erfahrungen von Hochschullehrenden sowie die Exploration der bedeutungstragenden Dimensionen des Phänomens erfordert – und nicht lediglich die quantitative Messung einzelner Variablen (Creswell & Poth, 2018: 42). Die in diesem Kapitel dargestellten statistischen Angaben dienen daher ausschließlich der

[1] Teshome (2025, pp. 88–89) und Tahir et al. (2025, pp. 65–66).

deskriptiven Charakterisierung der Teilnehmenden und wurden nicht für inferenzstatistische Analysen herangezogen. Alle im Folgenden präsentierten Daten entstammen vollständig der empirischen Feldforschung der vorliegenden Studie und wurden durch direkte Interviews mit Hochschullehrenden in Iran und Deutschland erhoben. Die Forschungspopulation umfasste 13 Mitglieder akademischer Fakultäten, die sowohl über Präsenz- als auch Online-Lehrpraxis verfügten und in die Gestaltung oder Durchführung digitaler Prüfungsformate eingebunden waren. In Iran erfolgten die Interviews persönlich in den Arbeitsräumen der Lehrenden an den Universitäten Teheran, Allameh Tabataba'i und Shahid Beheshti. Die Gespräche wurden mit Einverständnis der Teilnehmenden aufgezeichnet und anschließend vollständig transkribiert. In Deutschland wiederum wurden aufgrund geografischer Streuung und zeitlicher Einschränkungen halbstrukturierte Interviews über die Zusendung von Leitfragen durchgeführt; die Lehrenden übermittelten ihre Antworten schriftlich oder in Form von Audioaufnahmen via WhatsApp oder E-Mail. Sämtliche Antworten wurden nach Erhalt vollständig verschriftlicht und zur thematischen Analyse in das Programm MAXQDA importiert. Zur Sicherung der Authentizität und Glaubwürdigkeit der Daten wurde ein Beispiel des transkribierten Interviewleitfadens sowie eine Abbildung des Kodierprozesses im Anhang bereitgestellt, um die Nachvollziehbarkeit des Forschungsweges zu erhöhen. Die demografischen Merkmale der Teilnehmenden zeigten, dass 7 Personen (53,85 %) aus Iran und 6 Personen (46,15 %) aus Deutschland stammten. Neun Teilnehmende (69,23 %) waren männlich und vier (30,77 %) weiblich. Während die deutsche Stichprobe aus drei Männern und drei Frauen bestand, setzte sich die iranische Stichprobe aus einem weiblichen und sechs männlichen Lehrenden zusammen. Bezüglich des Bildungsniveaus verfügten 12 Personen (92,31 %) über eine Promotion, eine Person (7,69 %) über einen Masterabschluss. In Bezug auf die akademische Position verteilten sich die Teilnehmenden wie folgt: sechs Professor*innen, vier Assistant Professors, ein Associate Professor, eine Lehrkraft (Lecturer) und eine Person mit dem Rang Full Professor. Diese Vielfalt in akademischen Positionen und Qualifikationen ermöglichte den Zugang zu einem breiten Spektrum an Perspektiven. Für die Auswertung der Daten wurde die thematische Analyse nach Braun und Clarke (2006: 78) herangezogen. Der Analyseprozess umfasste fünf zentrale Schritte: wiederholtes Lesen der Daten zur vollständigen Vertrautheit, Generierung erster Codes, Gruppierung der Codes zu konzeptionellen Kategorien, Entwicklung von Subthemen und schließlich die Herausarbeitung übergeordneter Hauptthemen. Dieses Verfahren ermöglichte die Identifikation bedeutungstragender Muster und der vielfältigen Dimensionen der Wirksamkeit digitaler Prüfungsformate. Zur Gewährleistung der Validität und Vertrauenswürdigkeit der Befunde wurden die Kriterien von Lincoln und Guba (1985: 289) berücksichtigt. Die Glaubwürdigkeit (credibility) wurde durch Teilnehmendenvalidierung sowie parallele Codierung durch zwei unabhängige Forschende sichergestellt. Die Transferierbarkeit (transferability) wurde durch ausführliche Kontextbeschreibungen der institutionellen und kulturellen Rahmenbedingungen beider

Länder gestärkt. Die Verlässlichkeit (dependability) wurde durch umfassende Dokumentation sämtlicher Forschungsschritte im Feldtagebuch unterstützt; die Bestätigbarkeit (confirmability) wurde durch reflektive Notizen der Forschenden und Peer-Review-Verfahren gewährleistet. Die Durchführung der Studie in zwei kulturell und institutionell unterschiedlichen Kontexten ermöglichte einen systematischen Vergleich der Erfahrungen iranischer und deutscher Hochschullehrender. Die Ergebnisse zeigen, dass neben gemeinsam empfundenen Herausforderungen wie Sicherheitsbedenken, Fairnessfragen und Validitätsrisiken digitaler Prüfungen, auch deutliche kulturelle und strukturelle Unterschiede bestehen. Durch diese komparative Anlage liefert die Studie nicht nur ein vertieftes Verständnis digitaler Prüfungspraktiken in beiden Kontexten, sondern trägt auch zur Weiterentwicklung bildungspolitischer Maßnahmen sowie zur Verbesserung der Qualität digitaler Prüfungsformate im Hochschulbereich bei.

Tabelle 1. Demografische Merkmale der Teilnehmenden*(Quelle: Forschungsdaten, 2024–2025)*

Land	Nr.	Teilnehmencode	Geschlecht	Hochschule	Abschluss & Akademische Position
			Deutschland		
1	T.N1	Weiblich	Fachhochschule Polizei Sachsen-Anhalt		
2	T.N2	Männlich	Uni Duisburg / Uni Leipzig / Dr. phil / Uni Kassel	PhD	Professor-in
3	T.N3		CVJM-Hochschule Kassel / YMCA-University of Applied Sciences Kassel		
4	T.N4	Weiblich	Hochschule für Polizei und öffentliche Verwaltung NRW	PhD	
5	T.N5	Männlich	Universität Paderborn		
6	T.N6	Weiblich	CVJM-Hochschule Kassel / Universität Kassel		
			Iran		
7	T.N7	Männlich	Universität Teheran	PhD	Assistant Professor
8	T.N8				Associate Professor
9	T.N9			Master	Lecturer
10	T.N10		Allameh-Tabataba'i-Universität		Full Professor
11	T.N11			PhD	Assistant Professor
12	T.N12	Weiblich	Shahid-Beheshti-Universität		
13	T.N13	Männlich			

Tabelle 2. Deskriptive Statistik der Teilnehmenden *(Quelle: Forschungsdaten, 2024–2025)*

Variable	Gruppe	Anzahl	Prozent (%)

Land	Iran	7	53,85
	Deutschland	6	46,15
Geschlecht	Männlich	9	69,23
	Weiblich	4	30,77
Land × Geschlecht	Deutschland – weiblich	3	23,08
	Deutschland – männlich	3	23,08
	Iran – weiblich	1	7,69
	Iran – männlich	6	46,15
Bildungsgrad	Promotion (PhD)	12	92,31
	Master	1	7,69
Akademische Position	Professor	6	46,15
	Assistant Professor	4	30,77
	Associate Professor	1	7,69
	Lecturer	1	7,69
	Full Professor	1	7,69
Gesamt	—	13	100,00

4. Ergebnisse und Diskussion

Die Befunde dieser Untersuchung zeigen, dass die Erfahrung mit Online-Assessment in zwei unterschiedlichen kulturellen und bildungsbezogenen Kontexten – nämlich Iran und Deutschland – mit einer Reihe von Herausforderungen, Chancen und vorgeschlagenen Lösungsansätzen verbunden ist. Obwohl viele dieser Aspekte einen globalen Charakter besitzen, variieren ihre Intensität und Ausprägung je nach Entwicklungsstand der digitalen Infrastruktur, institutioneller Kultur und Bildungspolitik in jedem Land.

Herausforderungen des Online-Assessments

Das erste Hauptthema bezog sich auf die Herausforderungen des Online-Assessments. Eine der am häufigsten genannten Sorgen war die Prüfungssecurity und das Risiko des Betrugs. Die Teilnehmenden äußerten wiederholt die Befürchtung, dass unklar bleibt, ob die Person, die eine Online-Prüfung ablegt, tatsächlich der eingeschriebene Studierende ist oder ob jemand anderes an seiner Stelle antwortet. Im iranischen Kontext war diese Sorge aufgrund schwacher Videoüberwachungsinfrastruktur und begrenzter Remote-Control-Systeme noch ausgeprägter. Wie einer der iranischen Teilnehmenden betonte: „Das Problem des Betrugs ist in Iran ein ernstes Anliegen, insbesondere wenn keine ausreichenden Möglichkeiten zur visuellen Überwachung vorhanden sind“ (T.N7, Iran). Auch in Deutschland, obwohl die Überwachungsbedingungen besser sind, besteht weiterhin Zweifel an der Authentizität der Antworten (T.N2, Deutschland). Eine weitere Herausforderung war der Rückgang der Qualität von Interaktion und Feedback. Lehrende in beiden Ländern waren der Ansicht, dass Online-Umgebungen die Möglichkeit für individuelles, tiefgehendes Feedback einschränken und dass die Qualität der Interaktion zwischen Lehrenden und Studierenden in Online-Prüfungen – insbesondere bei offenen Fragen – deutlich geringer ist als im Präsenzunterricht. Eine deutsche Professorin kommentierte

dazu: „Wir können online nicht denselben tiefen und persönlichen Feedbackprozess realisieren wie im Präsenzunterricht“ (T.N1, Deutschland). Darüber hinaus stellten technische Probleme und infrastrukturelle Einschränkungen weitere zentrale Hindernisse dar. In Iran wurden unter anderem Internetabbrüche, unzureichende technische Ausstattung in den Haushalten der Studierenden und Softwareinkompatibilitäten berichtet (T.N9, Iran). Aber auch in Deutschland besteht trotz breiter verfügbarer Infrastruktur das Risiko, dass Plattformen bei großen Prüfungen ausfallen – ein Problem, das als ernsthafte Gefahr eingestuft wurde (T.N5, Deutschland).

Chancen des Online-Assessments

Trotz der genannten Herausforderungen eröffnete Online-Assessment auch bedeutende Chancen für die Hochschulsysteme. Der wichtigste Vorteil war die zeitliche und räumliche Flexibilität. Studierende konnten von jedem Ort und unter verschiedenen Bedingungen an Prüfungen teilnehmen, was besonders für berufstätige Studierende oder solche, die in anderen Städten wohnen, sehr wertvoll war (T.N6, Deutschland; T.N11, Iran). Eine zweite Chance war die Reduktion von Kosten und Ressourceneinsatz. Online-Prüfungen führten zu erheblichen Einsparungen bei Zeit und institutionellen Ausgaben, da keine physischen Räume oder umfangreiches Aufsichtspersonal benötigt wurden. Dieser Vorteil war nicht nur für Hochschulen, sondern auch für die Studierenden relevant, weil Transport- und Aufenthaltskosten entfielen (T.N3, Deutschland; T.N12, Iran). Eine weitere Möglichkeit bestand in der Bewertung großer Gruppen. Hochschulen konnten gleichzeitig Hunderte von Studierenden prüfen – etwas, das im Präsenzbetrieb aus physischen und logistischen Gründen praktisch unmöglich war (T.N4, Deutschland; T.N10, Iran). In Iran wurde diese Eigenschaft besonders während der COVID-19-Pandemie als zentral hervorgehoben.

Strategien zur Verbesserung der Wirksamkeit des Online-Assessments

Das dritte Thema betraf die vorgeschlagenen Strategien zur Verbesserung der Wirksamkeit digitaler Prüfungen. Teilnehmende betonten, dass das Prüfungsdesign stärker nutzerorientiert und hybrid gestaltet werden sollte. Die Neugestaltung von Prüfungen auf Grundlage offener, anwendungsbezogener Aufgaben sowie die Nutzung gemischter Formate (Präsenz + online) könnten mehr Sicherheit und Validität gewährleisten (T.N2, Deutschland; T.N13, Iran). Eine zweite Strategie bezog sich auf die Stärkung technologischer und sicherheitsbezogener Infrastruktur. Der Einsatz internationaler zertifizierter Software, Investitionen in leistungsfähige Server und die Entwicklung sicherer nationaler Plattformen – insbesondere für Iran – wurden als notwendig erachtet (T.N5, Deutschland; T.N7, Iran). Die dritte vorgeschlagene Strategie betraf die Befähigung von Lehrenden und Studierenden. Die Förderung technologischer Kompetenzen, Bewusstseinsbildung und digitale Schulung tragen dazu bei, Sorgen abzubauen und den Lehrenden Kompetenzen für das angemessene Design von Online-Prüfungen zu vermitteln. Wie ein iranischer Teilnehmender bemerkte: „Lehrende müssen lernen, wie

man Online-Prüfungen gestaltet; die bloße Übertragung einer Präsenzprüfung ins Digitale reicht nicht aus“ (T.N8, Iran).

Interpretation der Befunde

Die vergleichende Analyse zeigt, dass Sicherheitsbedenken, reduzierte pädagogische Interaktion und technische Begrenzungen in beiden Ländern präsent sind, ihre Intensität jedoch in Iran stärker ausgeprägt ist. Dies ist vor allem auf schwächere technologische Infrastruktur und geringere digitale Bereitschaft zurückzuführen. Gleichzeitig haben beide Länder ähnliche Vorteile erlebt – Flexibilität, geringere Kosten und die Möglichkeit groß angelegter Prüfungen. Diese Ergebnisse decken sich mit früheren Forschungen, die die Bedeutung innovativen Prüfungsdesigns zur Erhöhung der Validität und Verringerung von Betrugsrisiken hervorheben (Crisp & Ward, 2022; Jaggin, 2010). Basierend auf dem Konzept der consequential validity (Messick, 1995: 741; Kane, 2013: 2) zeigt sich zudem, dass die Wirksamkeit von Online-Assessment stark von bildungspolitischer Unterstützung und Investitionen in ICT-Infrastruktur abhängt. Die vorliegenden Daten, analysiert anhand des thematischen Modells nach Braun & Clarke (2006: 78–79) und der Kriterien nach Lincoln & Guba (1985: 289–331), bestätigen die Robustheit der Ergebnisse.

Tabelle 3. Qualitative Themenanalyse (Raw Data → Initial Codes → Sub-Themes → Main Themes)

Rohdaten (Zitate der Teilnehmenden)	Initiale Codes (Initial Codes)	Sub-Themes (Unterthemen)	Hauptthemen (Main Themes)
„Wir sind immer besorgt, ob der Studierende selbst die Prüfung ablegt oder jemand anderes.“ (T.N2 – Deutschland)	Betrug, Antwortauthentizität (Cheating, Authenticity)	Prüfungssicherheit (Exam Security)	**Herausforderungen des Online-Assessments (Challenges)**
„Im Iran gibt es nicht genügend Infrastruktur für Videoüberwachung.“ (T.N7 – Iran)	Fehlende Proctoring-Infrastruktur (Lack of Proctoring Infrastructure)	Prüfungssicherheit	
„Die Qualität des Feedbacks ist niedriger.“ (T.N8 – Iran)	Unzureichendes Feedback (Low-Quality Feedback)	Rückgang der Feedbackqualität (Feedback Decline)	
„Die Interaktion zwischen Lehrenden und Studierenden ist schwächer.“ (T.N1 – Deutschland)	Geringe Interaktion (Low Interaction)	Reduzierte pädagogische Interaktion (Reduced Pedagogical Interaction)	
„Das Internet wurde während der Prüfung unterbrochen.“ (T.N9 – Iran)	Internetabbrüche, Instabilität (Connectivity Issues)	Technische Probleme (Technical Difficulties)	
„Plattformen stürzen bei großen Prüfungen ab.“ (T.N5 – Deutschland)	Serverabsturz (Server Crash)	Plattforminstabilität (Platform Instability)	
„Studierende haben Zugang zu unerlaubten Materialien.“ (T.N11 – Iran)	Zugriff auf nicht genehmigte Ressourcen (Unauthorized Materials)	Fehlende Ressourcen-Kontrolle (Resource Control Issues)	
„Die Möglichkeiten der Videoüberwachung sind begrenzt.“ (T.N12 – Iran)	Unzureichende Überwachung (Insufficient Monitoring)	Unzureichende Aufsicht (Inadequate Supervision)	

„Multiple-Choice-Fragen können leicht geteilt werden.“ (T.N3 – Deutschland)	Teilen von Fragen (Question Sharing)	Inhaltsanfälligkeit (Content Vulnerability)	
„Studierende machen Screenshots und verbreiten sie.“ (T.N13 – Iran)	Diebstahl von Prüfungsinhalten (Exam Content Theft)	Inhaltssicherheit (Content Security)	
„Im Iran ist Bildungsgerechtigkeit gefährdet, weil nicht alle denselben Zugang haben.“ (T.N10 – Iran)	Ungleicher Zugang (Unequal Access)	Bildungsgerechtigkeit (Educational Justice)	
„Studierende mit schwachem Internet geraten unter Stress.“ (T.N7 – Iran)	Technologischer Stress (Tech Anxiety)	Stress und Angst der Studierenden (Student Stress)	
„Wenn der Server abstürzt, verlieren die Studierenden ihr Vertrauen.“ (T.N4 – Deutschland)	Misstrauen gegenüber dem System (Distrust in System)	Niedrige Zuverlässigkeit (Low Reliability)	
„Studierende sagen, Online-Prüfungen werden nicht ernst genommen.“ (T.N6 – Deutschland)	Niedrige soziale Glaubwürdigkeit (Low Social Credibility)	Prüfungscredibility (Exam Credibility)	
„Lehrende verfügen nicht über ausreichende Werkzeuge für Prüfungsdesign.“ (T.N8 – Iran)	Fehlende Fähigkeiten im Prüfungsdesign (Lack of Test Design Skills)	Geringe Kompetenz der Lehrenden (Low Faculty Competence)	
„Studierende können von überall an der Prüfung teilnehmen.“ (T.N6 – Deutschland)	Remote Access (Fernzugang)	Zeitlich-räumliche Flexibilität (Flexibility of Time & Place)	
„Diese Flexibilität ist für Studierende aus entlegenen Regionen im Iran sehr wertvoll.“ (T.N11 – Iran)	Zugang für benachteiligte Regionen (Access for Remote Areas)	Erweiterte Zugänglichkeit (Greater Accessibility)	
„Universitäten reduzieren die Kosten für Präsenzprüfungen.“ (T.N3 – Deutschland)	Institutionelle Kostenreduktion (Institutional Cost Reduction)	Kosteneffizienz (Cost Efficiency)	
„Studierende müssen keine Fahrtkosten mehr zahlen.“ (T.N12 – Iran)	Reduzierte individuelle Kosten (Reduced Student Expenses)	Individuelle Kostenersparnis (Personal Cost Reduction)	
„Wir konnten mehrere Hundert Studierende gleichzeitig prüfen.“ (T.N4 – Deutschland)	Skalierbarkeit (Scalability)	Großskalige Bewertung (Large-Scale Assessment)	**Chancen des Online-Assessments (Opportunities)**
„Diese Möglichkeit war während Corona im Iran sehr wichtig.“ (T.N10 – Iran)	Notwendigkeit in Krisen (Crisis Necessity)	Funktionalität in Krisenkontexten (Relevance in Crisis Contexts)	
„Ein interkultureller Vergleich wurde möglich.“ (T.N2 – Deutschland)	Vergleichende Perspektive (Comparative Perspective)	Kulturelle Vielfalt (Cultural Diversity)	
„Lehrende können die Ergebnisse leichter analysieren.“ (T.N5 – Deutschland)	Erleichterte Datenanalyse (Ease of Data Analysis)	Schnellere Auswertung (Faster Evaluation)	
„Online-Prüfungen sparen Zeit.“ (T.N9 – Iran)	Zeitersparnis (Time Efficiency)	Besseres Zeitmanagement (Time Management)	
„Auch Studierende mit motorischen Einschränkungen können teilnehmen.“ (T.N7 – Iran)	Inklusion (Inclusion)	Zugangsgerechtigkeit (Accessibility & Equity)	

„Daten werden für zukünftige Forschung gespeichert.“ (T.N13 – Iran)	Datenarchivierung (Data Archiving)	Sekundärnutzung (Secondary Use)	
„Studierende werden mit digitalen Umgebungen vertrauter.“ (T.N8 – Iran)	Digitale Kompetenz (Digital Literacy)	Entwicklung digitaler Fähigkeiten (Digital Competence)	
„Diese Erfahrung kann Teil des zukünftigen Blended-Learning sein.“ (T.N1 – Deutschland)	Verbindung zum Blended Learning (Blended Learning)	Zukunftsorientierung (Future Readiness)	
„Wenn das Prüfungsdesign verändert wird, steigt auch die Sicherheit.“ (T.N13 – Iran)	Innovationsorientiertes Design (Innovative Test Design)	Innovatives Prüfungsdesign (Innovative Assessment Design)	
„Ein Teil der Prüfungen muss weiterhin in Präsenz stattfinden.“ (T.N2 – Deutschland)	Hybrid Assessment (Hybridform)	Hybridmodell (Hybrid/Blended Models)	
„Ohne eine starke nationale Plattform gibt es immer Betrug.“ (T.N7 – Iran)	Entwicklung nationaler Plattformen (Local Platform Development)	Technologische Lokalisierung (Localization)	
„Investitionen in Infrastruktur sind notwendig.“ (T.N5 – Deutschland)	Digitale Infrastrukturinvestition (Digital Infrastructure Investment)	Infrastrukturstärkung (Infrastructure Development)	
„Lehrende müssen Prüfungsdesign erlernen.“ (T.N8 – Iran)	Kompetenzentwicklung der Lehrenden (Faculty Training)	Empowerment der Lehrenden (Faculty Empowerment)	
„Wir haben Workshops für Studierende durchgeführt.“ (T.N6 – Deutschland)	Schulung der Studierenden (Student Training)	Empowerment der Studierenden (Student Empowerment)	
„Offene Prüfungsfragen erschweren Betrug.“ (T.N3 – Deutschland)	Open-Ended Exams (Offene Aufgabenformen)	Diversifikation der Bewertung (Diversification of Assessment)	**Strategien zur Verbesserung (Strategies)**
„Projektbasierte Prüfungen sind nützlicher.“ (T.N9 – Iran)	Projektbasierte Bewertung (Project-Based Assessment)	Tiefes Lernen (Deeper Learning)	
„Kontinuierliche Bewertung ist besser als eine einzige Abschlussprüfung.“ (T.N12 – Iran)	Kontinuierliche Bewertung (Continuous Assessment)	Prozessorientierung (Process Orientation)	
„Es müssen ethische Standards definiert werden.“ (T.N11 – Iran)	Ethische Richtlinien (Ethical Guidelines)	Akademische Integrität (Academic Integrity)	
„Intelligente Videoüberwachung hilft.“ (T.N4 – Deutschland)	KI-Proctoring (AI Proctoring)	Fortschrittliche Technologie (Advanced Technology)	
„Universitäten müssen abgestimmte Richtlinien haben.“ (T.N10 – Iran)	Politische Harmonisierung (Policy Alignment)	Institutionelle Konsistenz (Institutional Consistency)	
„Die Überprüfung der Ergebnisse sollte gleichzeitig durch Lehrende und Software erfolgen.“ (T.N1 – Deutschland)	Doppelte Bewertung (Dual Evaluation)	Mehrschichtige Evaluation (Multi-Layered Evaluation)	
„Internationale Erfahrungen müssen lokal angepasst werden.“ (T.N2 – Deutschland)	Lokalisierung globaler Praktiken (Localization of Global Practices)	Kulturelle Anpassung (Cultural Adaptation)	

„Es sollte eine gemeinsame Expertengruppe für Prüfungsdesign gebildet werden.“ (T.N13 – Iran)	Kollektive Zusammenarbeit (Collaborative Design)	Wissenschaftliche Synergie (Collaborative Improvement)

5. Schlussfolgerung

Die Innovationsleistung dieser Untersuchung liegt in ihrem vergleichenden und erfahrungsorientierten Ansatz; das bedeutet, dass sie sich auf die Berichte von Hochschullehrenden aus Iran und Deutschland stützt und dadurch ein facettenreiches Bild der Herausforderungen, Chancen und Strategien zur Verbesserung des Online-Assessments in der Hochschulbildung zeichnet. Die Ergebnisse zeigen, dass sich die zentralen Herausforderungen im Wesentlichen auf drei Bereiche konzentrieren: Sicherheit und Authentizität der Antworten (Sorge vor Betrug, unzureichende Überwachung und Verletzbarkeit der Prüfungsinhalte; –T.N2 Deutschland, –T.N7 Iran), technische Infrastruktur und Systemstabilität (Internetunterbrechungen, Serverabstürze; –T.N4 Deutschland, –T.N9 Iran) sowie Interaktion und Feedback (Rückgang der individuellen Rückmeldungen und schwächere pädagogische Interaktion; –T.N1 Deutschland, –T.N8 Iran). Diese Befunde stehen im Einklang mit internationalen Studien, die die Notwendigkeit innovativer Prüfungsdesigns zur Stärkung der Validität und Minimierung von Betrug hervorheben (Crisp & Ward, 2022; Joughin, 2010). Gleichzeitig wurden bedeutende Chancen identifiziert. Zu den wichtigsten zählen zeitliche und räumliche Flexibilität, Kostenersparnis, Skalierbarkeit der Prüfungen und Funktionsfähigkeit in Krisensituationen (–T.N3 Deutschland; –T.N10 Iran). Darüber hinaus wurden positive Begleitwirkungen wie die Verbesserung der digitalen Kompetenz, die Inklusion von Studierenden mit körperlichen Einschränkungen und die Möglichkeit der Datenspeicherung für weiterführende Analysen hervorgehoben (–T.N7 Iran; –T.N13 Iran). Diese Ergebnisse stimmen mit deutschen Studien zu Online-Selbsttests und E-Prüfungen überein (Reiss et al., 2009: 3; Lehmann et al., 2015: 12; Grassinger & Schnabel, 2024: 5; Domen et al., 2021: 92). Aus der Perspektive der Bildungsgerechtigkeit zeigen die Befunde, dass die schwache Internetinfrastruktur und die ungleiche Verfügbarkeit technischer Ausstattung im Iran psychischen Druck und Misstrauen gegenüber den Systemen verstärken (–T.N10 Iran; –T.N7 Iran). Demgegenüber schaffen zeitliche und räumliche Flexibilität, Kostenreduktion sowie die Teilnahmemöglichkeiten für Studierende aus abgelegenen Regionen oder mit körperlichen Einschränkungen wichtige Voraussetzungen für größere Inklusion (–T.N6 Deutschland; –T.N11 Iran). Diese Dichotomie verdeutlicht, dass die Konsequenzvalidität (Messick, 1995: 741; Kane, 2013: 2) des Online-Assessments in hohem Maße von unterstützenden politischen Maßnahmen und Investitionen in die digitale Infrastruktur abhängt. Die Ergebnisse betonen ferner die hohe Bedeutung von Feedback und Systemvertrauen. Der Rückgang der Feedbackqualität und die Oberflächlichkeit der Interaktionen wurden in beiden Ländern berichtet; dennoch können Vorschläge wie mehrschichtige Überprüfung (Lehrende + Software), Schulungen zur Prüfungsentwicklung für Lehrende und die Nutzung gespeicherter Daten für ergänzende Analysen zur Verbesserung des Feedbackkreislaufs und zur

Stärkung des institutionellen Vertrauens beitragen (T.N1 – Deutschland; T.N8 – Iran; T.N13 – Iran). Diese Beobachtungen stehen im Einklang mit Shin (2022: 75) sowie Martin et al. (2020: 112–114), die auf die Herausforderung individualisierten Feedbacks und die Bedeutung digitaler Gerechtigkeit hinweisen. Auf theoretischer Ebene bestätigt die Analyse, dass Messicks (1995) Validitätskette und Kanes (2013) interpretative Argumentationsstruktur im Kontext von Online-Prüfungen vor neue Herausforderungen gestellt werden – insbesondere im Hinblick auf Interpretationsvalidität und Konsequenzvalidität. Darüber hinaus wurde die Datenanalyse entsprechend dem thematischen Analyseansatz von Braun und Clarke (2006: 78–79) und den Kriterien der Vertrauenswürdigkeit nach Lincoln und Guba (1985: 289–331) durchgeführt, was die Glaubwürdigkeit der Ergebnisse stärkt.

Die vorgeschlagenen Strategien dieser Studie lassen sich in drei Ebenen einteilen:

- **Makropolitik:** Entwicklung nationaler Plattformen, Investitionen in digitale Infrastruktur und institutionelle Kohärenz (–T.N5 Deutschland; –T.N7 Iran; –T.N10 Iran).
- **Humanressourcen:** Stärkung technologischer und prüfungsdidaktischer Kompetenzen der Lehrenden sowie Förderung digitaler Fähigkeiten bei Studierenden (–T.N6 Deutschland; –T.N8 Iran).
- **Innovatives Assessment-Design:** offene und projektbasierte Prüfungen, hybride Präsenz–Online-Modelle, kontinuierliche Bewertung und mehrschichtige Überprüfung (–T.N2 Deutschland; –T.N3 Deutschland; –T.N9 Iran).

Insgesamt zeigt die vorliegende Studie, dass die Erfahrungen mit Online-Assessments in Iran und Deutschland sowohl Gemeinsamkeiten als auch Unterschiede aufweisen: Sicherheits- und Infrastrukturprobleme sind im Iran stärker ausgeprägt, während beide Länder ähnliche Vorteile wie Flexibilität und Kostenersparnis erleben. Daraus lässt sich ableiten, dass nur durch innovative Gestaltungsansätze, eine robuste Infrastruktur und die Befähigung von Lehrenden und Studierenden ein verlässliches und dauerhaft tragfähiges Modell des Online-Assessments in der Hochschulbildung etabliert werden kann. Aus den Ergebnissen lassen sich mehrere praxisorientierte Maßnahmen ableiten, die zur Stärkung der Validität, Gerechtigkeit und Effizienz von Online-Assessments beitragen können. Auf makropolitischer Ebene sind der Ausbau digitaler Infrastruktur und Investitionen in sichere nationale Plattformen – wie dies in Deutschland durch Systeme wie ONYX und OPAL realisiert wurde (Domen et al., 2021: 92) – eine unverzichtbare Voraussetzung. Ebenso kann institutionelle Harmonisierung sowie die Entwicklung unterstützender Richtlinien zur Sicherung akademischer Integrität und Stärkung des öffentlichen Vertrauens – vergleichbar mit den Leitfäden deutscher Hochschulen zu Online-Selbsttests und E-Prüfungen (Reiss et al., 2009: 3; Grassinger & Schnabel, 2024: 5) – zur Förderung der Konsequenzvalidität beitragen. Auf institutioneller Ebene ermöglicht der Einsatz neuer Technologien wie KI-gestütztem Proctoring und mehrschichtiger Überprüfung eine Verringerung von Betrug und eine Stärkung des Vertrauens der Studierenden (Lenhard et al., 2007: 155). Die deutsche Erfahrung in der Etablierung wissenschaftlicher Netzwerke zwischen Lehrenden

zeigt ferner, dass gemeinsame Arbeitsgruppen den Prozess der Prüfungsneugestaltung und disziplinübergreifenden Erfahrungsaustauschs unterstützen können (Heil & Effenthaler, 2023: 16–18). Zudem tragen klar definierte Protokolle für Prüfungen in Krisensituationen – sowohl im Iran als auch in Deutschland – wesentlich zur Sicherung der Bildungskontinuität bei (–T.N10 Iran; –T.N4 Deutschland). Auf pädagogischer und praktischer Ebene ist die Befähigung der Lehrenden zur Entwicklung innovativer, digitaler Prüfungsformate von besonderer Bedeutung. Die Evidenz zeigt eindeutig, dass die bloße Übertragung von Präsenzprüfungen ins Online-Format nicht ausreicht; vielmehr benötigen Lehrende neue Kompetenzen, um projektbasierte, offene und hybride Prüfungsformate zu gestalten (–T.N8 Iran; –T.N2 Deutschland). Parallel dazu trägt die Förderung digitaler Kompetenz und wissenschaftlicher Redlichkeit unter Studierenden zur Verringerung von Bildungsungleichheiten und größerer Inklusion bei (Martin et al., 2020: 112–114). Darüber hinaus kann die Diversifizierung der Bewertungsmethoden – etwa durch kontinuierliche Bewertung, projektbasierte Prüfungen und hybride Präsenz–Online-Modelle, wie in der internationalen Literatur empfohlen (Crisp & Ward, 2022; Joughin, 2010) – zu tieferem Lernen und wirksamerem Feedback beitragen. Daraus folgt, dass der Weg zu einem nachhaltigen Modell des Online-Assessments nur durch die gleichzeitige Stärkung der drei Ebenen Politik, Institution und Pädagogik möglich ist. Ein solcher Ansatz trägt nicht nur zur Bewältigung aktueller Herausforderungen bei, sondern schafft auch die Grundlage dafür, Online-Assessments von einer Krisenlösung in ein dauerhaftes und transformatives Potenzial für die Hochschulbildung zu überführen.

Anhang: Beispiel eines vollständig transkribierten halbstrukturierten Interviewleitfadens sowie eine Abbildung der Kodierungsschritte der thematischen Analyse.

❊❊❊

1. The author declares that this article did not receive financial support from any institution, educational center, or approved research project.

2. The author declares that there is no conflict of interest related to this research.

3. The author declares that no generative artificial intelligence was used in writing this article and that the entire work is their own creation.

پرسش‌های مصاحبه کیفی برای استادان دانشگاه‌های ایران و آلمان

Qualitative Interviewfragen für Professorinnen und Professoren in Iran und Deutschland

این پرسش‌ها در چارچوب پژوهش با موضوع "واکاوی اثربخشی سنجش‌های آنلاین در آموزش عالی ایران و آلمان" طراحی شده است. هدف، گردآوری تجربه‌های زیسته استادان برای شناسایی چالش‌ها، فرصت‌ها و راهکارهای بهبود است.

- نام (اختیاری):
- دانشگاه (اختیاری):
- رشته / دروس تدریسی:

مصاحبه با استاد ایرانی

پژوهشگر :لطفاً برداشت کلی خودتان از اثربخشی سنجش آنلاین در دانشگاه ایران را بفرمایید.

استاد (ایران) :من فکر می‌کنم سنجش آنلاین از نظر انعطاف‌پذیری خیلی مفید بود؛ «دانشجویان می‌توانند از هرکجا در آزمون شرکت کنند. «برای دانشجویان شهرستانی این یک مزیت بزرگ است. اما در ایران، به دلیل مشکلات اینترنت و کمبود تجهیزات، گاهی عدالت آموزشی زیر سؤال می‌رود؛ «همه دسترسی یکسان ندارند.»

پژوهشگر :مهم‌ترین چالش‌ها برای شما چه بوده است؟

استاد (ایران) :اصلی‌ترین مسئله تقلب است.» «ممیز نگران مستیم که آیا دانشجو خودش امتحان می‌دهد یا فرد دیگری. «اضافه بر آن، «اینترنت در میانه آزمون قطع شد «برای چند دانشجو، و همین باعث استرس و نارضایتی آن‌ها شد. همچنین «سایت ابزار کافی برای طراحی آزمون ندارند. «او خیلی وقت‌ها ما فقط آزمون حضوری را به صورت آنلاین منتقل می‌کنیم، بدون نوآوری.

پژوهشگر :درباره بازخورد چه نظری دارید؟

استاد (ایران) :بازخورد در آزمون آنلاین محدود است.» «کیفیت بازخورد پایین‌تر است. «او دانشجوها می‌گویند تعامل ما با آن‌ها کمتر شده. در واقع، وقتی سرور یا سیستم از کار می‌افتد، اعتماد دانشجو هم آسیب می‌بیند.

پژوهشگر :آیا فرصت‌هایی هم ایجاد شده؟

استاد (ایران) :بله، فرصت‌ها هم قابل توجه‌اند.» «امتحانات آنلاین باعث صرفه‌جویی در زمان می‌شود. «او هم ما و هم دانشجوها از هزینه رفت‌وآمد رها می‌شویم. به‌ویژه در دوران کرونا، این ویژگی خیلی ارزشمند بود.

پژوهشگر :چه راهکارهایی برای بهبود پیشنهاد می‌دهید؟

استاد (ایران) :به نظر من باید «پلتفرم بومی قوی نوسعه پیدا کند. «او اساتید هم «مهارت آزمون‌سازی بیاموزند. «ترکیب آزمون‌های حضوری و آنلاین خیلی کمک می‌کند. همچنین استفاده از "آزمون‌های پروژه‌محور «می‌تواند هم تقلب را کاهش دهد و هم یادگیری عمیق‌تر ایجاد کند.

پرسش‌های مصاحبه کیفی برای استادان دانشگاه‌های ایران و آلمان

Qualitative Interviewfragen für Professorinnen und Professoren in Iran und Deutschland

Diese Fragen wurden im Rahmen einer Studie zum Thema „*Analyse der Effektivität von Online-Assessment-Systemen im Hochschulbereich Iran und Deutschland*" entwickelt. Ziel ist es, die gelebten Erfahrungen der Hochschullehrenden zu erfassen, um Herausforderungen, Chancen und Verbesserungsstrategien zu identifizieren.

- Name (optional):
- Universität (optional):
- Fach / Lehrveranstaltungen:

Interview mit einem deutschen Professor

Forscher: Wie definieren Sie die Wirksamkeit von Online-Prüfungen in der deutschen Hochschulbildung?

Professor (Deutschland): Für uns war die Flexibilität und der breite Zugang sehr wichtig. „**Die Studierenden können von überall teilnehmen.**" Gleichzeitig bleibt aber die Sorge bestehen: „**Wir sind immer besorgt, ob der Student selbst die Prüfung ablegt oder jemand anderes.**"

Forscher: Welche Herausforderungen waren für Sie am wichtigsten?

Professor (Deutschland): Obwohl unsere Infrastruktur stärker ist, gibt es ernsthafte Probleme. „**Die Plattformen stürzen bei großen Prüfungen ab.**" Dies führt zu einem Vertrauensverlust bei den Studierenden. Außerdem geben wir weniger individuelles Feedback; „**Wir können kein tiefes, individuelles Feedback geben.**"

Forscher: Wie beurteilen Sie die Qualität des Feedbacks und die Interaktion?

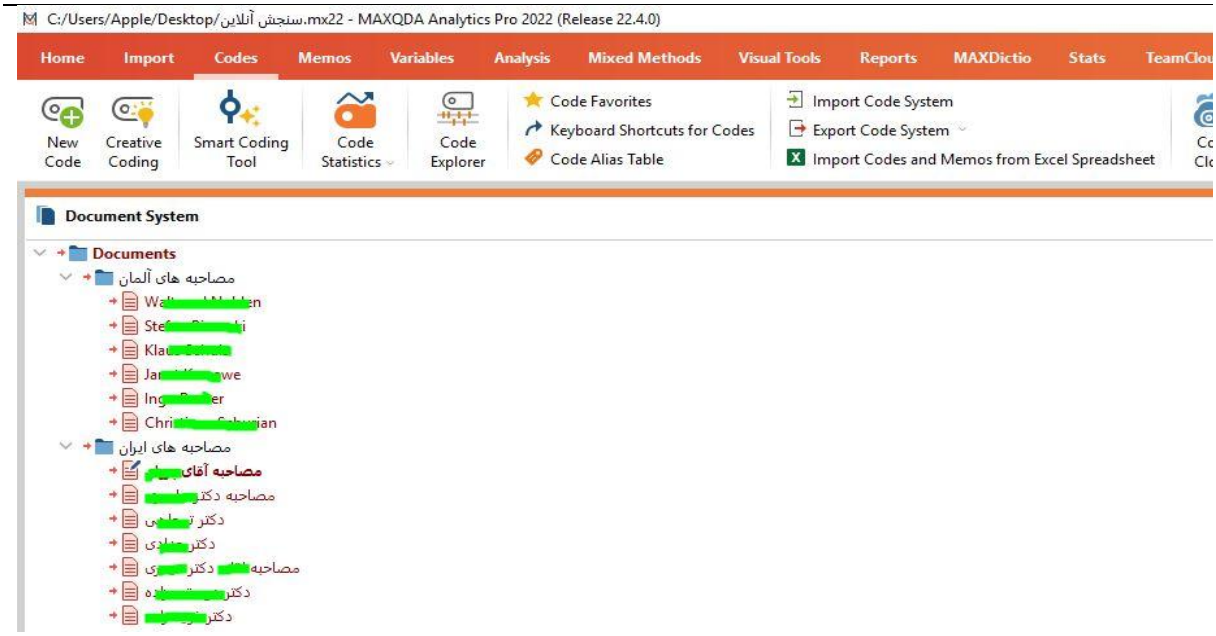
Professor (Deutschland): Der Austausch mit den Studierenden ist online deutlich reduziert. „**Die Interaktion zwischen Lehrenden und Studierenden ist schwächer.**" Und die Rückmeldungen wirken oberflächlicher als in Präsenz.

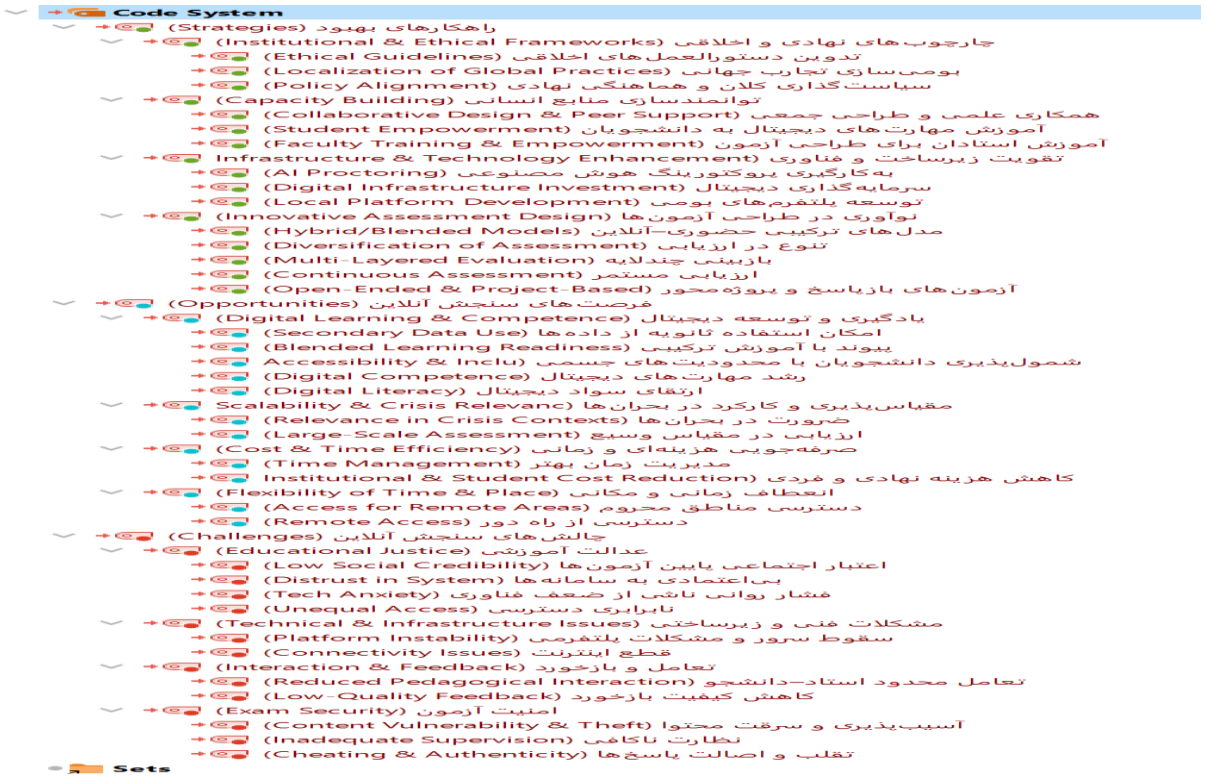
Forscher: Gab es auch Chancen oder Vorteile?

Professor (Deutschland): Auf jeden Fall. „**Wir konnten gleichzeitig mehrere hundert Studierende bewerten.**" Das wäre in Präsenz unmöglich gewesen. Außerdem sparen sowohl Universität als auch Studierende viel Zeit und Kosten.

Forscher: Welche Strategien schlagen Sie zur Verbesserung vor?

Professor (Deutschland): Ich denke, wir müssen uns in Richtung „**Hybridmodell**" bewegen: ein Teil Präsenz, ein Teil online. „**Projektbasierte Prüfungen sind sinnvoller und erhöhen die Tiefe des Lernens.**" Wenn die Prüfungen innovativer gestaltet werden, „**wird auch die Sicherheit höher.**"





Literaturverzeichnis

Azimi, H., Rahimi, M., & Doroudi, R. (2016). Academic dishonesty in Iran and Russia: A comparative study on the role of plagiarism detection and pre-tests. *Journal of Academic Ethics*, *14*(2), 121–134. https://doi.org/10.1007/s10805-016-9251-4

Azizi, M. (2022). Online assessment and distributive justice in Iranian higher education: Challenges and opportunities. *Journal of Educational Technology Development*, *18*(1), 39–54. https://doi.org/10.22054/edtech.2022.34567

Bauer, J., Figl, K., & Nistor, N. (2009). Peer assessment in higher education: Benefits and challenges. *Assessment & Evaluation in Higher Education*, *34*(1), 75–89. https://doi.org/10.1080/02602930801895736

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education*, *159*, 104024. https://doi.org/10.1016/j.compedu.2020.104024

Craig, E., & Kay, R. (2025). Enhancing student engagement in online assessments through peer review. *Journal of Online Learning Research*, *11*(1), 45–55.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Crisp, G., & Ward, C. (2022). Re-imagining assessment in higher education: Digital, authentic, and inclusive practices. *Assessment & Evaluation in Higher Education, 47*(4), 565–580. https://doi.org/10.1080/02602938.2021.1887812

Dashtestani, R. (2020). Online language learning in Iran: Students' and instructors' perceptions. *Education and Information Technologies, 25*(1), 81–99. https://doi.org/10.1007/s10639-019-09945-0

Dashtestani, R., & Karami, H. (2019). Validity and reliability issues in online assessment: A review of literature. *Asian EFL Journal, 21*(3), 85–110.

Davis, V. W., Smith, L. T., & Chan, M. (2023). Privacy, equity, and validity in online testing: A framework for higher education. *Assessment in Education: Principles, Policy & Practice, 30*(1), 118–134. https://doi.org/10.1080/0969594X.2022.2094567

Dohmen, D., Kröner, S., & Hartig, J. (2021). Digital examinations in German higher education: Experiences with ONYX, OPAL, and MAXIMA. *Zeitschrift für Hochschulentwicklung, 16*(2), 91–107. https://doi.org/10.3217/zfhe-16-02/05

Gamage, D., Ayres, J., & Behrend, M. (2022). A systematic review on the use of Moodle for online assessment in higher education. *Education and Information Technologies, 27*(1), 13–29. https://doi.org/10.1007/s10639-021-10683-3

Grassinger, R., & Schnebel, S. (2024). Online self-assessments in German higher education: Fostering reflection and learning outcomes. *Journal of Educational Measurement, 61*(1), 1–19. https://doi.org/10.1111/jedm.12345

Heil, C., & Ifenthaler, D. (2023). Quality and acceptance of online assessment in German universities: Policies and practices. *British Journal of Educational Technology, 54*(1), 13–29. https://doi.org/10.1111/bjet.13224

Jeremias, F., & Sellmer, J. (2019). Implementing e-assessments in higher education: Lessons learned from German universities. *Zeitschrift für Hochschulentwicklung, 14*(3), 97–110.

Joughin, G. (2010). The hidden curriculum revisited: A critical review of research into the influence of summative assessment on learning. *Assessment & Evaluation in Higher Education, 35*(3), 335–345. https://doi.org/10.1080/02602930903221493

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Köller, O., Watermann, R., & Wigfield, A. (2020). Adaptive testing and diagnostic assessment in German higher education. *Studies in Educational Evaluation, 66,* 47–58. https://doi.org/10.1016/j.stueduc.2020.100876

Lehmann, T., Hahn, M., & Halle, T. (2015). Peer assessment and its impact on higher education learning outcomes. *Assessment & Evaluation in Higher Education, 40*(1), 11–22. https://doi.org/10.1080/02602938.2014.976111

Lenhard, W., Baier, H., Hoffmann, J., & Schneider, W. (2007). Computer-based latent semantic analysis in online assessments. *Educational Psychology, 27*(2), 155–176. https://doi.org/10.1080/01443410601066742

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Loureiro, A., & Gomes, C. (2022). Online learning and peer interaction: The role of anonymity in feedback quality. *Computers in Human Behavior, 136,* 233–241. https://doi.org/10.1016/j.chb.2022.107382

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Mounesi Qaramalki, A., Āhangarī, S., & Haddīdi Tamjid, N. (2024). Online foreign language testing: Validity and cognitive load considerations. *Iranian Journal of Applied Linguistics*, *27*(2), 115–134. https://doi.org/10.22054/ijals.2024.34567

Muzaffar, M., Ali, S., & Raza, T. (2020). Security challenges in online assessments: A systematic review. *International Journal of Emerging Technologies in Learning, 15*(10), 6–12. https://doi.org/10.3991/ijet.v15i10.13795

Mutimukwe, C., Lubega, J., & Muyinda, P. (2020). Online assessment in higher education: Equity and privacy concerns. *International Journal of Education and Development using ICT, 16*(2), 63–74.

Petri, K., & Kersting, M. (2024). Online self-assessments as guidance tools in German higher education. *Frontiers in Education, 9,* 1234567. https://doi.org/10.3389/feduc.2024.1234567

Rahimi, M., Aghajani, M., & Khodabakhsh, S. (2019). Technical challenges of online exams in Iranian universities. *Educational Technology Research, 11*(2), 53–61. https://doi.org/10.22054/edtech.2019.27890

Reiss, K., Sälzer, C., & Schiepe-Tiska, A. (2009). Online self-assessment in German higher education: Potentials and limitations. *Zeitschrift für Pädagogik, 55*(1), 1–15.

Shin, H. (2022). Redefining feedback in online higher education: From individualization to dialogic engagement. *Assessment & Evaluation in Higher Education, 47*(7), 1095–1110. https://doi.org/10.1080/02602938.2021.1938123

Strickroth, S., Hesse, F. W., & Fischer, F. (2011). Adaptive online testing in higher education: Potentials and challenges. *Computers & Education, 57*(1), 62–74. https://doi.org/10.1016/j.compedu.2010.12.004

Tahir, R., Yousaf, Z., & Khan, H. (2025). Online exams and the digital divide: A case study from South Asia. *International Review of Research in Open and Distributed Learning, 26*(1), 61–70. https://doi.org/10.19173/irrodl.v26i1.7654

Teshome, A. (2025). Online assessment in African universities: Opportunities and risks. *African Educational Review, 17*(2), 87–99. https://doi.org/10.1080/18146627.2025.123456