

Zusammenfassung

Die Möglichkeiten der Digitalisierung im Bildungsbereich werden bereits seit langer Zeit in Forschung und Praxis adressiert. Diese erstrecken sich hierbei über alle Bestandteile der Curriculum-Instruction-Assessment (CIA) Triade nach PELLEGRINO (2010). So ändern sich durch die Digitalisierung nicht nur die erforderlichen Kompetenzen von zukünftigen Arbeitnehmern, sondern auch die Art und Weise des Lehrens, Lernens und Prüfens. Diese Entwicklung wird durch den technologischen Fortschritt im Bereich der Künstlichen Intelligenz (KI) zusätzlich beschleunigt. Sowohl die STÄNDIGE KONFERENZ DER BILDUNGS- UND KULTUSMINISTER (2022) als auch die STÄNDIGE WISSENSCHAFTLICHE KOMMISSION (2022) weisen in diesem Zusammenhang auf die Notwendigkeit einer verstärkten Adressierung digitaler Prüfungen hin.

Eine im Rahmen dieser Dissertation durchgeführte Literaturanalyse zeigt, dass die aktuelle Forschung zur digitalen Prüfungsdurchführung häufig die Nutzungsperspektive fokussiert. So erfolgt primär eine isolierte Betrachtung einzelner Einflussfaktoren auf die Prüfungsteilnehmer (z.B. Stress, Familiarität etc.). Dabei bleibt eine umfassende Betrachtung potentieller Zusammenhänge zwischen diesen Faktoren weitestgehend aus. Bei der digitalen Prüfungskorrektur wird hier dem KI-Einsatz ein hohes Potential bei der Freitextkorrektur nachgesagt. Es wird gezeigt, dass aktuell lediglich die Korrektorgenauigkeit, jedoch nicht die Konzeption von Korrektursystemen, adressiert wird. Durch diesen rein technischen Fokus findet zudem keine Betrachtung der Nutzungsperspektive (z.B. Vertrauen) statt.

Aufbauend auf diesen Ergebnissen wurden in dieser kumulativen Dissertation fünf Studien zur digitalen Klausurdurchführung und -korrektur durchgeführt. In Verbindung mit den Erkenntnissen aus der Literaturanalyse wurden anhand der Ergebnisse dieser fünf Studien insgesamt 13 Empfehlungen für die Praxis abgeleitet. Diese zeigen, dass nutzungsorientierte Einflussfaktoren durch die Prüfungsverantwortlichen bereits vor der Prüfungsdurchführung adressiert werden können. Hierdurch kann der Einfluss konstrukt-irrelevanter Faktoren auf die Prüfungsergebnisse reduziert und somit die Testqualität erhöht werden. Im Bereich der digitalen Prüfungskorrektur wurde gezeigt, dass trotz technischem Fortschritt eine menschliche Korrekturbeteiligung das Vertrauen in die KI-basierte Korrektur erhöhen kann. Aufbauend auf diesen Erkenntnissen wurden konkrete Gestaltungsempfehlungen für semi-automatische, KI-basierte Korrektursysteme abgeleitet. Hierdurch wird der allgemeine Transfer technischer Forschungsergebnisse zur KI-basierten Klausurkorrektur in Produktivsysteme vereinfacht.

Abschließend werden weitere Ansatzpunkte für die zukünftige Forschung abgeleitet. Potentiale werden hier insbesondere der Entwicklung von Large-Language-Modells zugesagt.

Abstract

The opportunities for digitization in education have been addressed in research and practice for a long time. These extend across all components of the Curriculum-Instruction-Assessment (CIA) triad according to PELLEGRINO (2010). Thus, digitization is not only changing the required competencies of future workers, but also the way of teaching, learning, and testing. This development is further accelerated by the technological progress in the field of artificial intelligence (AI). In this context, both the STÄNDIGE KONFERENZ DER BILDUNGS- UND KULTUSMINISTER (2022) and the STÄNDIGE WISSENSCHAFTLICHE KOMMISSION (2022) point to the need for increased addressing of digital assessment.

A literature analysis conducted as part of this dissertation shows that current research on digital exam execution often focuses on the usage perspective. Thus, primarily an isolated consideration of individual factors influencing the examinees (e.g., stress, familiarity, etc.) takes place. A comprehensive consideration of potential interrelationships between these factors is largely omitted. In the case of digital exam scoring, the use of AI is said to have a high potential in essay scoring. It is shown that currently only the scoring accuracy, but not the design of essay scoring systems, is addressed. This purely technical focus also means that the user perspective (e.g., trust) is not taken into account.

Building on these findings, five studies on digital exam execution and scoring are conducted in this cumulative dissertation. In conjunction with the findings from the literature analysis, a total of 13 recommendations for practice were derived based on the results of these five studies. These show that examiners can address usage-oriented factors even before digital exams are conducted. This can reduce the influence of construct-irrelevant factors on test results and thus increase test quality. In the area of digital exam scoring, it is shown that despite technological advances, human scoring involvement can increase confidence in AI-based scorings. Based on these findings, specific design recommendations for semi-automatic AI-based scoring systems are derived. This simplifies the general transfer of technical research results on AI-based exam scoring into productive systems.

Finally, further starting points for future research are derived. In particular, the development of large language models (LLM) is expected to have potential.