

Digitales Prüfen

Didaktik, Umsetzung und Evidenz für die neue Prüfungsnormalität an Hochschulen

Malte Persike

Zusammenfassung

Das Prüfen mit digitalen Mitteln an Hochschulen hat in den vergangenen Jahren kontinuierlich an Bedeutung gewonnen. Viele Hochschulen führen digitale Prüfungen in großer Zahl durch und nutzen dabei die immer umfangreicher werdenden digitalen Möglichkeiten. Der Artikel beginnt mit einer Begriffsbestimmung der verschiedenen Formen digitaler Prüfungen. Er berücksichtigt dabei sowohl prüfungsdidaktische und technische Ordnungsdimensionen als auch bestehende Einstufungen nach dem Hochschulprüfungsrecht. Daraus wird eine Arbeitsdefinition abgeleitet, die Begriffe wie die digitale und elektronische Prüfung, das E-Assessment, den digitalen Test und Hybridformen umfasst. Im Anschluss werden formative und summative Prüfungsformen vorgestellt und voneinander abgegrenzt. Zu den offenen Prüfungsbedingungen wird eine ausführliche Besprechung der empirischen Forschungsliteratur vorgenommen. Der Umgang mit Täuschungen, von Täuschungsvermeidung, über Täuschungsüberwachung bis hin zur Täuschungsentdeckung, wird in einem eigenen Kapitel erläutert. Der Artikel schließt mit einer Diskussion der Bedingungen von Akzeptanz für digitale Prüfungsformen. Umfangreiche Tabellenanhänge zu Täuschungsszenarien sowie aktuell genutzten E-Prüfungssystemen an deutschen Hochschulen vervollständigen den Text.

Schlüsselwörter

Elektronische Prüfung, digitale Prüfung, E-Assessment, Proctoring

Digital assessment – didactics, implementation and evidence for the new testing normality at universities

Keywords

Digital exam, electronic exam, e-assessment, proctoring

1 Einführung

Die Bedeutung des Prüfens mit digitalen Mitteln an Hochschulen ist seit Jahren stetig gewachsen. Dieser Trend hatte sich bereits mit der Diskussion um digitale Badges und Microdegrees beziehungsweise Nanodegrees (Hochschulrektorenkonferenz 2020) abgezeichnet, deren Erwerb zumeist mit Hilfe digitaler Prüfungen stattfindet. Badges sind digitale Plaketten, die die Teilnahme an Weiterbildungen, das Ablegen von Leistungsprüfungen oder allgemein das Vorhandensein bestimmter Kompetenzen belegen. Unter Micro- und Nanodegrees versteht man Zertifikate oder Abschlüsse, die nach der Teilnahme an einem Bildungsangebot mit reduziertem zeitlichen oder inhaltlichen Umfang im Vergleich zu regulären Lehrveranstaltungen an Hochschulen bestehen. Wenngleich im Zuge der Etablierung dieser Zertifizierungsinstrumente auch die Nutzung digitaler Prüfungsformen angewachsen ist, hat vor allem die Corona-Pandemie maßgebliche Impulse zum Einsatz digitaler Prüfungen in der Breite deutscher Hochschulen geliefert. Viele Hochschulen haben zum ersten Mal digitale Prüfungen in größerem Umfang durchgeführt und an Hochschulen, die bereits über eine etablierte Infrastruktur für digitale Prüfungen verfügen, hat sich deren Anzahl um ein Vielfaches erhöht. Parallel zu dieser raschen Entwicklung sind auch die Anforderungen für elektronische Prüfungs durchführungen effizienter machen, örtlich und zeitlich flexibilisieren, die Situierung und Kontextualisierung von Aufgabenformen ermöglichen, eine Kalibrierung auf das spezifische Kompetenzniveau der Lernenden erlauben und eine schnellere Ergebnisrückmeldung leisten. Diese organisatorischen, psychologischen, analytischen und pädagogischen Vorteile wurden bereits im Jahr 2008 von der EU-Kommission formuliert (Scheuermann und Pereira 2008). Die Einführung digitaler Prüfungsformen darf aber nicht zulasten der etablierten didaktischen und diagnostischen Anforderungen an Prüfungen gehen. Elektronische Prüfungen müssen weiterhin klassische Testgütekriterien wie Objektivität, Reliabilität und Validität erfüllen (Moosbrugger und Kelava 2008). Sie müssen transparent sein und unter reproduzierbaren Bedingungen stattfinden. Gleichzeitig müssen sie hochschuldidaktische Innovationen abbilden, Ansätze wie Constructive Alignment fördern (Biggs 1996), kompetenzorientiert und diversitätsgerecht angelegt werden (Urfer-Schumacher 2016). Nur dann können sie die für die Hochschulen wichtigen Funktionen der Allokation und Selektion erfüllen, aber auch die besonders bedeutsame Steuerungsfunktion für Lehren und Lernen sowie für die modularisierte Gestaltung von Studiengängen übernehmen.

2 Begriffsbestimmung

Der in diesem Kapitel bislang verwendete Terminus »digitale Prüfung« zerfällt im praktischen Gebrauch in eine Vielzahl verwandter Konzepte und Bezeichnungen. Für die noch keine eindeutige begriffliche Klärung und Ausdifferenzierung existiert. Eine Arbeitsdefinition soll im Folgenden vorgenommen werden, wobei zwischen einer prüfungsdidaktischen und einer prüfungsrechtlichen Definitions perspektive unterscheiden wird.

2.1 Prüfungsdidaktische und technische Definition

Der Wunsch nach einer geeigneten Taxonomie für die Art und Weise, wie an Hochschulen geprüft wird (zum Beispiel Reis und Ruschin 2008; Wildt und Wildt 2011), umfasst auch die digitalen Prüfungsvorgänge. Digitale Prüfungen lassen sich gut in den von Huber und Reinmann (2019) aufgespannten Ordnungsrahmen für Prüfungen einpassen. Dieser unterscheidet zwischen drei Grundformen: mündliche Prüfungen, schriftliche Prüfungen und praktische beziehungsweise enaktive Prüfungen. Zu den letzteren gehören Demonstrationen, bei denen Studierende den Prüfenden eine Aktivität zeigen oder vorführen, und Produktionen, bei denen die Prüfenden lediglich das Ergebnis eines praktischen Tuns bewerten, ohne den Produktionsvorgang selbst zu verfolgen.

Mündliche Prüfungen werden grundsätzlich nicht als digitale Prüfungen aufgefasst, auch wenn sie auf digitalem Wege stattfinden, zum Beispiel als Videokonferenz. Deshalb fallen digitale Prüfungen in aller Regel unter die schriftlichen oder die praktischen beziehungsweise enaktiven Prüfungen, wobei die zunehmende Vielfalt der digital unterstützten Prüfungsformen die Grenzen zwischen diesen Kategorien verschwimmen lässt. So wird beispielsweise eine papiergebundene Klausur im Fach Informatik in aller Regel den schriftlichen Prüfungen zugeordnet (Pörzgen, Olivier, Sackbrook und Pinkwart 2012). Diese Einschätzung kann sich verändern, wenn die Klausur mit Hilfe eines Computer Algebra Systems (CAS) umgesetzt wird. Hier steigt der enaktive Anteil unter Umständen stark an. Es können Simulationen durchgeführt und komplexe Modelle entwickelt werden. Eine solche mit digitalen Mitteln durchgeführte Klausur trüge viele Kennzeichen der enaktiv-produzierenden Prüfung. Die zunehmende Durchlässigkeit der Kategorien verweist auf die Notwendigkeit zur Überarbeitung der gängigen Prüfungstaxonomien. Im praktischen Alltag der Prüfungsdurchführung an Hochschulen ist aber oft nicht die Klassifikation der Prüfungsform an sich, sondern vor allem die Abgrenzung der digitalen von der traditionell analogen Prüfung von Interesse. Prüfungsrechtlich und organisatorisch kann es notwendig sein zu beantworten, wann es sich bei einem gegebenen Prüfungsvorgang um eine digitale Prüfung handelt und welche Unterformen dabei zu differenzieren sind. Dies macht eine stärker an der technischen als der didaktischen Gestaltung orientierte Definition des Begriffs der digitalen Prüfung sinnvoll. Eine solche soll im Folgenden versucht werden.

Die verschiedenen Formen der digitalen Prüfung werden auch in Deutschland häufig mit dem aus der englischsprachigen Literatur entlehnten Begriff des Electronic Assessment (E-Assessment) zusammengefasst (Ridgway, McCusker und Pead 2004). Aus prüfungsdidaktischer Sicht meint der Begriff E-Assessment alle hauptsächlich kognitiven Leistungsmessungen, deren Vorbereitung, Durchführung und Verarbeitung mit Hilfe digitaler Informations- und Kommunikationstechnologien vorgenommen werden (Schmees und Horn 2014). Dabei lassen sich zwei Auslegungsansätze identifizieren, die sich im Grad der Digitalität des Prüfungsverlaufs unterscheiden. In der restriktiven Auslegung müssen alle genannten Phasen des Prüfungsverlaufs sowohl für die Prüfenden als auch für die Studierenden digital unterstützt sein. So handelt sich bei einer Prüfung, die mit einer Software absolviert und von den Studierenden als digitales Dokument abgegeben wird, noch nicht notwendigerweise um ein E-Assessment. Wenn der Prüfende die digitalen Dokumente zum Zwecke der Korrektur ausdrückt und auf

Papier korrigiert, wäre der digitale Prozess unterbrochen, womit gemäß der restriktiven Definition kein E-Assessment mehr vorläge. Nur wenn alle Schritte des Prüfungsverlaufs von der Vorbereitung bis zur Verarbeitung elektronisch erfolgen, handelt es sich um vollwertige E-Assessments. Praxisnäher ist die Auslegung aus Studierendenperspektive. Hier gilt eine Leistungsprüfung dann als E-Assessment, wenn alle Phasen des Prüfungsverlaufs aus Studierendensicht digital durchgeführt werden. Die oben als Beispiel verwendete softwarebasierte Prüfung wäre in dieser Auslegung zweifelsfrei ein E-Assessment, da Unterbrechungen des digitalen Prüfungsprozesses auf Seiten der Prüfenden keine Rolle spielen.

Ein deutschsprachiges Pendant zum E-Assessment zu finden, gestaltet sich schwierig. Als Kandidaten kommen sowohl die Begriffe digitale Prüfung wie auch elektronische Prüfung in Betracht. Beide sind aber aus zwei wesentlichen Gründen keine adäquate Übersetzung. Erstens ist die Prüfung im deutschsprachigen Gebrauch konzeptuell enger gefasst als das Assessment. Prüfungen bezeichnen meist nur studienverlaufsrelevante summative Prüfungsleistungen wie Abschluss- oder Examensklausuren. Formative Assessments hingegen werden eher selten als Prüfungen, sondern meist als »Tests« bezeichnet. Zweitens kommt dem Begriff der elektronischen Prüfung aus prüfungsrechtlicher Sicht eine ausgezeichnete Stellung zu.

2.2 Einstufung nach dem Hochschulprüfungsrecht

Im Hochschulprüfungsrecht wird neben den grundsätzlich nicht-elektronischen mündlichen Prüfungen lediglich zwischen schriftlichen Prüfungen auf der einen Seite und elektronischen Prüfungen (E-Prüfungen) beziehungsweise elektronischen Klausuren (E-Klausuren) auf der anderen Seite unterschieden. E-Prüfung und E-Klausur sind dabei synonym zu verwenden. Die E-Prüfung wird in den einschlägigen Rechtsvorschriften meist in direkter Verbindung mit sogenannten Antwort-Wahl-Aufgaben besprochen, also Multiple oder Single Choice Aufgaben, die mit der Maus angeklickt werden. Dieser Definitionsversuch geht auf frühe Formen der E-Prüfungen im Hochschulbereich zurück, die mit überwiegender Mehrheit im Antwort-Wahl-Format implementiert waren. Diese enge juristische Definition der E-Prüfung ist aus prüfungsdidaktischer Sicht nicht zufriedenstellend und auch im Rechtsgebrauch weitgehend überholt (Morgenroth 2020). Nach aktuell herrschender Rechtsauffassung hat sich der Begriff der E-Prüfung erweitert und schließt einen weiteren Kreis an elektronischen Prüfungsformen ein. Deren gemeinsames Merkmal ist, dass die Verteilung, Durchführung und Verarbeitung der Prüfung in *demselben* informationstechnischen System erfolgen muss. Nur dieser letzte Teil unterscheidet sie von der oben abgeleiteten prüfungsdidaktischen Arbeitsdefinition des E-Assessment. Erst dann, wenn die Eingaben der Studierenden im Rahmen einer Prüfung das E-Prüfungssystem nicht verlassen, kommt prüfungsrechtlich die Einordnung als E-Prüfung überhaupt infrage (Niehues, Fischer und Jeremias 2018). Praktisch läuft dies darauf hinaus, dass nur Prüfungen, die in einem dezidierten E-Prüfungssystem durchgeführt werden, im prüfungsrechtlichen Sinne als E-Prüfungen gelten. Eine elektronisch verteilte Office-Datei, die von Studierenden mit einem Textverarbeitungssystem bearbeitet, per Upload wieder abgegeben und von den Prüfenden dann in einer separaten Software

korrigiert wird, ist demzufolge keine E-Prüfung. Der Verarbeitungsprozess ist zwar eine nicht unterbrochene Kette digitaler Arbeitsschritte, allerdings finden diese nicht in ein- und demselben Softwaresystem statt. Gleichermaßen gilt für Prüfungen, die von Studierenden mit digitalen Werkzeugen wie einer Statistik- oder Konstruktionssoftware bearbeitet und später in elektronischer Form wieder abgegeben werden. Auch hierbei ist es unerheblich, dass der gesamte Prozess der Verteilung, Durchführung und Verarbeitung digital abläuft. Die Tatsache, dass die Prüfung nicht innerhalb eines einzigen integrierten E-Prüfungssystems stattfindet, rückt sie prüfungsrechtlich in den Formenkreis normaler schriftlicher Prüfungen. Angesichts der kontinuierlichen Weiterentwicklung elektronischer Aufgabenformate führt dies notwendigerweise auf Anwendungsfälle, für die noch keine Rechtsklärung hinsichtlich der Einordnung als E-Prüfung oder schriftlicher Prüfung besteht. Verschiedene E-Prüfungssysteme wie zum Beispiel das an der RWTH Aachen University entwickelte Dynexite zielen auf die Ermöglichung komplexer elektronischer Aufgabenformate. Ein Beispiel ist die Stellung von Programmieraufgaben, die in einer in das Prüfungssystem integrierten Programmierumgebung bearbeitet und später ausgewertet werden können. Obwohl eine Rechtsklärung aussteht, fiele eine solche Prüfung vermutlich in den Geltungsbereich der E-Prüfung, da ihre Vorbereitung, Durchführung und Verarbeitung in ein- und demselben Prüfungssystem stattfinden. Es gäbe allerdings nur noch rein formale Unterschiede zu einer Prüfung mit derselben Programmierumgebung, die als separate Software auf dem Prüfungscomputer und nicht als Teil eines dezidierten Prüfungssystems liefe. Eine solche Prüfung wäre aber gemäß der oben skizzierten Rechtsauffassung der E-Prüfung eher nicht als solche, sondern als konventionelle schriftliche Prüfung aufzufassen.

2.3 Arbeitsdefinition

Zur informierten und transparenten Kommunikation über E-Assessments sind eine gemeinsame begriffliche Basis sowie ein geteiltes mentales Modell der jeweiligen Begriffe unumgänglich. Es ist elementar, dass im Diskurs über Konzepte wie E-Prüfungen oder Take-Home Klausuren alle beteiligten Akteure über denselben Gegenstand sprechen. Vor dem Hintergrund der Vielzahl gebräuchlicher Termini wird deshalb eine pragmatische Definition vorgeschlagen, die die Praxis digital unterstützter Prüfungsvorgänge an Hochschulen abbildet.

- 1) Das E-Assessment als Oberbegriff bezeichnet alle hauptsächlich kognitiven Formen der Leistungsmessung, deren Durchführung und Verarbeitung aus Studierendensicht mit digitalen Mitteln erfolgen. Die Beschränkung auf »hauptsächlich kognitive« Leistungsmessungen grenzt E-Assessments von digital unterstützten psychomotorischen Leistungsmessungen ab, zum Beispiel digitale Zeitnahmen bei Sportprüfungen. E-Assessments reichen grundsätzlich von formativen Leistungsmessungen wie freiwilligen semesterbegleitenden Tests oder Hausarbeiten bis hin zu summativen Leistungsmessungen wie elektronischen Abschlussklausuren sowie von hoch standardisierten Prüfungsformen im Antwort-Wahl-Verfahren bis

zu offenen Eingabeformaten wie Programmieraufgaben, die mit Softwareentwicklungsgebungen bearbeitet werden.

- 2) Die digitale Prüfung umfasst alle summativen E-Assessments wie Abschluss- oder Examensklausuren.
- 3) Die elektronische Prüfung (E-Prüfung) bezeichnet im Sinne der prüfungsrechtlichen Normen nur solche summativen E-Assessments, deren Vorbereitung, Durchführung und Verarbeitung innerhalb ein- und desselben E-Prüfungssystems stattfinden.
- 4) Der digitale Test oder elektronische Test (E-Test) meint alle formativen E-Assessments wie semesterbegleitende Quizzes oder digitale Hausübungen.
- 5) Hybridformen verknüpfen E-Assessments mit analogen Prüfungslementen. So kann im Rahmen eines E-Assessments die Anfertigung einer papiergebundenen Zeichnung als Teilleistung erfolgen. In diesem Fall spricht man von hybriden E-Assessments beziehungsweise hybriden digitalen oder elektronischen Prüfungen oder Tests.

Hinzu treten häufig gebrauchte Bezeichnungen wie Fernprüfungen, Take-Home-Klausuren, proktorierte Prüfungen, Closed-Book-, Cheat-Sheet-, Open-Book- oder Open-Web-Klausuren. Die Begriffe heben jeweils bestimmte Eigenschaften von E-Assessments hervor, zum Beispiel die Art der Täuschungsüberwachung oder die Durchführungsbedingungen. Eine ausführlichere Besprechung auch dieser Termini erfolgt in späteren Abschnitten des vorliegenden Kapitels.

3 Formative E-Assessments

Formative E-Assessments, auch als digitale Tests oder elektronische Tests (E-Tests) bezeichnet, sind alle elektronischen Leistungsprüfungen, die den Lernprozess von Studierenden über einen längeren Zeitraum hinweg begleiten und mit Hilfe von wiederholten Ergebnisrückmeldungen verbessern sollen (Ehlers, Guetl, Höntzsch, Usener und Gruttmann 2013). E-Tests können dabei durchaus studienverlaufsrelevante Wirkung haben, indem ihre Absolvierung zum Beispiel als Zulassungsvoraussetzung für eine Prüfungsanmeldung oder zur Vergabe von Bonuspunkten auf eine spätere Klausur genutzt wird. Traditionell zählen zu den semesterbegleitenden Leistungsprüfungen die Hausübungen in der Mathematik, an deren Erfüllung oftmals die Zulassung zur Klausuranmeldung geknüpft ist.

Formative E-Assessments haben enge Verbindungen zum Konzept des *Mastery Learning* (Bloom 1974) und *Mastery Testing* (Kulik und Kulik 1987). Studierende werden hierbei durch kontinuierliche Tests entlang des Lernverlaufs und das daraus abgeleitete Feedback so lange auf eine summative Kompetenzprüfung vorbereitet, bis sie ein Kompetenzniveau erreicht haben, dass eine erfolgreiche Teilnahme verspricht. E-Tests führen diese Idee ins Digitale. Unterschiedliche Voraussetzungen auf Seiten der Lernenden werden durch E-Tests erfasst, Kompetenzdefizite rückgemeldet und so ein zielgerichteter Lernprozess ermöglicht. Grundsätzlich unterscheiden sich formative E-Assessments hinsichtlich der denkbaren Prüfungsgestaltung nicht von summativen

E-Assessments. Zur Gestaltung von formativen E-Assessments kommt eine Vielzahl von Prüfungsvarianten in Betracht, darunter Präsentationen, Disputationen, geschlossene, kognitive Testaufgaben, Hausarbeiten, Planspiele oder Portfolios (Huber und Reinmann 2019).

Allerdings gehört zu den konstituierenden Merkmalen der meisten E-Tests, dass sie wiederholt während des Lernverlaufs der Studierenden durchgeführt werden, oft wöchentlich oder in noch kürzerem Takt. Dies verursacht einen hohen Ressourcenaufwand für Korrektur und Feedback-Gabe. Daher greifen Lehrende zur effizienten Durchführung von E-Tests meist auf Aufgabenformate zurück, die eine einfache automatische Korrektur und Ergebnisrückmeldung erlauben. Dazu gehören zum einen die genannten geschlossenen Aufgabenformate wie etwa Antwort-Wahl-Aufgaben im Multiple oder Single Choice Design (Butler 2018), Zuordnungs- und Klassifikationsaufgaben und zum anderen halboffene Aufgabenformate wie Ergänzungsaufgaben und Lückentexte. E-Tests, die aus solchen Aufgabenformaten zusammengesetzt sind, werden häufig als Quizzes bezeichnet.¹

3.1 Empirische Evidenz am Beispiel von Quizzes

Die empirische Befundlage zur Wirksamkeit von Quizzes im Hochschulkontext ist reichhaltig und überwiegend positiv. Die Absolvierung von Quizaufgaben durch Studierende während der Lernphase verbessert signifikant die Prüfungsleistungen (Gholami und Moghaddam 2013) und erhöht den Wissensbehalt auch über längere Zeiträume (Roediger, Agarwal, McDaniel und McDermott 2011). Erwartungsgemäß gilt dies umso stärker, je höher die Ähnlichkeit der adressierten Kompetenzen zwischen den Quizfragen und den späteren Prüfungsfragen ist (Nguyen und McDaniel 2015), allerdings lässt sich eine generell lernförderliche Wirkung auch bei deutlich unterschiedlichen Kompetenzniveaus messen. Diese übergreifend lernförderliche Wirkung von Quizfragen geht zum Teil auf den lernpsychologischen »Testing Effect« zurück (McDaniel, Anderson, Derbish und Morrisette 2007). Allein das Ablegen eines Tests über gelerntes Material fördert das spätere Lernen und Behalten des Materials in einer abschließenden Prüfung, und zwar auch dann, wenn die Testergebnisse oder die korrekten Antworten auf die Testfragen vor dem Ablegen der abschließenden Prüfung nicht mitgeteilt wurden. Darüber hinaus erhöht die Absolvierung von Quizfragen auch sekundäre Ergebniskriterien wie Zufriedenheit, Engagement und Selbstwirksamkeits-erwartung, aber auch den Workload bei Studierenden (DeSouza und Fleming 2003). Gleichzeitig scheinen sie eine wichtige Vorbedingung für die Wirksamkeit von Blended Learning-Formaten zu sein, deren lernförderliche Wirkung ohne die Flankierung durch Quizfragen nur deutlich schwächer ausgeprägt ist (Spanjers, Könings, Leppink, Verstegen, de Jong, Czabanowska et al. 2015).

¹ Eine interaktive Entscheidungshilfe zur Auswahl von innovativen Prüfungsformen und E-Assessments liefert die Assessment Toolbox der Universität Bern (<https://assessment.unibe.ch/TestingFormats>) [10.03.2021].

4 Summative E-Assessments

Die wichtigste Funktion summativer E-Assessments ist es, das Kompetenzniveau von Studierenden zu einem festgelegten Zeitpunkt bilanzierend zu erfassen und die Ergebnisse gegenüber Stakeholdern wie der Hochschule oder Arbeitgebern zu dokumentieren. Summative E-Assessments sind retrospektiv und insofern final, als von dem bis zu einem bestimmten Zeitpunkt entwickelten Kompetenzniveau auf zukünftige Leistungen geschlossen und auf dieser Basis über den Zugang zu sozioökonomischen Ressourcen entschieden wird. Dem summativen Prüfen kommt somit in erster Linie eine Selektions- beziehungsweise Sozialisierungsfunktion zu (Flechsig 1976).

4.1 Closed-Book-, Cheat-Sheet-, Open-Book- und Open-Web-Prüfungen

Im Zuge der Verbreitung summativer E-Assessments haben sich offenere Bedingungen der Prüfungsdurchführung weiter etabliert, die über traditionelle Closed-Book-Prüfungen (CBP) hinausgehen. Bei Closed-Book-Prüfungen können Studierende während der Prüfungsteilnahme auf keinerlei eigene Materialien zurückgreifen. Alle zulässigen Materialien und Hilfestellungen wie zum Beispiel Formelsammlungen werden von der Prüfungsaufsicht gestellt. Offenere Prüfungsbedingungen wurden bereits in den 1960er Jahren diskutiert und erprobt (Feldhusen 1961). Die wesentlichen Formen sind Cheat-Sheet-Prüfungen (CSP), Open-Book-Prüfungen (OBP) und Open-Web-Prüfungen (OWP). Die drei Formen unterscheiden sich im Grad, zu dem selbst erstellte oder mitgeführte Hilfsmittel während der Klausurteilnahme von Studierenden benutzt werden dürfen. In typischen Cheat-Sheet-Prüfungen dürfen Studierende eine meist begrenzte Anzahl von Hilfszetteln in der Klausur verwenden, oft als Spickzettel bezeichnet. Lehrende können an solche Hilfszettel sowohl in Form als auch Inhalt weitgehend beliebige Vorgaben stellen. In der Praxis begegnet man dabei häufig einer Begrenzung der Seitenzahl und -größe (beispielsweise »drei beidseitig beschriebene DIN A4 Zettel«) oder auch der Bedingung, dass die Zettel handgeschrieben sein müssen. Damit soll gewährleistet werden, dass die Studierenden diese Zettel selbst erstellen und somit bestenfalls bereits beim Erstellvorgang Lerneffekte eintreten (Mathew 2012). Open-Book-Prüfungen gehen einen Schritt weiter. Hier sind nicht nur wenige, von Studierenden oft selbst zu erstellende Materialien als Hilfestellungen während der Prüfungsteilnahme erlaubt, sondern ganze Skripte oder Bücher. Ob die Menge erlaubter Dokumente auf eine konkrete Auswahl von Werken beschränkt ist oder beliebige Lehrbücher in beliebiger Menge mitgeführt werden dürfen, ist vom konkreten Einzelfall abhängig. In Open-Web-Prüfungen sind schließlich vollständig beliebige Hilfsmaterialien und insbesondere auch die Nutzung des Internets während der Prüfungsteilnahme erlaubt, solange diese asynchroner Natur sind. Ein Austausch zwischen Studierenden während der Prüfung, zum Beispiel über soziale Netzwerke oder Filesharingdienste, hat synchronen Charakter und wäre demzufolge untersagt.

Cheat-Sheet-Prüfungen und insbesondere die Open-Book- beziehungsweise Open-Web-Prüfungen betonen die Fähigkeit, Wissen unter authentischen Bedingungen zu lokalisieren, zu differenzieren und anzuwenden (Williams und Wong 2009). Zudem reduzieren sie die Bedeutung des Auswendiglernens für die Prüfungsleistung. Die Eignung

solcher Prüfungsbedingungen entfaltet sich deshalb vor allem auf höheren Kompetenzebenen etablierter Lernzieltaxonomien, wo nicht mehr reines Faktenwissen durch Nachschlagen oder eine Internetrecherche produziert werden kann (Imrie 1995).

An offene Prüfungsbedingungen werden zahlreiche und zum Teil konfligierende hochschuldidaktische Erwartungen geknüpft (Gupta 2007). Open-Book-Prüfungen sollen das Angsniveau von Studierenden senken, weil sie in der Verfügbarkeit von Referenzmaterialien Beruhigung finden. Zugleich kann der thematische Umfang einer Open-Book-Prüfung breiter sein, weil von den Prüflingen erwartet wird, dass sie alle Informationen in den Referenzquellen recherchieren können und nicht nur die vernünftigerweise erinnerbaren Kernaspekte reproduzieren könnten. Dies soll einen zusätzlichen Anreiz für ein gründliches Studium der verfügbaren Materialien liefern und somit die inhaltliche Beschäftigung mit den Prüfungsthemen intensivieren. Open-Book- und Open-Web-Prüfungen können somit auch realistischere Prüfungsfragen zulassen, die spätere Einsatzszenarien in Studium oder Beruf besser abbilden. Studierende werden zudem auf die akademische Arbeitspraxis vorbereitet, die weniger ein Auswendiglernen von Fachinhalten erfordert als vielmehr ein kompetenten Umgang mit Informationsquellen. Zudem besteht bei Open-Book- und Open-Web-Prüfungen aus hochschuldidaktischer Sicht die Hoffnung, dass Studierende bei der Prüfungsvorbereitung stärker auf höhere kognitive Fähigkeiten fokussiert sind und gegebenenfalls sogar ihre Lernstrategien anpassen, um die Anforderungen offener Prüfungen besser erfüllen zu können (Theophilides und Koutselini 2000).

Auf der Gegenseite stehen diverse Herausforderungen, die mit Open-Book- oder Open-Web-Prüfungen verknüpft sind. Zunächst wird von jedem Prüfling erwartet, dass ein dauerhafter Zugang zu denselben Referenzquellen besteht. Auch unterschätzen Studierende zum Teil deutlich die Zeit, die sie benötigen, um während der Prüfung in den Referenzmaterialien zu recherchieren. In Verbindung mit der geringeren Motivation für Studierende, die Materialien vor der Prüfung zu organisieren kann es dazu kommen, dass Studierende sich stark auf die Referenzmaterialien verlassen und sich im Extremfall während der Prüfung zum ersten Mal intensiver mit den Referenzmaterialien befassen. Auf der anderen Seite des Spektrums kann die Arbeitsbelastung bei denjenigen Studierenden erheblich ansteigen, die viel Zeit für die Vorbereitung und Auswahl des Referenzmaterials aufwenden. Im Übrigen erhöhen offene Prüfungsbedingungen auch den Arbeitsaufwand der Aufsichtspersonen, wenn sichergestellt werden muss, dass die Studenten während der Teilnahme nur die zulässigen Materialien verwenden.

4.2 Empirische Befundlage zu offenen Prüfungsbedingungen

Diese theoretischen Erwartungen an offene Prüfungsbedingungen werden inzwischen von einer umfassenden empirischen Befundlage komplementiert. Die Ergebnisse widersprechen zum Teil den Erwartungen und sind gerade für Lehrende teilweise überraschend. Zunächst steigt die prüfungsdidaktische Qualität von Klausuren mit wachsendem Grad der Offenheit an (Msila 2014). Lehrende verlagern abzuprüfende Lernziele auf höhere Lernziellebenen, um eine direkte Recherche zu verhindern. Somit gelingt es durch eine Veränderung der bloßen Prüfungsbedingungen, eine kompetenzorientier-

tere Gestaltung der Prüfung durch Lehrende zu erreichen. Die diagnostische Qualität der Prüfungen ist vom Grad der Offenheit unberührt. Es gelingt sowohl in Closed-Book-Prüfungen als auch in offeneren Formaten, die zu erhebenden Kompetenzen entlang psychometrischer Gütekriterien wie Reliabilität und Validität mit vergleichbarer Güte zu messen (Heijne-Penninga, Kuks, Schönrock-Adema, Snijders und Cohen-Schotanus 2008).

Studierende wenden für Closed-Book-Prüfungen tendenziell längere Vorbereitungszeiten auf, die zudem durch intensivere Nutzung der Lernmaterialien gekennzeichnet ist. Entsprechend ist die Lernzeit für Closed-Book-Prüfungen signifikant höher als für offene Formate (Durning, Dong, Ratcliffe, Schuwirth, Artino, Boulet, et al. 2016). Lernstrategien im kognitiven oder motivationalen Bereich sind aber zwischen den verschiedenen Formaten nicht systematisch verschieden. Allein die Tatsache, dass eine offene Prüfungsform gewählt wird, greift also nicht maßgeblich in das Lernverhalten der Studierenden ein (ebd.). Weitgehend übereinstimmende Evidenz liegt hinsichtlich der Prüfungsleistungen vor: Studierende erzielen in Closed-Book-Prüfungen tendenziell bessere Ergebnisse als in offenen Formaten. Dies ist nicht allein ein Effekt längerer Vorbereitungszeit für Closed-Book-Prüfungen, sondern auch darauf zurückzuführen, dass Studierende oft keinerlei Erfahrung im Umgang mit Open-Book- oder Open-Web-Prüfungen haben. Bucharbeit beziehungsweise die Recherche von Informationen während einer stressbelasteten Situation müssen geübt werden und verursachen bei unerfahrenen Prüflingen oftmals mehr Aufwand als sie Nutzen bringen (Myyry und Joutsenvirta 2015).

Die von vielen Lehrenden erwartete affektive Erleichterung durch offene Prüfungsbedingungen findet in begrenztem Umfang statt. Es finden sich messbare, aber quantitativ eher geringe Unterschiede bei Stressbelastung und Prüfungsangst zwischen Closed-Book-Prüfungen auf der einen Seite und Cheat-Sheet-, Open-Book- beziehungsweise Open-Web-Prüfungen auf der anderen Seite (Baillie und Toohey 1997). Gleichwohl ist die Anwesenheitsquote in Open-Book-Prüfungen höher als in Closed-Book-Prüfungen, bei denen deutlich mehr Absagen durch Studierende erfolgen (Francis 1982). Die Bearbeitungsdauer bei Open-Book ist aber im Mittel höher (Weber, McBee und Krebs 1983). Studierende verbringen zwischen 5 % und 10 % ihrer Klausurzeit mit der Recherche von Informationen. Die aufgewendete Zeit hat oftmals aber keinen Zusammenhang mit der erzielten Note (Westerkamp et al. 2013). Dennoch werden Cheat-Sheet-, Open-Book- und Open-Web-Prüfungen von Studierenden klar bevorzugt, da sie sich Vorteile wie die Reduktion der Stressbelastung während der Prüfung versprechen – die von der empirischen Evidenz nicht gedeckt sind.

Verengt man den Fokus auf Cheat-Sheet-Prüfungen, reproduziert sich die heterogene Evidenzlage. Entgegen der Erwartung vieler Lehrender stellen sich bei Cheat-Sheet-Prüfungen keine systematischen Leistungssteigerungen gegenüber Closed-Book-Prüfungen ein (Gharib, Phillips und Mathew 2012). Dieser Befund ist nicht zuletzt deshalb erstaunlich, weil die Vorbereitungszeit für Cheat-Sheet-Prüfungen tendenziell höher ist, da sie ja auch noch die Vorbereitung der Hilfszettel umfasst. Die Hoffnung, dass durch die eigene Erstellung der Hilfszettel eine tiefere Verarbeitung der Lerninhalte stattfindet, die zu einer Verbesserung der Prüfungsleistung führt, wird trotz höherer Lernzeit nicht erfüllt. Negative affektive Wirkungen wie

die Stressbelastung und Prüfungsangst sind bei Cheat-Sheet-Prüfungen zwar leicht reduziert im Vergleich zu Closed-Book-Prüfungen, aber höher als bei Open-Book- und Open-Web-Prüfungen. Schließlich scheint das Ausmaß von Täuschungsfällen bei Open-Book- und Open-Web-Prüfungen nicht höher zu sein als in Closed-Book-Prüfungen, obschon der Umfang zulässiger Materialien das Verbergen nicht erlaubter Hilfsmittel vereinfacht (Gharib und Phillips 2013).

4.3 Begriffsklärung: Fernprüfungen und Take-Home-Prüfungen

Fernprüfungen und Take-Home-Prüfungen sind summative Leistungskontrollen, welche in einem vorgegebenen Zeitintervall ortsunabhängig und unter frei wählbaren Arbeitsbedingungen durchgeführt werden. In der Literatur werden die Begriffe der Fernprüfungen und Take-Home-Prüfungen oftmals nicht deckungsgleich verwendet. Bei Fernprüfungen wird üblicherweise angenommen, dass eine Prüfungsaufsicht die Durchführung der Prüfung überwacht, zum Beispiel mit Hilfe einer Videoübertragung. Bei Take-Home-Prüfungen hingegen ging man ursprünglich davon aus, dass eine Prüfungsaufsicht während der Durchführung gerade nicht zugegen ist und auch sonst keine dezidierten Maßnahmen zur Täuschungsüberwachung ergriffen werden. Die Studierenden absolvieren die Prüfung zum gegebenen Zeitpunkt an einem frei wählbaren Ort ohne Live-Überwachung. An vielen Hochschulen im deutschsprachigen Raum ist der Begriff der Take-Home-Prüfung aber inzwischen zum Gattungsbegriff geworden, der sowohl Fernprüfungen (mit Überwachung) als auch beliebige andere Formen der ortsunabhängigen Leistungskontrolle umfasst. Für dieses Kapitel definieren wir Take-Home-Prüfungen deshalb als jede Form der zeit- und ortsflexiblen Leistungskontrolle, unabhängig davon, ob sie mit oder ohne Täuschungsüberwachung während der Durchführung stattfindet.²

4.4 Take-Home-Prüfungen im Allgemeinen

Take-Home-Prüfungen an Hochschulen sind international bereits seit längerer Zeit ein etabliertes Werkzeug der Kompetenzprüfung, insbesondere im Rahmen der beruflichen Weiterbildung und der Fernhochschulen (zum Beispiel Caruth und Caruth 2013). In Deutschland hat ihr Einsatz erst im Rahmen der Corona-Pandemie eine nennenswerte Verbreitung erreicht. Die zögerliche Umsetzung ist unter anderem dadurch begründet, dass die Durchführung von Take-Home-Prüfungen mit einer Reihe von Herausforderungen behaftet ist. Take-Home-Prüfungen müssen dem Gleichbehandlungsgrundsatz Rechnung tragen, der sich sowohl aus dem in Art. 3 Abs. 1 GG normierten allgemeinen Gleichheitssatz als auch aus anderen einfachgesetzlichen Bestimmungen ergibt. Es muss sichergestellt sein, dass bei allen Studierenden vergleichbare Bedingungen zur Prüfungsteilnahme herrschen. Sowohl bei der technischen wie auch der räumlichen Infrastruktur von Studierenden ist davon nur in seltenen Fällen auszugehen,

² Weiterführende Informationen liefert unter anderem das Themendossier »Prüfungen im Digitalen« des Hochschulforum Digitalisierung: <https://hochschulforumdigitalisierung.de/de/dossiers/pruefungen-im-digitalen> [10.03.2021].

allerdings sind solche Beschränkungen nach geltender Rechtsprechung kein Grund für die Gewährung eines Nachteilsausgleichs. Anders als bei gesundheitlicher Beeinträchtigung leitet sich aus dem Prüfungsrecht keine Lösung für den Umgang mit unterschiedlichen Prüfungsbedingungen auf Seiten der Studierenden ab. Hochschulen sind hier individuell gefordert, Maßnahmen zu entwerfen, die auf eine Homogenisierung der Voraussetzungen im Rahmen von Take-Home-Prüfungen zielen.

Aus prüfungsdidaktischer Sicht stellen Take-Home-Prüfungen keine größeren Herausforderungen an Form und Inhalt der Aufgabenformulierung als die klassische schriftliche Präsenzprüfung. Die Forschungslage zu Take-Home-Prüfungen selbst ist noch lückenhaft (Bengtsson 2019). Generell wird aber davon ausgegangen, dass Take-Home-Prüfungen häufig als Open-Book- oder Open-Web-Prüfungen ausgelegt sind, sodass eine Übertragbarkeit der damit assoziierten empirischen Befunde auf Take-Home-Prüfungen möglich erscheint. Lehrende betreiben auch bei Take-Home-Prüfungen erheblich höheren Aufwand bei der Erstellung von Aufgaben, um kompetenzgerechte Aufgabenstellungen zu entwickeln. Entgegengesetzt zum Aufwand für Lehrende ist die Vorbereitungsintensität von Studierenden für die Vorbereitung auf Take-Home-Prüfungen oftmals niedriger als für Prüfungen im schriftlichen Präsenzformat. Die Leistungsdauer ebenso wie die Leistungsgüte sind bei Take-Home-Prüfungen aber höher (Marsh 1984). Entsprechende Evidenz findet sich nicht nur in Feldstudien, sondern auch in kontrollierten experimentellen Untersuchungen, bei denen Studierende in überwachten Prüfungen schlechtere Leistungen erzielen als bei Take-Home-Prüfungen (Weber, McBee und Krebs 1983). Die Forschungsliteratur liefert Hinweise darauf, dass Ursachen dafür in der Nutzung unzulässiger Hilfsmittel im Rahmen von Take-Home-Prüfungen liegen könnten. Wird die Leistung von Studierenden, die zuvor entweder eine Präsenz- oder Take-Home-Prüfung absolviert haben, kurze Zeit später mit einem weiteren Test überprüft, der für beide Gruppen im Präsenzsetting durchgeführt wird, zeigen die Studierenden, die zuvor eine Take-Home-Prüfung zu bearbeiten hatten, deutlich schlechtere Leistungen als Personen, die sich auf eine Präsenzprüfung vorbereitet hatten (Marsh 1984).

4.4 Take-Home-Prüfungen als E-Assessments

Im Kontext von Take-Home-Prüfungen kommt der Nutzung von E-Assessments eine maßgebliche Bedeutung zu. Hochschulen versuchen, den Schwund an Kontrolle und Standardisierung durch die Nutzung elektronischer Prüfungssysteme (E-Prüfungssysteme) wenigstens zum Teil auszugleichen. Zudem erleichtern E-Prüfungssysteme die Durchführung von Take-Home-Prüfungen in verschiedenen Phasen. Lehrende können mit Hilfe von E-Prüfungssystemen auf einfache Weise mehrere Varianten derselben Klausur herstellen, indem Aufgabenpools oder parametrisierbare Aufgaben genutzt werden. Darüber hinaus ist der Prozess der Verteilung der Prüfungsunterlagen und deren späterer Einsammlung weitgehend automatisiert. Sollen bearbeitete Prüfungsunterlagen elektronisch abgegeben werden, ist dies ohne die Nutzung von E-Prüfungssystemen ein fehlerträchtiger Prozess, der eine Reihe technisch anspruchsvoller Schritte umfasst. Studierende müssen aus ihren Prüfungsunterlagen digital versendbare Dokumente generieren. Dies ist gerade bei

papiergebundenen Take-Home-Prüfungen keine triviale Aufgabe, da sie die Digitalisierung der Prüfungsunterlagen mit einem Dokumentenscanner oder einem Smartphone erfordert. Dabei entstehen unter Umständen hohe Datenmengen, deren Versand fehlschlagen kann. Zudem wünschen Studierende bei Take-Home-Prüfungen eine unmittelbare und verlässliche Information darüber, dass sie ihre bearbeitete Prüfung ordnungsgemäß abgegeben haben. Die Nutzung eines dezidierten E-Prüfungssystems vereinheitlicht den Prozess und schafft so Prüfungssicherheit sowohl für Lehrende als auch Studierende bei der Durchführung von Take-Home-E-Assessments.

5 Umgang mit Täuschungen

Normverletzendes Verhalten von Studierenden wie zum Beispiel Schummeln stellt die größte Herausforderung für die Einführung von Take-Home-Prüfungen dar (Tao und Li 2012). Das International Center for Academic Integrity (ICAI) geht davon aus, dass 40-60 % der Studierenden mindestens einmal im Rahmen einer schriftlichen Prüfung getäuscht haben.³ Zu den beliebten Täuschungsmethoden gehören die Nutzung unerlaubter Materialien, technische Maßnahmen zur Kompromittierung der Prüfungssituation (»Hacking«) sowie verabredete Kollaboration (Hellas, Leinonen und Ihantola 2017). In Tabelle 1 findet sich eine systematische Aufbereitung und Kategorisierung möglicher Täuschungsszenarien, gemeinsam mit einer Aufwandsschätzung und der Nennung von Möglichkeiten der Überwachung und Entdeckung von Täuschungen beziehungsweise Sicherstellung von Beweismitteln. Die zentrale Forderung vieler Lehrender bezieht sich deshalb auf die Bereitstellung von hinreichend effektiven Maßnahmen bei Take-Home-Prüfungen a) zur Täuschungsvermeidung, b) zur Täuschungsüberwachung und c) zur Täuschungsentdeckung. Täuschungsvermeidung bezeichnet Vorkehrungen, die es Studierenden erheblich erschweren, erfolgreiche Täuschungsversuche überhaupt zu unternehmen. Täuschungsüberwachung meint die Beobachtung der Studierenden während der Prüfungsdurchführung, in der Regel unterstützt durch eine Prüfungsaufsicht. Maßnahmen zur Täuschungsentdeckung schließlich machen es der Prüfungsaufsicht einfacher, Täuschungsversuche während ihres Auftretens oder im Nachhinein aufzudecken und zu belegen.

5.1 Täuschungsvermeidung

E-Assessments bieten vielfältige Optionen zur Täuschungsvermeidung. Ihr gemeinsames Merkmal ist es, dass sie den erwartbaren Nutzen von Täuschungsmanövern stark einschränken. Ein von vielen Lehrenden verwendetes Mittel ist die Adressierung bestimmter Kompetenzebenen in den Aufgabeninhalten, mit denen Täuschungsversuche vermieden werden sollen. Hier können mehrere Ansätze gewählt werden (Von Gruenigen, Souza, Pradarelli, Magid, und Cieliebak 2018). Auf der einen Seite können Lehrende versuchen, ihre Aufgaben auf höheren Kompetenzebenen zu formulieren, die Interpretations- oder Transferleistungen erfordern, sodass ein reines Abschreiben nicht

³ <https://www.academicintegrity.org/statistics/> [10.03.2021].

zu einer nennenswerten Verbesserung der Prüfungsleistung führen kann beziehungsweise schnell bemerkt würde. Auf der Gegenseite steht der Versuch, möglichst niedrige Kompetenzebenen wie die reine Wissensproduktion anzusprechen und dafür den zur Verfügung stehenden Zeitrahmen auf ein absolutes Minimum zu verkürzen. Solche Varianten der Hochgeschwindigkeitsprüfung sind aus prüfungsdidaktischer Sicht für die meisten Anwendungsfälle im Hochschulkontext ungeeignet, da sie höherwertige Lernzielebenen naturgemäß außer Acht lassen. Eine zweite prominente Methode der Täuschungsvermeidung ist die Individualisierung von Prüfungsaufgaben oder ganzer Prüfungen. Sie kann händisch oder im Fall von E-Assessments vollautomatisch durch die Erstellung mehrerer Varianten derselben Klausur erfolgen. Dazu werden entweder einzelne Aufgaben in einer Weise parametrisiert, dass Teile der Aufgabenformulierung oder das zur Aufgabe gehöriges Datenmaterial mit einem Zufallsalgorithmus durch das E-Prüfungssystem automatisch generiert werden. Diese Parametrisierung erfolgt nach vorgegebenen Regeln, sodass der Variantenreichtum kontrolliert und unterschiedliche Schwierigkeitsgrade durch ungünstige Zufallsauswahlen vermieden werden können. Überdies kann die Zusammensetzung der Prüfung an sich automatisch gewählt werden. Aus einem umfangreichen Aufgabenpool wird dazu eine Stichprobe von Aufgaben gezogen, sodass möglichst keine Prüfung für dieselbe Kohorte identische Aufgaben enthält. Das Kopieren der Lösungen anderer wird damit erheblich erschwert. Ferner erlaubt der Einsatz datenanalytischer Verfahren auf den Prüfungsergebnissen das (teil-)automatische Aufdecken unerlaubter Zusammenarbeit zwischen Studierenden (Li, Luo, Sikdar, Nizam, Gao, Shan, et al. 2019). Sowohl bei der Parametrisierung als auch bei der Pool-basierten Prüfungserstellung muss streng auf eine Vergleichbarkeit zwischen den erstellten Klausur Varianten geachtet werden. Dies betrifft nicht nur psychometrische Eigenschaften einer Klausur wie den mittleren Schwierigkeitsgrad aller Aufgaben, die Verteilung der Schwierigkeitsgrade zwischen den Aufgaben oder die Dauer der Aufgabenbearbeitung, sondern auch prüfungsdidaktische Aspekte wie die Abdeckung derselben Lernziele. Im Übrigen ist die vielfach verbreitete Abgabe von Ehrenklärungen eher wenig effektiv, wie eine Vielzahl von Täuschungsskandalen aus dem internationalen Raum auch an hochreputierten Hochschulen zeigt.⁴ Auch Studierende stehen der Durchführung von Take-Home-E-Assessments ohne weitergehende Maßnahmen zur Täuschungsüberwachung äußerst kritisch gegenüber (Jervis und Brown 2020).

5.2 Täuschungsüberwachung

Die Maßnahmen zur Täuschungsüberwachung werden häufig mit dem Begriff des Proctoring zusammengefasst. Dieser meint nicht mehr als die Durchführung einer

4 Zum Beispiel Harvard University (https://en.wikipedia.org/wiki/2012_Harvard_cheating_scandal) [10.03.2021], Duke University (<https://www.nytimes.com/2007/05/01/us/01duke.html>) [10.03.2021], Stanford University (<https://www.mercurynews.com/2015/03/27/stanford-university-looks-into-allegations-of-cheating-by-students/>) [10.03.2021], University of Sidney (<https://www.smh.com.au/cqstatic/givnzb/Taskforce.pdf>) [10.03.2021].

Prüfungsaufsicht zum Zwecke der Gewährleistung akademischer Integrität. In schriftlichen oder mündlichen Präsenzprüfungen erschöpft sich das Proctoring zumeist in der Anwesenheit einer oder mehrerer Aufsichtspersonen. Bei Take-Home-Prüfungen ist diese Prüfungsaufsicht im Vergleich zum Präsenzsetting deutlich erschwert, da kein direkter Zugriff auf den physischen Arbeitsbereich der Studierenden besteht. Mit Methoden des Online-Proctoring wird hier versucht, eine effektive Aufsicht der Prüflinge aus der Ferne zu ermöglichen. Online-Proctoring lässt sich in vier verschiedene Stufen gliedern, die durch einen steigenden Grad an Eingriff in die Privatsphäre der Studierenden gekennzeichnet sind (Sietses 2016).

Ab der zweiten Stufe kommt dabei spezialisierte Software zur Durchführung des Proctoring zum Einsatz. Solche Online-Proctoring-Systeme bieten Schnittstellen zu gängigen E-Prüfungssystemen, sodass spätestens auf dieser Ebene der Einsatz von E-Assessments als Format für Take-Home-Prüfungen dominiert.⁵ Proctoring Software ermöglicht Aufzeichnungen in verschiedenster Form, darunter Video- und Audioaufzeichnungen, Bildschirmfotos oder -videos, die Übersicht auf dem PC der Prüflinge laufender Prozesse sowie ein Logging aller Nutzerinteraktionen wie Maus- und Tastaturlklicks. In vielen Online-Proctoring-Systemen werden Überwachungsalgorithmen aus dem Bereich maschinellen Lernens eingesetzt, um auffälliges Verhalten der Studierenden zu entdecken und entsprechende Hinweise an die Prüfungsaufsicht zu geben. Typische Anwendungen sind die Auswertung von Augenbewegungen oder der Abgleich von Klickgeräuschen, die das Computermikrofon aufzeichnet, mit den tatsächlichen Eingaben. Zusätzlich kann ein Lockdown des Computers erfolgen, sodass unzulässige Aktivitäten wie das Besuchen bestimmter Webseiten oder der Start von Anwendungen unterbunden werden. Entsprechend tief muss die Software in den PC der Prüflinge eingreifen.

Online-Proctoring kann dabei grundsätzlich asynchron oder synchron erfolgen. Beim asynchronen Proctoring findet während der Prüfungsteilnahme keine Überwachung durch eine Prüfungsaufsicht statt. Stattdessen wird durch das Online-Proctoring-System eine Aufzeichnung durchgeführt und persistent gespeichert, die ex post durch Aufsichtspersonen auf Täuschungsversuche begutachtet wird, oft unterstützt durch automatische Algorithmen zur Täuschungsentdeckung, die auffälliges Verhalten der Studierenden in einer Timeline markieren und so das Auffinden erleichtern. Die Ahndung eines festgestellten Täuschungsversuchs erfolgt hier erst nach dem Ablegen der Prüfung. Synchrones Proctoring oder Live-Proctoring kommt ohne eine persistente Speicherung der Aufzeichnung einer Prüfungsdurchführung aus. Hier ist während der Abnahme einer Prüfung eine menschliche Prüfungsaufsicht zugegen, die durch automatisierte Algorithmen des Online-Proctoring-Systems unterstützt wird. Entdeckt die menschliche Prüfungsaufsicht einen Täuschungsversuch oder signalisiert die Software einen entsprechenden Verdacht, kann live im aufgezeichneten Material zurückgespult und das verdächtige Verhalten geprüft werden. Die Feststellung des Täuschungsversuchs findet hier noch während der Prüfungsdurchführung statt.

⁵ Eine Übersicht kommerzieller und nicht-kommerzieller Proctoring-Software liefert <https://docs.google.com/spreadsheets/d/1TQcwXozYqwocmHoSUQNeoR71opUWj9WtXnmBWRH9e48/> [10.03.2021].

Auf der ersten Stufe des Proctoring steht die Videoüberwachung von Prüflingen in Videokonferenzen mit maximal 10 bis 15 Studierenden unter Verwendung nur einer Videokamera. Eine verlässliche Täuschungüberwachung ist in solchen Szenarien nur schwer möglich. Studien mit studentischen Probandinnen und Probanden an verschiedenen Hochschulen, darunter an der TU München und an der RWTH Aachen haben festgestellt, dass Täuschungsversuche der oftmals gut vorbereiteten Studierenden zu einem guten Teil selbst bei aufmerksamer Prüfungsaufsicht unbemerkt bleiben.⁶ Die zweite Stufe des Online-Proctoring umfasst synchrones oder asynchrones Proctoring mit der Aufzeichnung einer Videokamera sowie des Bildschirms der Prüflinge. Zusätzlich werden die besuchten Webseiten geloggt. Auf der dritten Stufe findet zusätzlich ein Logging der laufenden Applikationen und ein Lockdown des Computers der Prüflinge statt. Auf der höchsten Stufe werden zwei Kameras für die Videoüberwachung eingesetzt (zum Beispiel einmal die Webcam des Computers und einmal eine rückwärtig positionierte Smartphone-Kamera). Zusätzlich erfolgt ein vollständiges Aktivitäts-Logging, das sämtliche Nutzerinteraktionen mit dem Computer und alle Systemprozesse aufzeichnet.

5.3 Täuschungsentdeckung

Vorkehrungen zur Täuschungsentdeckung sollen es der Prüfungsaufsicht erleichtern, während der Prüfungsdurchführung unbemerkt gebliebene Täuschungsversuche aufzudecken. In diesen Bereich fallen unter anderem die weiter oben beschriebenen Maßnahmen zur maschinellen Entdeckung von Täuschungsversuchen durch Proctoring-Software. Auch die Herstellung paralleler Varianten einer Prüfung ist ein gängiges Mittel, um feststellen zu können, ob abgegebene Prüfungsunterlagen schlichtweg physisch kopiert wurden. Dies kann zum einen durch die Erstellung verschiedener ähnlicher Aufgabenvarianten geschehen, sodass der Inhalt der Bearbeitung durch einen Studierenden/eine Studierende anzeigt, auf welche der Varianten geantwortet wurden. Zum anderen bietet sich die Individualisierung nicht-inhaltsrelevanter Gestaltungsmerkmale der Klausurmaterialien. Dies können leichte Veränderungen der Abstände oder Formulierungen sein, anhand derer das physische Kopieren von Lösungen festgestellt werden kann. Auch die Einbettung eines versteckten Identifikationscodes in den digitalen Aufgabendokumenten ist möglich. Zusätzlich kommt die Verwendung von Plagiatserkennungssoftware in Betracht, welche die abgegebenen Aufgabenbearbeitungen automatisiert mit einem großen Pool von Quellen vergleicht. Hier ist von zentraler Bedeutung, dass die Software nicht nur gegen Quellen aus dem Internet prüft, sondern auch innerhalb der abgegebenen Klausurmaterialien, um sowohl das Plagiieren öffentlich verfügbarer Lösungen als auch das Abschreiben unter den Prüflingen aufdecken zu können.

6 Siehe unter anderem <https://hochschulforumdigitalisierung.de/de/blog/online-proctoring> [31.05.2021]; https://video.cls.rwth-aachen.de/wp-content/uploads/2021/02/Analyse_Taeuschungsszenarien_Pruefungen_published.pdf [10.03.2021].

5.4 Rechtliche Bewertung

Insbesondere der Aspekt der Täuschungsüberwachung bei Take-Home-Prüfungen greift in mehrere Rechtsgebiete ein, darunter das Datenschutzrecht, Persönlichkeitsrechte wie das Recht am eigenen Bild und Wort sowie das Urheberrecht. Zum aktuellen Zeitpunkt im April 2021 ist die rechtliche Situation rund um die Täuschungsüberwachung an deutschen Hochschulen weiterhin unübersichtlich, insbesondere mit Bezug auf die Zulässigkeit verschiedener Formen des Online-Proctoring. Einige Bundesländer haben Sonderverordnungen erlassen, die den Rechtsrahmen für überwachte Take-Home-Prüfungen deutlich erweitern. Ferner haben sich verschiedene Gerichte im Rahmen von Eilverfahren mit der Täuschungsüberwachung im Rahmen digitaler Prüfungen beschäftigt, darunter das OVG Münster (Az. 14 B 278/21.NE) sowie das OVG Schleswig (Az. 3 MR 7/21). Die Beschlüsse geben erste Hinweise darauf, wie die Gerichte die Rechtslage in Hauptsacheverfahren beurteilen könnten. Die Einschätzungen ähneln dem Gutachten der Rechtsinformationsstelle der Digitalen Hochschule NRW (RiDHnrw) aus dem Jahr 2020. (Albrecht und Fischer 2020). Fachexperten gehen deshalb davon aus, dass die Hochschulen bereits auf kurze Sicht eine deutliche Erweiterung ihrer Möglichkeiten erhalten werden, Take-Home-Prüfungen und die erforderlichen Maßnahmen zum wirksamen Umgang mit Täuschungsversuchen rechtssicher umsetzen zu können.

6 Implementierung von E-Assessments

6.1 Realisierung mit E-Prüfungssystemen

Zur Durchführung von E-Assessments können vielfältige Software-Plattformen genutzt werden. Sämtliche etablierten Lernmanagementsysteme wie Moodle, Ilias oder Blackboard stellen Funktionalitäten für die Durchführung von E-Assessments mit verschiedenen Aufgabenformaten bereit. E-Assessments sind aber nicht auf dezidierte Systeme zur Durchführung elektronischer Prüfungen beschränkt. Auch Audience Response Systeme wie Mentimeter, Particify oder QuizAcademy bieten nutzenstiftende Plattformen für die Durchführung von E-Assessment, ebenso wie Umfrage-Plattformen wie Soscisurvey, Limesurvey oder Google Forms. Zur Verwaltung studentisch generierter E-Assessments bieten sich Plattformen wie PeerWise oder das Moodle Plugin Student-Quiz an. Komplexere Aufgabenformate lassen sich mit kollaborativen Plattformen wie Padlet, Miro oder IdeaPad implementieren. Auch kollaborative Entwicklungsumgebungen aus dem Statistik- oder Programmierbereich wie zum Beispiel Jupyter Notebooks eignen sich zur Durchführung von E-Assessments.

Sobald es aber um die rechtssichere Abnahme von E-Prüfungen im Sinne des Hochschulprüfungsrechts geht, benötigen Hochschulen E-Prüfungssysteme, die genau für diesen Zweck konfiguriert sind. Ein Lernmanagementsystem wie Moodle oder ILIAS mit den prinzipiell reichhaltigen E-Assessment Funktionen ohne weitere Anpassungen für diesen Zweck einzusetzen, ist kein zielführender Weg. Lernmanagementsysteme sind nicht notwendigerweise als E-Prüfungssysteme geeignet und benötigen zum

Teil umfassende Konfigurationsänderungen, um für die regelhafte Durchführung von E-Prüfungen verwendet werden zu können. So müssen E-Prüfungssysteme eine Reihe technischer und organisatorischer Maßnahmen implementieren, die maßgeblich durch das Datenschutzrecht und das Prüfungsrecht vorgegeben sind. Sie müssen eine revisionssichere Bereitstellung, Bearbeitung, Abgabe und Ausfallgeschützte Speicherung von Prüfungen ermöglichen. Neben der Datensicherheit müssen sie Prüfungsdaten geschützt ablegen, sodass nur autorisierte Personen Zugriff auf die Daten erlangen können. Entsprechend wird ein systematisches Pentesting (penetration testing) zur Gefahrenabwehr empfohlen, bevor ein E-Prüfungssystem in der Breite eingesetzt wird. E-Prüfungssysteme sollten zudem Snapshots der Prüfungs-durchführung anfertigen, zum Beispiel in Form regelmäßiger Screenshots, um bei einem Absturz des E-Prüfungssystems den bisherigen Prüfungsverlauf zweifelsfrei dokumentieren zu können. Sie sollten Schnittstellen zur verlässlichen Übertragung von Prüfungsergebnissen an das Campusmanagement bieten. Zudem sollten sie mit Blick auf aktuelle Entwicklungen im Bereich der Täuschungsüberwachung bei elektronischen Take-Home-Prüfungen Schnittstellen zu Online-Proctoring-Systemen bereitstellen und zusätzlich selbst Werkzeuge zur Täuschungsvermeidung bieten, zum Beispiel einen Browser-Lockdown während der Prüfungs-durchführung.⁷ Schließlich ist die Einplanung einer aktiven Weiterentwicklung unbedingt erforderlich, um der wachsenden Aufgabenvielfalt im Bereich der E-Assessments auch in Zukunft Rechnung tragen zu können.

E-Prüfungssysteme stellen in der Regel mehrere Aufgabenformate bereit. Zum Quasi-Standard zählen dabei zunächst die sogenannten geschlossenen Aufgaben wie Ja-/Nein-Fragen, Single- oder Multiple-Choice Fragen sowie Zuordnungsaufgaben, bei denen ein Satz von Antwortoptionen einer oder mehreren Antwortkategorien zugeordnet werden muss. Ferner bieten praktisch alle etablierten E-Prüfungssysteme die sogenannten halboffenen Aufgabenformate. Dazu zählen Lückentexte, Aufgaben zur Ergebniseingabe und Cloze-Aufgaben, bei denen eine meist aus nur einem Wort bestehende Antwort frei eingegeben werden kann. Offene Aufgabenformate sind Eingabemöglichkeiten für längere Texte oder auch die Upload-Möglichkeit für beliebige Dateiformate. Tabelle 2 gibt eine Übersicht der wichtigsten an deutschen Hochschulen eingesetzten kommerziellen und nicht-kommerziellen E-Prüfungssysteme mitsamt jeweils der enthaltenen Aufgabenformate.

6.2 Realisierung als alternative Prüfungsformen

Ein häufig geäußerter Vorwurf an E-Prüfungssysteme ist die eingeschränkte Aufgabenvielfalt. Insbesondere den geschlossenen und halboffenen Aufgabenformaten wird eine Eignung eher zur Erfassung von Kompetenzen auf niedrigen Lernzielebenen wie der Wissensreproduktion und Verständnisprüfung zugesprochen (Bender, Hilkenmeier und Schaper 2015). Die Prüfung höherer Lernzielebenen gestaltet sich mit den etablierten E-Prüfungssystemen (Tab. 2) derzeit noch schwierig. Es existieren erste Ansätze für konkrete Fachdisziplinen, auch komplexe Lernzielanforderungen

⁷ SafeExamBrowser (<https://safeexambrowser.org/>) [10.03.2021].

mit E-Prüfungssystemen abzuprüfen. Ein Beispiel ist die nahtlose Integration spezifischer Softwaretools wie statistische oder informatische Programmierumgebungen als Aufgabenformate in E-Prüfungssystemen, allerdings befinden sich solche Vorhaben aktuell erst im Entwicklungsstadium.

Zudem muss konstatiert werden, dass praktisch alle Lösungen zur Täuschungsvermeidung, -überwachung und -entdeckung im Rahmen von E-Assessments an ihre Grenzen stoßen, wenn während des Ablegens der Prüfung unbeteiligte Dritte eine individuelle Hilfestellung leisten oder sogar das komplette Absolvieren der Prüfung für den eigentlichen Prüfling übernehmen (Tab. 1). Selbst fortgeschrittene Methoden des Online-Proctoring können bei entsprechender technischer Expertise nicht verhindern, dass ein Prüfling die Prüfung nur zum Schein bearbeitet, während dritte Personen die eigentliche Prüfungsbearbeitung übernehmen.

Deshalb erfährt die Nutzung alternativer Prüfungsformen wachsende Aufmerksamkeit (Billerbeck, Barnat und Knutzen 2016). Zu den prominenten Vertretern alternativer Prüfungsformen zählen die digitalen Prüfungsportfolios, App-basierte E-Assessments, Serious Games, praktische Demonstrationen oder Präsentationsleistungen. Ihr kennzeichnendes Merkmal ist, dass Art und Umfang der Prüfungsleistungen systematisches Täuschungsverhalten unrealistisch machen. Alternative Prüfungsformen lösen überdies eine bereits seit langem formulierte Forderung an Prüfungskonzepte im Hochschulbereich ein. Sie zielen auf den Wechsel von zeitgebundenen summativen Prüfungen hin zu kontinuierlichen Leistungskontrollen, die formativ in den Semesterverlauf eingebettet sind und perspektivisch sogar komplett auf summative Prüfungen verzichten können (»embedded assessment«; Redecker und Johannessen 2013). Mit Hilfe von Leistungsdaten, die während des Semesters erhoben werden, findet hier eine kumulative Bewertung des Kompetenzniveaus und dessen Entwicklung bei den Studierenden statt, die eine abschließende Prüfung weitgehend überflüssig macht. Auf zwei alternative Prüfungsformen soll im Folgenden exemplarisch näher eingegangen werden.

6.2.1 Prüfungsportfolios

Prüfungsportfolios sind (digitale) Sammlungen von Leistungsartefakten, deren Zusammenstellung klaren Vorgaben folgt und die Erreichung festgelegter Lernziele sowie den Erwerb unterschiedlicher Kompetenzen dokumentiert (Lombardi 2008). Ursprünglich vor allem im erziehungswissenschaftlichen Bereich verbreitet (Barton und Collins 1993), finden sich Prüfungsportfolios inzwischen quer durch alle Fachdisziplinen, zum Beispiel Mathematik (Asturias 1994), Ingenieurwissenschaften (Christy und Lima 1998), Sprachkurse (Song und August 2002), Sportwissenschaften (Mohnsen und Mendon 1997) und Wirtschaftswissenschaften (Walstad 2001). Traditionell werden Portfolios ergänzend zu summativen Abschlussprüfungen eingesetzt. Dort ermöglichen sie insbesondere eine formative Steuerung des Kompetenzerwerbs und sollen somit eine bessere Vorbereitung auf die summative Kompetenzprüfung leisten (Dysthe und Engelsen 2004). Da Prüfungsportfolios explizit auf die Messung höherwertiger Lernzielebenen ausgerichtet und im Wesentlichen einer Take-Home-Prüfung im Open-Web-Format entsprechen, sind sie nicht auf die Abfrage recherchierbaren Wissens

ausgelegt, sondern entfalten ihr Potential vor allem auf höheren Kompetenzebenen (Davies und LeMahieu 2003). Zwar sind viele Leistungsartefakte von Prüfungsportfolios im Kern wieder Take-Home-Prüfungen mit den typischen Herausforderungen bei der Täuschungsüberwachung, aber alleine aufgrund des hohen Umfangs und der Vielfalt der enthaltenen Leistungsartefakte errichteten Prüfungsportfolios oftmals eine hohe Schwelle der Täuschungsvermeidung. Täuschungswillige Studierende müssen eine erhebliche Menge unzulässiger Ressourcen über einen längeren Zeitraum hinweg zum Einsatz bringen. Die dabei entstehenden stilistischen und inhaltlichen Brüche sind zudem oftmals leicht zu entdecken (Cizek 1999).

Tatsächlich gelingt es mit der Nutzung von Prüfungsportfolios in verschiedenen Szenarien, spätere summative Prüfungsleistungen zu verbessern (Händel, Wimmer und Ziegler 2020), allerdings findet sich dieser leistungsförderliche Effekt nur in einem Teil der Forschungsliteratur (siehe zum Beispiel Astwood und Slater 1997). Gründe dafür liegen vor allem in der didaktischen Anlage des Prüfungsportfolios. Der Grad der Auseinandersetzung der Studierenden mit den Aufgaben des Prüfungsportfolios und entsprechend die spätere Leistungsgüte sind unter anderem abhängig von Art und Qualität des Feedbacks und der Lernstrategien der Studierenden (Segers, Gijbels und Thurlings 2008). Positive Einstellungen von Studierenden gegenüber Prüfungsportfolios finden sich überdies oft erst nach intensiver Einführung des Formats der Prüfungsportfolios (Davis, Ben-David, Harden, Howie, Ker, McGhee et al. 2001). Studien weisen ferner darauf hin, dass die Beurteilung von Prüfungsportfolios durch Lehrende mit zufriedenstellender psychometrischer Qualität und einer guten Übereinstimmung mit späteren Abschlussnoten gelingt (Chang und Wu 2012).

6.2.2 E-Assessments mit Drittapplikationen

E-Assessments finden im Zuge des digitalen Wandels immer öfter unter Verwendung von Drittapplikationen statt, die zuvor auch in der Kompetenzvermittlung eingesetzt worden ist (Minocha 2009). Unter Drittapplikationen versteht man Software, deren primärer Zweck nicht das Durchführen von E-Assessments ist. Zu unterscheiden ist zwischen domänenpezifischer und universeller Software. Unter domänenpezifischer Software versteht man spezialisierte Software, die zur Lösung fachbezogener Anforderungen eingesetzt wird. Beispiele für solche Werkzeuge sind Geoinformationssysteme (GIS), Software aus dem Feld der statistischen Analyse, Computer Algebra Systeme (CAS), Konstruktionssoftware (CAD/CAM-Systeme) und Entwicklungsumgebungen für Programmiercode (Integrated Development Environment, IDE), aber auch stark spezialisierte Werkzeuge wie datenbankbasierte Abfragesprachen oder Software für Cloud Computing (Al-Samarraie und Saeed 2018). Universelle Software auf der anderen Seite sind unabhängig von der Fachdomäne einsetzbar. Im Bereich des digitalen Prüfens sind dies vor allem Office-Anwendungen.

Drittapplikationen realisieren die Forderung des Constructive Alignment nach einer engen Verzahnung von Lernzielen, Lernmethoden und Prüfungsvorgängen (Biggs und Tang 2011). Von zentraler Bedeutung bei der Auswahl geeigneter Drittapplikationen für E-Assessments ist ihre Authentizität für den Lern-, Forschungs- oder Arbeitsalltag (Jones und Bissell 2011). Je stärker die Verwendung der Software als Teil einer

Prüfung an ihren realweltlichen Einsatz angelehnt ist, desto höher ist die Passung von Prüfungssituation, Prüfungsanforderungen und Prüfungsaufgaben im Sinne des Constructive Alignment (Iller und Wick 2009). Gleichzeitig bringt der Einsatz von Drittapplikationen als Prüfungswerkzeuge verschiedene Herausforderungen mit sich (Halbherr, Dittmann-Domenichini, Piendl und Schlienger 2016). Die Applikationen müssen allen Studierenden zur Verfügung stehen, was zwar in Rechnerpools der Hochschule, nicht aber bei Take-Home-Prüfungen sicher gewährleistet werden kann. Die Software sollte automatisch Zwischenergebnisse speichern, um eine Wiederherstellbarkeit der Eingaben bei Abstürzen zu ermöglichen. E-Assessments mit Drittapplikationen sollten zudem als Open-Web-Prüfungen angelegt sein, da die Nutzung integrierter Hilfesysteme und oft auch der Weg ins Internet nicht vermieden werden können. Die Applikationen müssen den Export der Bearbeitungen als Dateien ermöglichen, wobei eine Versionskompatibilität zwischen den Installationen der Studierenden und der Prüfenden sicherzustellen ist. Und schließlich muss die technisch stabile und revisionssichere Versendung der exportierten Dateien an die Prüfenden ermöglicht werden. Diese Anforderungen zu erfüllen ist nicht in jedem Falle trivial. Trotzdem birgt die Nutzung von Drittapplikationen enorme Potentiale zur kompetenzorientierten Situierung von Prüfungen in sich stetig stärker digitalisierenden Lernwelten (Hillier und Fluck 2017).

7 Akzeptanz

Entscheidende Bedeutung bei der nachhaltigen Etablierung von E-Assessments kommt dem Grad an Akzeptanz bei den beteiligten Akteurinnen und Akteuren zu. Die Einführung neuer Prüfungsformen ist nicht selten von Skepsis begleitet, die didaktisch, technologisch, datenschutzrechtlich oder ethisch begründet sein kann. Die Technologieakzeptanzforschung hat eine Reihe empirisch gut validierter Modelle hervorgebracht, um die Akzeptanz der Einführung neuer Technologien sowohl bei Lehrenden (Kuikka, Kitala und Laakso 2014) als auch bei Studierenden zu erklären (Doukas und Andreatos 2007). Es zeigt sich, dass die Akzeptanz von E-Assessments von einer Vielzahl an Variablen abhängig ist (Zheng und Bender 2019) und zudem das Potential für signifikante »Enacting«-Effekte bietet. Studierende, die ursprünglich geringe oder negative Erwartungen haben, zeigen nach der Nutzung der elektronischen Infrastruktur eine signifikante Steigerung ihrer Akzeptanz (Looi, Sun, Wu, Seow, Chia, Wong et al. 2014). Diese Ergebnisse sprechen dafür, dass die Akzeptanz der Einführung von E-Assessments sowohl bei Lehrenden wie auch Studierenden nicht allein durch kommunikative Formate und Schulungen erhöht werden kann, sondern maßgeblich auch praktischen Umgang mit den Systemen erfordert.

8 Zusammenfassung

Die Bedeutung des Prüfens mit digitalen Mitteln hat an Hochschulen in den vergangenen Jahren stetig zugenommen. Mit der Corona-Pandemie hat sich in diese Entwicklung noch einmal erheblich beschleunigt. Die systematische Beschäftigung mit dem

Themenfeld wird durch eine Vielzahl an Begriffen erschwert, die überdies nicht immer deckungsgleich verwendet werden. E-Assessments, digitale und elektronische Prüfungen oder Tests, hybride Formate, Fernprüfungen und Take-Home-Klausuren, Closed-Book-, Cheat-Sheet-, Open-Book- oder Open-Web-Prüfungen sind nur einige der Begriffe, mit denen die verschiedenen Akteurinnen und Akteure im deutschsprachigen Hochschulwesen operieren. Das vorliegende Kapitel unternimmt deshalb eine vorläufige Arbeitsdefinition grundlegender Termini auf dem Gebiet der digital gestützten Leistungsprüfungen. E-Assessments haben gegenüber klassischen Prüfungsformen Vorteile aus organisationaler, psychologischer, analytischer und pädagogischer Perspektive. So sind E-Assessments ein zentrales Werkzeug zur Realisierung von orts- und zeitflexiblen Take-Home-Prüfungen. Auch ermöglichen sie alternative Prüfungsformen wie digitale Prüfungsportfolios oder authentische kognitive Leistungsmessungen mit Softwareapplikationen. Zugleich aber stellen sie die Hochschulen vor neue Herausforderungen. Diese gründen nur zu einem kleinen Teil in den von vielen Lehrenden geäußerten Bedenken beim Umgang mit wachsenden Täuschungsmöglichkeiten. Vielmehr steigen durch E-Assessments die Anforderungen nicht nur an die technische Infrastruktur der Hochschulen und ihrer Studierenden, sondern auch an die prüfungsdidaktischen Kompetenzen der Lehrenden. Sie müssen sich mit den neuen Prüfungsformen vertraut machen und ebenso willens wie in der Lage sein, ihre langjährig erprobten klassischen Prüfungen zu digitalisieren. Die empirische Befundlage zu den Effekten der verschiedenen Formen von E-Assessments bietet hier aber eine gute Orientierungsmöglichkeit. Insgesamt ist zu konstatieren, dass Hochschulen, Lehrende und Studierende derzeit in großer Breite und mit hohem Tempo damit beginnen, die enormen Potentiale digital gestützter Prüfungen auszuloten. E-Assessments werden kompetenzorientierter, technisch komplexer und didaktisch kreativer. Die Zeit, in der die Durchführung elektronischer Multiple-Choice Aufgaben als Speerspitze digitalen Prüfens galt, ist zum Glück vorbei.

Literatur

- Albrecht, J., Fischer, M. (2020). *Gutachten zur datenschutzrechtlichen Zulässigkeit von Überwachungsfunktionen bei Online-Klausuren. Projekt Rechtsinformationsstelle Digitale Hochschule NRW.* https://www.itm.nrw/wp-content/uploads/RiDHnrw_11.06.20_Gutachten-zur-datenschutzrechtlichen-Zulässigkeit-von-%C3%A4rgerlichkeiten.pdf [10.03.2021].
- Al-Samarraie, H., & Saeed, N. (2018). A systematic review of cloud computing tools for collaborative learning: Opportunities and challenges to the blended-learning environment. *Computers & Education*, 124, 77-91.
- Astwood, P. M. & Slater, T. F. (1997). Effectiveness and management of portfolio assessment in high-enrollment courses. *Journal of Geoscience Education*, 45 (3), 238-242.
- Asturias, H. (1994). Implementing the »Assessment Standards for School Mathematics«. Using Students' Portfolios to Assess Mathematical Understanding. *The Mathematics Teacher*, 87 (9), 698-701.

- Baillie, C. & Toohey, S. (1997). The ›Power Test‹. Its impact on student learning in a materials science course for engineering students. *Assessment & Evaluation in Higher Education*, 22 (1), 33-48.
- Billerbeck, K., Barnat, M. & Knutzen, S. (2016). Kompetenzorientierung auf dem Prüfstand. Erprobung von Indikatoren für den Kulturwandel. *Die Hochschullehre*, 2. https://www.hochschullehre.org/wp-content/files/diehochschullehre_2016_billerbeck_barnat_knutzen.pdf [01.06.2021].
- Barton, J. & Collins, A. (1993). Portfolios in teacher education. *Journal of teacher education*, 44 (3), 200-210.
- Bender, E., Hilkenmeier, F. & Schaper, N. (2015). Zusammenhänge von kontinuierlichen Lernanreizen und der Erreichung kompetenzorientierter Lernziele in der Hochschullehre. *Die Hochschullehre*, 1, 1-12.
- Bengtsson, L. (2019). Take-home exams in higher education: a systematic review. *Education Sciences*, 9 (4), 267.
- 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: What the Student Does* (4. Auflage). Maidenhead: Open University Press.
- Bloom, B. S. (1974). Time and learning. *American Psychologist*, 29 (9), 682-688.
- Butler, A. C. (2018). Multiple-choice testing in education: Are the best practices for assessment also good for learning? *Journal of Applied Research in Memory and Cognition*, 7 (3), 323-331.
- Caruth, G. D. & Caruth, D. L. (2013). Distance education in the United States: From correspondence courses to the Internet. *Turkish Online Journal of Distance Education*, 14 (2), 141-149.
- Chang, C. C. & Wu, B. H. (2012). Is teacher assessment reliable or valid for high school students under a web-based portfolio environment? *Journal of Educational Technology & Society*, 15 (4), 265-278.
- Christy, A. D. & Lima, M. (1998). The use of student portfolios in engineering instruction. *Journal of Engineering Education*, 87 (2), 143-148.
- Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. London: Routledge.
- Davies, A. & LeMahieu, P. (2003). Assessment for learning: Reconsidering portfolios and research evidence. In M. Segers, F. Dochy, & E. Cascallar (Hg.), *Optimising new modes of assessment: In search of qualities and standards* (141-169). Dordrecht: Springer.
- Davis, M. H., Ben-David, M. F., Harden, R. M., Howie, P., Ker, J., McGhee, C., & Snadden, D. (2001). Portfolio assessment in medical students' final examinations. *Medical Teacher*, 23 (4), 357-366.
- DeSouza, E. & Fleming, M. (2003). A comparison of in-class and online quizzes on student exam performance. *Journal of Computing in Higher Education*, 14 (2), 121-134.
- Doukas, N. & Andreatos, A. (2007). Advancing electronic assessment. *International Journal of Computers Communications & Control*, 2 (1), 56-65.
- Durning, S. J., Dong, T., Ratcliffe, T., Schuwirth, L., Artino, A. R., Boulet, J. R. & Eva, K. (2016). Comparing open-book and closed-book examinations: a systematic review. *Academic Medicine*, 91 (4), 583-599.

- Dysthe, O. & Engelsen, K. (2004). Portfolios and assessment in teacher education in Norway: a theory-based discussion of different models in two sites. *Assessment & evaluation in higher education*, 29 (2), 239-258.
- Ehlers, J. P., Guetl, C., Höntzscher, S., Usener, C. A. & Gruttmann, S. (2013). Prüfen mit Computer und Internet. Didaktik, Methodik und Organisation von E-Assessment. In M. Ebner & S. Schön (Hg.), *L3T. Lehrbuch für Lernen und Lehren mit Technologien*. Epubli.
- Feldhusen, J. F. (1961). An evaluation of Open college students' reactions to open book examinations. *Educational and Psychological Measurement*, 21 (3), 637-646.
- Flechsig, K.-H. (1976). *Prüfungen und Evaluation, Blickpunkt Hochschuldidaktik* 40, 303-336.
- Francis, J. (1982). A case for open-book examinations. *Educational Review*, 34 (1), 13-26.
- Gharib, A. & Phillips, W. (2013). Test anxiety, student preferences and performance on different exam types in introductory psychology. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 3 (1), 1.
- Gharib, A., Phillips, W. & Mathew, N. (2012). Cheat Sheet or Open-Book? A Comparison of the Effects of Exam Types on Performance, Retention, and Anxiety. *Online Submission*, 2 (8), 469-478.
- Gholami, V. & Moghaddam, M. M. (2013). The effect of weekly quizzes on students' final achievement score. *International Journal of Modern Education and Computer Science*, 5 (1), 36.
- Gupta, M. S. (2007). Open-Book Examinations for Assessing Higher Cognitive Abilities. *IEEE Microwave Magazine*, 8 (6), 46-50.
- Halbherr, T., Dittmann-Domenichini, N., Piendl, T. & Schlienger, C. (2016). Authentische, kompetenzorientierte Online-Prüfungen an der ETH Zürich. *Zeitschrift für Hochschulentwicklung*, 11 (2) 247-269.
- Händel, M., Wimmer, B. & Ziegler, A. (2020). E-portfolio use and its effects on exam performance—a field study. *Studies in Higher Education*, 45 (2), 258-270.
- Heijne-Penninga, M., Kuks, J. B. M., Schönrock-Adema, J., Snijders, T. A. B. & Cohen-Schotanus, J. (2008). Open-book tests to complement assessment-programmes: analysis of open and closed-book tests. *Advances in Health Sciences Education*, 13 (3), 263-273.
- Hellas, A., Leinonen, J. & Ihantola, P. (2017). Plagiarism in take-home exams: Help-seeking, collaboration, and systematic cheating. *Proceedings of the 2017 ACM conference on innovation and technology in computer science education* (238-243). New York: Association for Computing Machinery.
- Hillier, M., & Fluck, A. E. (2017). Transforming exams—how IT works for BYOD e-Exams. In *Proceedings ASCILITE2017: 34th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (100-105). Queensland: University of Southern Queensland.
- Hochschulrektorenkonferenz (2020). *Micro-Degrees und Badges als Formate digitaler Zusatzqualifikation*. <https://www.hrk.de/positionen/beschluss/detail/micro-degrees-und-badges-als-formate-digitaler-zusatzqualifikation/> [15.12.2020].
- Huber, L. & Reinmann, G. (2019). Vom forschungsnahen zum forschenden Lernen an Hochschulen. *Wege der Bildung durch Wissenschaft*. Berlin: Springer.

- Iller, C., & Wick, A. (2009). Prüfungen als Evaluation der Kompetenzentwicklung im Studium. *Das Hochschulwesen*, 57 (6), 195-201.
- Imrie, B. W. (1995). Assessment for learning: quality and taxonomies. *Assessment & Evaluation in Higher Education*, 20 (2), 175-189.
- Jeremias, X. V. & Sellmer, B. (2019). Chancen und Herausforderungen von E-Assessments im Anwendungskontext. In T. Barton, C. Müller und C. Seel (Hg.), *Hochschulen in Zeiten der Digitalisierung. Lehre, Forschung und Organisation* (123-137). Wiesbaden: Springer.
- Jervis, C. G. & Brown, L. R. (2020). The prospects of sitting «end of year» open book exams in the light of COVID-19: A medical student's perspective. *Medical Teacher*, 42 (7), 830-831.
- Jones, A., & Bissell, C. (2011). The social construction of educational technology through the use of authentic software tools. *Research in Learning Technology*, 19 (3).
- Kuikka, M., Kitola, M. & Laakso, M. J. (2014). Challenges when introducing electronic exam. *Research in Learning Technology*, 22.
- Kulik, C. L. C. & Kulik, J. A. (1987). Mastery testing and student learning: A meta-analysis. *Journal of Educational Technology Systems*, 15 (3), 325-345.
- Li, M., Luo, L., Sikdar, S., Nizam, N. I., Gao, S., Shan, H., Kruger, M. et al. (2021). Optimized collusion prevention for online exams during social distancing. *Nature Partner Journals Science of Learning*, 6 (1), 1-9.
- Lombardi, J. (2008). To portfolio or not to portfolio: Helpful or hyped? *College teaching*, 56 (1), 7-10.
- Looi, C. K., Sun, D., Wu, L., Seow, P., Chia, G., Wong, L. H., & Norris, C. (2014). Implementing mobile learning curricula in a grade level: Empirical study of learning effectiveness at scale. *Computers & Education*, 77, 101-115.
- Marsh, R. (1984). A comparison of take-home versus in-class exams. *The Journal of Educational Research*, 78 (2), 111-113.
- Mathew, N. (2012). Student Preferences and Performance: A Comparison of Open-Book, Closed Book, and Cheat Sheet Exam Types. *Proceedings of the National Conference on Undergraduate Research* (NCUR). Asheville: University of North Carolina.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H. & Morrisette, N. (2007). Testing the testing effect in the classroom. *European journal of cognitive psychology*, 19 (4-5), 494-513.
- Moosbrugger, H. & Kelava, A. (2008). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hg.), *Testtheorie und Fragebogenkonstruktion* (7-26). Berlin, Heidelberg: Springer.
- Minocha, S. (2009). Role of social software tools in education: A literature review. *Education and Training*, 51 (5/6), 353-369.
- Morgenroth, C. (2017). *Hochschulstudienrecht und Hochschulprüfungsrecht* (2. Auflage). Baden-Baden: Nomos.
- Msila, V. (2014). Assessment in a Transforming Higher Learning Institution: A Case for Open-Book Examinations. *Mediterranean Journal of Social Sciences*, 5 (14), 365.
- Myyry, L. & Joutsenvirta, T. (2015). Open-book, open-web online examinations: Developing examination practices to support university students' learning and self-efficacy. *Active Learning in Higher Education*, 16 (2), 119-132.

- Niehues, N., Fischer, E. & Jeremias, C. (2018). *Prüfungsrecht* (7. Auflage). München: C. H. Beck.
- Nguyen, K. & McDaniel, M. A. (2015). Using quizzing to assist student learning in the classroom: The good, the bad, and the ugly. *Teaching of Psychology*, 42 (1), 87-92.
- Pörzgen, S., Olivier, H., Sackbrook, J. L., & Pinkwart, N. (2012). Papier oder elektro-nisch? Eine Prozesskostenanalyse von Klausuren im universitären Umfeld. In U. Goltz, M. Magnor, H.-J. Appelrath, H. K. Matthies, W.-T. Balke & L. Wolf (Hg.), *INFORMATIK 2012 (1792-1806)*. Bonn: Gesellschaft für Informatik e.V.
- Redecker, C. & Johannessen, Ø. (2013). Changing assessment – Towards a new assessment paradigm using ICT. *European Journal of Education*, 48 (1), 79-96.
- Reinmann, G. (2014). *Kompetenzorientierung und Prüfungspraxis an Universitäten: Ziele heute und früher, Problemanalyse und ein unzeitgemäßer Vorschlag*. Preprint.
- Reis, O., & Ruschin, S. (2008). Zur Vereinbarkeit von Prüfungssystem und Kompetenz-orientierung. *Personal-und Organisationsentwicklung/-politik*, 1 (2), 17-21.
- Ridgway, J., McCusker, S. & Pead, D. (2004). *Literature review of e-assessment*. Bristol: Futurelab.
- Roediger III, H. L., Agarwal, P. K., McDaniel, M. A. & McDermott, K. B. (2011). Test-en-hanced learning in the classroom: long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 17 (4), 382.
- Scheuermann, F., A. G. Pereira (2008): *Towards a research agenda on computer-based assessment – challenges and needs for european educational measurement*. Luxembourg: Office for Official Publications of the European Communities.
- Schmees, M. & Horn, J. (2014). *E-Assessments an Hochschulen: Ein Überblick: Szenarien. Pra-xis. E-Klausur-Recht*. Münster: Waxmann.
- Segers, M., Gijbels, D. & Thurlings, M. (2008). The relationship between students' per-ceptions of portfolio assessment practice and their approaches to learning. *Educational Studies*, 34 (1), 35-44.
- Sietses, L. (2016). *White Paper Online Proctoring. Questions and answers about remote proctor-ing*. Utrecht: SURF.
- Song, B. & August, B. (2002). Using portfolios to assess the writing of ESL students: a powerful alternative? *Journal of second language writing*, 11 (1), 49-72.
- Spanjers, I. A., Könings, K. D., Leppink, J., Verstegen, D. M., de Jong, N., Czabanowska, K. & van Merriënboer, J. J. (2015). The promised land of blended learning: Quizzes as a moderator. *Educational Research Review*, 15, 59-74.
- Tao, J. & Li, Z. (2012). A Case Study on Computerized Take-Home Testing: Benefits and Pitfalls. *International Journal of Technology in Teaching & Learning*, 8 (1).
- Theophilides, C. & Koutsolini, M. (2000). Study behavior in the closed-book and the open-book examination: A comparative analysis. *Educational Research and Evaluation*, 6 (4), 379-393.
- Urfer-Schumacher, M. (2016). Problembasiert, kompetenzorientiert prüfen. *Zeitschrift für Hochschulentwicklung*, 11 (3), 67-84.
- Von Gruenigen, D., e Souza, F. B. D. A., Pradarelli, B., Magid, A. & Cieliebak, M. (2018). Best practices in e-assessments with a special focus on cheating prevention. In 2018 *IEEE Global Engineering Education Conference (EDUCON)*, 17.-20.04.2018 (893-899).

- Walstad, W. B. (2001). Improving assessment in university economics. *The Journal of Economic Education*, 32 (3), 281-294.
- Weber, L. J., McBee, J. K. & Krebs, J. E. (1983). Take home tests: An experimental study. *Research in Higher Education*, 18 (4), 473-483.
- Westerkamp, A. C., Heijne-Penninga, M., Kuks, J. B. & Cohen-Schotanus, J. (2013). Open-book tests: Search behaviour, time used and test scores. *Medical teacher*, 35 (4), 330-332.
- Wildt, J. & Wildt, B. (2011). Lernprozessorientiertes Prüfen im »Constructive Alignment«. *Neues Handbuch Hochschullehre*, H 6.1, 46.
- Williams, J. B. & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology*, 40 (2), 227-236.
- Zheng, M. & Bender, D. (2019). Evaluating outcomes of computer-based classroom testing: Student acceptance and impact on learning and exam performance. *Medical teacher*, 41 (1), 75-82.

