

Jan P. EHLERS<sup>1</sup>, Torsten CARL, Karl-Heinz WINDT, Daniel MÖBS,  
Jürgen REHAGE, Andrea TIPOLD (Hannover / Bergisch Gladbach)

## **Blended Assessment: Mündliche und elektronische Prüfungen im klinischen Kontext**

### **Zusammenfassung**

Die Curriculumsreform in der Tiermedizin und die Einführung des „praktischen Jahres“ an der Tierärztlichen Hochschule Hannover machten es notwendig, die klinischen Prüfungen in zwei Teilen durchzuführen. Eine Kombination aus schriftlicher und mündlicher Prüfung wurde gewählt. In der vorliegenden Studie werden Erfahrungen mit verschiedenen Prüfungsformen hinsichtlich ihrer Effizienz und Qualität verglichen: mündliche Prüfungen, papiergestützte schriftliche Prüfungen und elektronische Prüfungen. Zusätzlich wurden Studierende und Dozierende zu den Prüfungsformen befragt. Durch Einführung der elektronischen Teilprüfungen wurde eine Effizienz- und Qualitätssteigerung erreicht. Die Akzeptanz bei den Studierenden ist gut, bei den Prüfenden herrschen noch gewisse Ressentiments. Mit der Einführung elektronischer Prüfungen wurde der richtige Weg eingeschlagen; um die Qualität weiter zu steigern, sind aber Fortbildung der PrüferInnen und die Weiterentwicklung der Itemformate notwendig.

### **Schlüsselwörter**

Tiermedizin, Prüfungen, Itemanalyse, eAssessment, Staatsexamen

## **Blended Assessment: Oral and Electronic Examinations in a Clinical Setting**

### **Abstract**

The new veterinary curriculum and the introduction of the “practical year” at the University for Veterinary Medicine in Hannover required a bipartite clinical examination. A combination of written and oral examination was established. The current study compares experiences with different formats of tests regarding efficiency and quality factors: oral, paper-based written and electronic assessment. In addition students and teachers were interviewed. The use of e-assessment raised efficiency and quality. The acceptance by students was good but the teachers remained cautious. For further improvement of quality faculty development in the field of test- and item generation and the development of additional items is necessary. In summary, this study proves that electronic assessment is feasible in clinical examinations.

### **Keywords**

Veterinary medicine, examinations, item review, e-assessment, state examination

---

<sup>1</sup> e-Mail: [Jan.Ehlers@tiho-hannover.de](mailto:Jan.Ehlers@tiho-hannover.de)

# 1 Einleitung

Traditionell werden Examensprüfungen im Tiermedizinstudium in Deutschland mündlich durchgeführt. Bei nur wenigen Prüfungen wurden in den letzten Jahren auch schriftliche Leistungsnachweise eingefordert. Die mündlichen Prüfungen finden vornehmlich in Gruppen von 4-5 Studierenden bei einem Prüfer statt. Erst bei Wiederholungsprüfungen sind Beisitzer vorgesehen. An diesem etablierten Verfahren gab es aber sowohl aus Studierenden- als auch aus Dozierendensicht Kritik (HUCKLENBROICH, 2004; KLEE, 2007).

Die Neuordnung der klinischen Lehre und die Einführung des klinisch-praktischen Jahres (WAGELS et al., 2008) an der Tierärztlichen Hochschule Hannover (TiHo) machte es notwendig, die klinischen Prüfungen des Staatsexamens aufzuteilen, um die allgemeine Approbation (TAppV, 2007) sicher gewährleisten zu können. Sie bestehen seit 2005 aus einem tierartübergreifenden, schriftlichen Teil vor dem praktischen Jahr und einem tierartspezifischen, praktischen Teil im 11. Semester. Alle Studierenden müssen in jeweils drei Disziplinen (Chirurgie, Innere Medizin und Reproduktion) diese Prüfungen ablegen. Die Endnoten setzen sich zu 40 Prozent aus dem Ergebnis der schriftlich-theoretischen und zu 60 Prozent aus dem Ergebnis der mündlich-praktischen Prüfung zusammen.

In der Medizin werden verschiedene Prüfungsformen eingesetzt, die unterschiedliche Qualitäten des Wissens abprüfen (MILLER, 1990, s. Abb. 1):

- für das Prüfen von deskriptivem Wissen (Knows) Multiple Choice (MC)-Fragen oder mündliche Prüfungen
- für prozedurales Wissen ebenfalls mündliche Prüfungen oder Key-Feature-Fragen (ROTHHOFF et al., 2006)
- für Leistungen "in vitro" Objective Structured Clinical Examinations (OSCE, NIKENDEI & JÜNGER, 2006)
- für Leistungen "in vivo" MiniCEX (JOLLY, 2007).

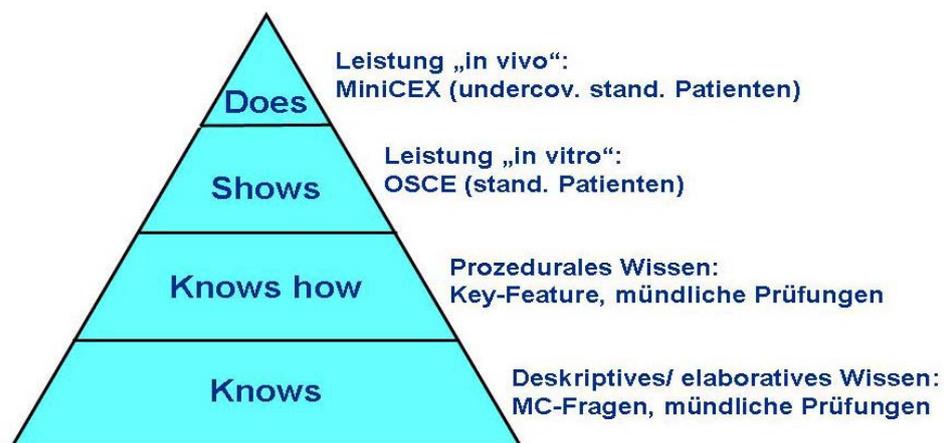


Abb. 1: Wissenspyramide (MILLER 1990) modifiziert mit geeigneten Prüfungsformen nach KOPP et al. 2006 (MiniCEX: Mini Clinical Evaluation Exercise; OSCE = Objective Structured Clinical Examinations; undercov.: undercover, verdeckte; stand.: standardisierte; MC: Multiple Choice)

Die klinisch-praktische Prüfung an der TiHo entspricht einer Kombination von mündlicher Prüfung und MiniCEX. Damit wird praxisnah klinisches Wissen geprüft.

Prüfungen für das Staatsexamen sollten die Hauptgütekriterien (Objektivität, Validität und Reliabilität) möglichst gut erfüllen und Nebengütekriterien (Utilität, Testfairness, Testökonomie, Transparenz, Unverfälschbarkeit und Zumutbarkeit) beachten. Für die Medizin (FISCHER, 2008) und Tiermedizin (BAILLIE & RHIND, 2008) wurden bereits Leitlinien veröffentlicht, die den Prüfenden Hilfestellungen bieten sollen.

Wenn es um schriftliche Prüfungen geht, werden viele Vorteile für elektronische Prüfungen beschrieben (BJÖRNSSON, 2008). In der Tiermedizin wurden bereits erfolgreich elektronische Prüfungen durchgeführt (BRIESE & HARTUNG, 2008), allerdings teilweise mit kleinen Kohorten und im Vergleich zu Papierprüfungen erhöhtem Aufwand (POSPISCHIL et al., 2007).

Neben den erweiterten Möglichkeiten durch elektronisches Prüfen kann hierdurch auch eine Effizienzsteigerung durch Zeit- und Kostenersparnis erreicht werden (FISCHER et al., 2007; WOLF, 2007). Da oft nicht die Räumlichkeiten zur Vorhaltung eines ausreichenden Computerprüfungsraumes bestehen, gibt es auch das Geschäftsmodell, elektronische Prüfungen als Dienstleistung zu kaufen (KRÜCKEBERG et al., 2008).

Die vorliegende Studie soll untersuchen, ob rein mündliche Prüfungen in ein elektronisch-schriftliches Format überführt werden können und welche Auswirkungen dies auf die Erfüllung der Gütekriterien hat. Ziel ist, die Qualität der Prüfungen zu verbessern und den Aufwand für die Prüfenden zu verringern.

## 2 Methode

In den Jahre 2005 bis 2008 wurden die theoretischen klinischen Prüfungen an der Stiftung Tierärztliche Hochschule Hannover auf drei verschiedene Arten durchgeführt:

- 2005 und 2006: Schriftlich mit kurzen Freitext-Antworten
- 2007: Schriftlich mit teilweise bebilderten Multiple-Choice-Fragen (TypA positiv, 1 Antwort, 2 Distraktoren, RODRIQUEZ, 2005)
- 2008: Elektronisch mit teilweise bebilderten Multiple-Choice-Fragen (TypA positiv, 1 Antwort, 2 Distraktoren)

Die Freitextklausuren wurden parallel in vier Hörsälen geschrieben. Die fertigen Klausuren wurden per Hand von jeweils 3 Personen mit Prüfungserlaubnis aus jeder Klinik korrigiert und bewertet.

Die schriftlichen MC-Klausuren wurden ebenfalls parallel in vier Hörsälen geschrieben. Es gab drei verschiedene Versionen der Klausur mit unterschiedlicher Fragen- und Antwortreihenfolge, um ein Abschreiben zu verhindern. Die fertigen Klausuren wurden eingescannt und mit der Software McSurvey ausgewertet.

Die elektronischen MC-Klausuren wurden in zwei Kohorten direkt nacheinander in einem Hörsaal geschrieben. Es gab für alle Studierenden verschiedene Versionen der

Klausur mit randomisierter Fragen- und Antwortenreihenfolge, um ein Abschreiben zu verhindern. Die Klausuren wurden in Echtzeit von dem System ausgewertet.

Die Ergebnisse der schriftlichen Prüfungen wurden anhand der richtigen Antworten, der Trennschärfe und Schwierigkeit der Fragen, Cronbachs alpha (Maß für die Reliabilität eines Werkzeugs, Idealwerte  $> 0,8$ ) der Klausur sowie der Dauer der Erstellung und Korrektur verglichen. Die Noten der schriftlichen Prüfungen wurden mit der Endnote verglichen. Dies war nur für die Klausuren von 2005 bis 2007 möglich, da zu den schriftlichen Prüfungen 2008 die Endnoten erst nach den mündlichen Prüfungen 2009/10 vorliegen werden. Zum weiteren Vergleich wurden die Noten der rein mündlichen Prüfungen aus dem Jahr 2005 herangezogen, die noch nicht unterteilt waren.

Um die Akzeptanz von elektronischen Prüfungen zu evaluieren, wurden insgesamt 677 Teilnehmerinnen und Teilnehmer an den elektronischen klinischen Prüfungen jeweils direkt im Anschluss an die Klausuren mit Q [kju:] quantitativ befragt. Unter anderem sollten folgende Fragen mit 1 (ja, sehr gut) bis 6 (nein, gar nicht) bewertet werden:

1. War das System einfach zu benutzen?
2. Halten Sie die Benutzung eines elektronischen Systems für schriftliche Prüfungen für sinnvoll?
3. Halten Sie MC-Fragen prinzipiell für sinnvoll für Prüfungen?
4. Sollen auch weitere Fragenformate eingesetzt werden?
5. Sollten mehr schriftliche statt mündliche Prüfungen gehalten werden?
6. Empfinden Sie schriftliche Prüfungen als fairer im Vergleich zu mündlichen?
7. Haben Sie Vertrauen in die Technik?
8. Wie bewerten Sie die komplette Prüfung und deren Ablauf?

Für eine Betrachtung der Änderung der Prüfungsform aus Dozierendensicht wurden insgesamt 35 Prüfende aus vier Klinken um die Teilnahme an einer qualitativen Befragung gebeten. Der Interviewleitfaden sah folgende Punkte vor:

1. Was sind die Stärken von mündlichen Prüfungen?
2. Was sind die Schwächen von mündlichen Prüfungen?
3. Was sind die Stärken von schriftlichen Prüfungen?
4. Was sind die Schwächen von schriftlichen Prüfungen?
5. Was sind die Stärken von elektronischen Prüfungen?
6. Was sind die Schwächen von elektronischen Prüfungen?
7. Wie sieht für Sie die ideale Prüfungsform aus?

## 3 Ergebnisse

### 3.1 Vergleich der Klausurformen

In den Freitextklausuren 2005 und 2006 sollten bei der Auswertung per Hand nur ganze Punktzahlen pro Frage vergeben werden. Bei der Auswertung durch die Fachvertreterinnen und Fachvertreter wurden aber doch 0,25 oder 0,5 Punkte als teilweise richtig beantwortete Frage vergeben, da die Studierenden trotz der Aufforderung, nur Stichworte zu verwenden, oft grenzwertige Formulierungen benutzten. Es war eine Begutachtung durch mehrere Prüfende notwendig, was zur Zeitverzögerung bei der Korrektur der Klausuren führte. Insgesamt dauerte es vier Wochen, bis die Ergebnisse der Prüfungen vorlagen.

Eine Itemanalyse war nicht möglich, da keine Übersicht über die Ergebnisse der einzelnen Fragen angefertigt wurde. Die durchschnittlichen Ergebnisse lagen in der Chirurgie bei 68% richtigen Antworten, in der Inneren Medizin bei 67% und in der Reproduktion bei 64%.

Im Sommer 2007 wurde das erste Mal eine Multiple-Choice-(MC)-Klausur durchgeführt. Nach Einweisung der Prüfenden durch die AG Reformstudiengang der Berliner Charité fiel die Entscheidung für das Verfahren Typ Apos mit nur zwei Distraktoren. Zur Erstellung und Auswertung wurde die durch eigene Programme ergänzte McSurvey-Software eingesetzt. So konnten drei Klausuren mit den gleichen Fragen und Antworten in randomisierter Anordnung geschaffen werden. Der Ablauf der Klausuren war für alle Beteiligten problemlos. Die technische Erstellung jeder einzelnen Klausur (Ausdruck, Sortierung) nahm zwei Tage in Anspruch. Einige Fragen wurden durch Schwarz-Weiß-Kopien von Befundfotos ergänzt. Die Qualität war als nicht besonders gut einzustufen.

Die Ergebnisse lagen innerhalb von 30 Minuten nach Klausurende vor und konnten in das TiHo-Portal „StudIS“ übertragen werden. Die durchschnittlichen Erfolgsquoten der MC-Klausuren zeigten ein identisches Ergebnis wie bei den Klausuren mit Freitextfragen (Chirurgie 68% richtige Antworten, Innere Medizin 66% und Reproduktion 64%).

Der durchschnittliche Schwierigkeitsindex der Klausurfragen lag zwischen 66,7 und 70,4 (Tab. 1), die durchschnittliche Fragentrennschärfe der einzelnen Klausuren zwischen 0,22 und 0,24 und Cronbachs alpha zwischen 0,66 und 0,74 (Tab. 1).

Im Sommer 2008 wurde erstmalig das elektronische Prüfungssystem Q[kju:] eingesetzt, das als komplette Dienstleistung inkl. Tablet-PC bei der Codiplan GmbH erworben wurde. Es kam der gleiche Fragentyp wie bei den vorherigen MC-Klausuren zur Anwendung. Die Klausur wurde direkt in der online-Plattform Q[kju:] Online erstellt, sodass keine weitere Vorbereitungszeit notwendig war. Auch in diesen Klausuren wurden Fragen durch Bilder ergänzt, die in guter Qualität und in Farbe bildschirmfüllend angezeigt wurden.

Die Ergebnisse der Klausuren lagen direkt bei Prüfungsende vor und konnten dann über das TiHo-Portal mitgeteilt werden. Auch mit dieser Prüfungsform konnten vergleichbare Ergebnisse erzielt werden (Chirurgie 66% richtige Antworten, Innere Medizin 64% und Reproduktion 67%).

Der durchschnittliche Schwierigkeitsindex der Klausurfragen lag zwischen 63,9 und 66,5 (Tab.1), die durchschnittliche Fragentrennschärfe der einzelnen Klausuren zwischen 0,22 und 0,24 und Cronbachs alpha zwischen 0,68 und 0,72 (Tab. 1).

	Chirurgie 2007	Chirurgie 2008	Innere 2007	Innere 2008	Reproduk- tion 2007	Reproduk- tion 2008
Durchschnittlicher Schwierigkeitsindex (%)	70,4	65,8	66,7	63,9	68,8	66,5
Durchschnittliche Trennschärfe	0,24	0,22	0,24	0,24	0,22	0,23
Cronbachs $\alpha$	0,72	0,68	0,74	0,72	0,66	0,68

Tab. 1: Ergebnisse der Item- und Klausuranalyse der Klausuren 2007 und 2008

### 3.2 Vergleich der Klausurergebnisse in den einzelnen Disziplinen

#### Chirurgie

An den Chirurgieprüfungen 2005 bis 2008 nahmen zwischen 206 und 219 Studierende teil. Die Durchschnittsnote der rein mündlichen Prüfung 2005 lag bei 2,7. Die schriftliche Durchschnittsnote der Prüfung 2006 lag bei 3,4, die kombinierte Endnote besserte sich auf 2,7. Die gleiche Tendenz zeigte sich auch 2007 mit einer Verbesserung von 3,8 auf 2,8 und 2008 von 2,9 auf 2,5.

Die Korrelationskoeffizienten der schriftlichen Teilnoten und der Endnoten waren über die Jahre ähnlich gut (2006:  $r=0,72$ ; 2007:  $r=0,67$ ; 2008:  $r=0,55$ ).

#### Innere Medizin

An den Prüfungen in Innerer Medizin 2005 bis 2008 nahmen zwischen 204 und 211 Studierende teil. Die Durchschnittsnote der rein mündlichen Prüfung 2005 lag bei 2,6. Die schriftliche Durchschnittsnote der Prüfung 2006 lag bei 3,4, die kombinierte Endnote besserte sich auf 2,6. Die gleiche Tendenz zeigte sich auch 2007 mit einer Verbesserung von 3,2 auf 2,5 und 2008 von 3,3 auf 2,7.

Die Korrelationskoeffizienten der schriftlichen Teilnoten und der Endnoten waren über die Jahre ähnlich hoch (2006:  $r=0,68$ ; 2007:  $r=0,74$ ; 2008:  $r=0,60$ ).

#### Reproduktionsmedizin

An den Prüfungen in Reproduktionsmedizin 2005 bis 2008 nahmen zwischen 198 und 215 Studierende teil. Die Durchschnittsnote der rein mündlichen Prüfung 2005 lag bei 2,5. Auch für 2006 liegen aus verwaltungstechnischen Gründen nur die Endnoten vor. Die Durchschnittsnote war 2,9. Die schriftliche Durchschnittsnote der Prüfung 2007 lag bei 3,6, die kombinierte Endnote besserte sich auf 2,7. Auch 2008 wurde eine Verbesserung von 3,1 auf 2,7 verzeichnet.

Die Korrelationskoeffizienten der schriftlichen Teilnoten und der Endnoten waren über die beiden Jahre ähnlich gut (2007:  $r=0,72$ ; 2008:  $r=0,67$ ).

### 3.3 Ergebnisse der Studierendenbefragung

Bei den quantitativen Befragungen wurden insgesamt 658 Fragebögen (97,3% Rücklauf) beantwortet (Tab. 2). Q [kju:] wurde mit einer Durchschnittsbewertung von 1,4 als sehr einfach zu benutzen und als sinnvoll für den Einsatz in schriftlichen Prüfungen (2,5) eingeschätzt. Nur 49% der Studierenden empfanden MC-Fragen als prinzipiell sinnvoll für Prüfungen und 72,9% der Befragten sprechen sich für den Einsatz weiterer Frageformate wie z.B. Bildanalyse oder fallorientierte Fragen aus.

Mit einer Durchschnittsbewertung von 4,0 sprechen sich die Studierenden gegen schriftliche Prüfungen aus, bewerten diese aber als tendenziell fairer als mündliche Prüfungen (3,2).

Vertrauen in die verwendete Technik bestätigten 86,7% der befragten Studierenden. Nach einer Schulnote gefragt, gaben sie dem gesamten Prüfungsablauf eine Durchschnittsnote von 2,7.

Frage	n	1	2	3	4	5	6	8	ja %	nein %
War das Codiplan-System "Q-Exam" einfach zu benutzen?	658	436	186	25	4	3	4	1,4	98,3	1,7
Haben Sie Vertrauen in die eingesetzte Technik?	655	180	228	160	48	22	17	2,3	86,7	13,3
Finden Sie die Benutzung eines elektronischen Systems für schriftliche Prüfungen sinnvoll?	658	170	229	138	54	32	35	2,5	81,6	18,4
Halten Sie Multiple-Choice-Fragen prinzipiell sinnvoll für Prüfungen?	656	40	93	190	105	75	153	3,8	49,2	50,8
Sollten mehr schriftliche statt mündliche Prüfungen gehalten werden?	655	67	95	113	85	118	177	4,0	42,0	58,0
Empfinden Sie schriftliche Prüfungen fairer im Vergleich zu mündlichen?	655	127	145	128	92	68	95	3,2	61,1	38,9
Sollten weitere Frageformate (z.B. fallorientierte Fragen) eingesetzt werden?	649	47	195	231	92	51	33	3,0	72,9	27,1
Bitte bewerten Sie die komplette Prüfung und deren Ablauf mit einer Schulnote.	654	34	274	232	79	27	8	2,7	82,6	17,4

Tab. 2: Befragung der Studierenden nach den elektronischen Klausuren.  
(Likertwerte von 1=ja, sehr/sehr gut bis 6=nein, gar nicht/schlecht, ja: Prozentsatz)

der Studierenden, die mit 1-3 gewertet haben, nein: Prozentsatz der Studierenden, die mit 4-6 gewertet haben)

### 3.4 Ergebnisse der Dozierendeninterviews

Von den angefragten Prüfenden stellten sich elf (31,4%) aus allen vier Kliniken für eine qualitative Befragung zur Verfügung.

Als Hauptvorteil der mündlichen Prüfungen wurde beschrieben, dass nicht nur Faktenwissen sondern auch Zusammenhänge, Verständnis und Argumentationslogik geprüft werden und dadurch die Kenntnisse der Studierenden differenzierter betrachtet werden. Außerdem sei es möglich, den Studierenden Hilfestellungen zu geben und individuell auf sie einzugehen. Die mündliche Prüfungssituation entspreche den Anforderungen im tierärztlichen Beruf und sei ein gutes Training. Sie könne direkt praktisch am Patienten durchgeführt werden und sei ein direkter Teil der Lehre.

Als größte Schwäche mündlicher Prüfungen beschrieben fast alle Befragten die Subjektivität der Ergebnisse. Die Bewertung sei nur schwer vergleichbar, da sie auch vom Prüfer oder der Prüferin, den Fragen, dem Eindruck des Geprüften und anderen äußeren Einflüssen abhängt. Da nur ein Ausschnitt aus den Disziplinen geprüft werden kann, hängt das Ergebnis auch vom Glück bei der Patientenauswahl ab. Durch die zeitintensiven mündlichen Prüfungen ist die Belastung der Prüfenden erhöht. Die zu Prüfenden zeigen auch mehr Angst bei mündlichen als bei schriftlichen Prüfungen.

Als Hauptvorteil schriftlicher Prüfungen nannten alle befragten Dozierenden die Objektivität, die Vergleichbarkeit der Klausuren und den Ausschluss äußerer Einflüsse. Als weitere Vorteile wurden die erhöhte Effizienz, die Möglichkeit über das ganze Gebiet zu prüfen und das Feedback, das man von allen Studierenden auf seinen Unterricht bekommt, beschrieben. Außerdem fände bei schriftlichen Prüfungen ein gemeinsames Besprechen der Fragen unter den Prüfenden und ein Review statt.

Als Nachteile der schriftlichen Prüfungen beschrieben alle die Gefahr, nur Detailwissen und kein Verständnis oder komplexe Zusammenhänge abzufragen. Weiterhin wurde als Schwäche der Schriftform beschrieben, dass hier die Prüfung keine Lehre sei, da es kein Feedback gäbe, dass Fragen aufwändig erstellt werden müssen und hierfür besondere Qualifikationen notwendig seien, dass Missverständnisse in Prüfungen nicht ausgeräumt werden können, dass nicht praktisch am Patienten geprüft werden könne und dass auch hier ein gewisser subjektiver Einfluss durch Schreibweise und Stil oder schlechte Deutschkenntnisse entstehen können.

Nur wenige Befragte gaben an, keine Erfahrungen mit elektronischen Prüfungen zu haben. Andere verwiesen auf ähnliche Vor- und Nachteile wie bei den schriftlichen Prüfungen. Als Hauptvorteil der elektronischen Form wurde aber von den meisten Dozierenden genannt, dass eine Erweiterung der Items stattfinden kann. Bisher durch Einsatz von multimedialen Elementen und zukünftig wahrscheinlich durch Simulationen und Adaptivität werden auch klinische, fallorientierte Fragen ermöglicht. Außerdem weise das elektronische System eine erhöhte Effizienz bei der Fragegestaltung sowie Prüfungsdurchführung und -auswertung auf. Da alle die gleichen

Fragen aus dem ganzen Lehrgebiet bekommen, sei die Prüfung objektiver und fairer. Weiterhin kommt es durch den Einsatz von E-Prüfungen zu einem verstärkten Pre- und Post-Review der Fragen, letzteres mit automatischer Itemanalyse.

Die Prüfenden sahen elektronische Prüfungen ähnlich kritisch wie die schriftlichen. Als Hauptnachteil wurde die Beschränkung auf Faktenwissen und die Schwierigkeit, Hintergrundwissen zu prüfen, beschrieben. Die unterschiedliche Medienkompetenz und Computeraffinität habe einen direkten Einfluss auf das Prüfungsergebnis der Studierenden. Weiterhin sei es notwendig, dass die Prüfer und Prüferinnen mehr Know-How für solche Prüfungen bekämen. Der technische Support sei aufwändig und das System sehr teuer. Schließlich sei es schade, dass die Prüfung nicht mehr als Lehre betrachtet werden kann und der direkte Kontakt entfällt.

Die Meinungen über die Struktur einer idealen klinischen Prüfungen differierten sehr. Sie gingen von der Aussage, so eine Prüfung gäbe es nicht, bis zu sehr genauen Vorstellungen über den Anteil einzelner Prüfungsformate. Größte Einigkeit wurde erzielt, dass die Prüfungen aus einem schriftlichen Theorieteil und einer praktischen Prüfung am Patienten bestehen sollte. Dieser schriftliche Teil kann elektronisch mit MC-Fragen gestaltet sein, sollte aber auch noch weitere Formate enthalten (Bildanalyse, fallorientierte Fragen, Essay-Questions). Der mündlich-praktische Teil solle ausreichende Breite prüfen und den Studierenden genügend Zeit geben. Mehrfach wurde die Relevanz eines Lernzielkatalogs betont, um Lehre und Prüfungen zu parallelisieren.

## 4 Diskussion

Aufgrund der Reform des Curriculums und der Stärkung der klinisch-praktischen Ausbildung war die Stiftung Tierärztliche Hochschule Hannover gezwungen, die Prüfungsform der klinischen Prüfungen im Staatsexamen zu ändern. Bei der Umwandlung der rein mündlichen Prüfungen in einen tierartübergreifenden, schriftlich-theoretischen und einen tierartspezifischen, mündlich-praktischen Teil sollte besonderes Augenmerk auf die Steigerung der Qualität und die Effizienz gelegt werden.

Die Qualität der neuen Prüfungsformen soll anhand der Gütekriterien dargestellt werden. Sowohl Studierende als auch Dozierende beschrieben bei dem neuen Verfahren eine deutlich ausgeprägtere Objektivität als bei den mündlichen Prüfungen. Äußere Einflüsse durch die Prüfenden oder Korrigierenden konnten bei den elektronischen Klausuren vollkommen ausgeschlossen werden.

Die Reliabilität der Prüfung wird durch Cronbachs alpha errechnet. Idealerweise liegt dieser über 0,8 (MÖLTNER et al., 2006). Bei der weiteren Betrachtung der Itemanalysen war festzustellen, dass viele Fragen nicht dem idealen Schwierigkeitsindex (20-80%) aufwiesen und zu schwer oder zu leicht waren (LIENERT & RAATZ, 1998). Weiterhin könnte die Reliabilität der Prüfung durch eine genauere Anpassung der verfügbaren Zeit pro Item verbessert werden. In diesem Bereich ist eine weitere Verbesserungen der schriftlichen Prüfungen notwendig.

Die Inhaltsvalidität der klinischen Prüfungen wurde durch die Kombinationsprüfung deutlich verbessert, da das gesamte Gebiet der Disziplin und verschiedene Arten von Wissen abgeprüft werden können (LUTZ, 2004). Durch eine weitere

konkrete Ausarbeitung eines Lernzielkataloges und die Definition sogenannter "Day-One-Skills" kann auch die Konstruktvalidität der Prüfungen weiter ermittelt werden. Die enge Korrelation der schriftlichen und praktischen Prüfungen zeigt hier aber schon eine sehr gute Tendenz an. Diese belegt ebenfalls die Kriteriumsvalidität. Dass grundsätzlich eine Verbesserung zwischen der schriftlichen und mündlichen Prüfung stattfindet, ist beabsichtigt und liegt an dem dazwischen stattfindenden praktischen Semester und der damit verbundenen Spezialisierung auf eine Tierart.

Im Rahmen der Diskussion um die Utilität der neuen Prüfungsform zeigte sich, dass es mit einer schriftlichen Prüfung und in diesem Fall einer Multiple-Choice-Prüfung nicht möglich ist, auch prozedurales Wissen und Fertigkeiten zu prüfen (MILLER, 1990). Daher wurde weiterhin an einer mündlichen Teilprüfung am Patienten festgehalten.

Zur Testfairness wurde von den Prüfenden befürchtet, dass das elektronische Verfahren computeraffine Studierende bevorzugt. Die Befragung der Studierenden zeigte hingegen, dass alle gut mit dem System umgehen konnten. Die Computernutzung bei den Studierenden der Tiermedizin ist sehr ausgeprägt (EHLERS et al., 2002). Ähnlich wie in anderen Berufsgruppen (KERRES, 2009) findet das elektronische Prüfen daher eine hohe Akzeptanz.

Die Testökonomie war natürlich auch einer der Gründe, dieses neue System einzuführen. Die Effizienz eines elektronischen Prüfungsverfahrens ist deutlich größer als bei mündlichen oder papier-gestützten Prüfungen (FISCHER et al. 2007, WOLF, 2007). Da es noch keinen vorhandenen Fragenpool gab, war der Aufwand der Prüfungserstellung noch hoch, wird aber im Laufe der Zeit durch Wiederverwendung einzelner Fragen weiter nachlassen.

Durch das Erwerben der Prüfungen als Dienstleistung konnte auf ein etabliertes System zurückgegriffen werden, dass die Unverfälschbarkeit garantiert.

Die Befragung der Studierenden und auch die Erfahrungen an anderen Bildungstätten (BENNING, 2009) haben gezeigt, dass die Zumutbarkeit dieses Verfahrens gegeben ist. Die psychische Belastung und Prüfungsangst wird als deutlich geringer als in mündlichen Prüfungen beschrieben.

Die Durchführungen der ersten schriftlichen und elektronischen Prüfungen zeigten aber auch, dass es dringend notwendig ist, die Prüfer und Prüferinnen stärker didaktisch und prüfungsdidaktisch zu schulen, damit mehr als Detailwissen mit Multiple-Choice-Fragen abgefragt werden kann. Schon bei der ersten Durchführung der elektronischen Klausuren wurde den Dozierenden eine Fortbildungsveranstaltung zum Thema MC-Prüfungen durch die Charité (GEORG et al., 2006) angeboten. Diese Fortbildung bestand aus zwei Teilen.

Im ersten Teil wurde im Frontalunterricht erklärt, was MC-Fragen leisten können und worauf bei der Erstellung zu achten ist (KREBS, 2004). Der zweite Teil fand in Seminarform statt. Hier wurden eigene Fragen entwickelt und besprochen. Neben der fachlichen Tiefe ging es vor allem um formale und technische Fragen. Von der Stiftung Tierärztliche Hochschule Hannover wurde weiterhin festgelegt,

dass zuerst ein solcher Kurs über die Gestaltung von MC-Fragen durchgeführt werden soll, bevor eine Klinik oder ein Institut diese Prüfungsform durchführt.

Auch in der Befragung der Dozierenden zeigte sich, dass einige Vorteile der mündlichen Prüfungen, z.B. das Weiterhelfen, nur schwer mit dem Gleichheitsgrundsatz zu vereinen sind. Der Vorteil, dass mündliche Prüfungen in Kleingruppen den direkten Kontakt zwischen Studierenden und Dozierenden ermöglicht, sollte weiter in die Lehre (z.B. POL-Unterricht) verlegt werden und nicht erst in der abschließenden Prüfungssituation entstehen.

Die Befragung der Prüfenden hat eine große Nachfrage nach formativen Prüfungen mit gutem Feedback verdeutlicht. Daher werden studienbegleitende Testate mit einem Feedbacksystem (EHLERS et al., 2007) vermehrt vorgesehen.

Für die Tiermedizin zeigt das MC-Verfahren nur eine enge Einsatzbreite, daher wird die Ausbildungsforschung vermehrt in die Etablierung neuer Itemformen gerichtet. Zunächst sollen dabei die Fragentypen Bildanalyse und Key-Feature verwirklicht werden.

Als großer Vorteil des elektronischen Systems wird das Pre- und Post-Review der Fragen in Form einer Itemanalyse gesehen. An der Medizinischen Hochschule Hannover, die das gleiche System für elektronische Prüfungen nutzt, wird von Studierenden kritisch angemerkt, dass viele Dozierende die Itemanalyse kaum nutzen und somit ein Teil des Qualitätsmanagements von Prüfungsfragen verloren geht (CABEZON, 2009). Anhand einer neuen Prüfungsordnung wird daher an der Stiftung Tierärztliche Hochschule Hannover der Ablauf der Prüfungsgestaltung und -bewertung festgelegt.

Die ersten Ergebnisse der neuen klinischen Kombinationsprüfungen zeigen die großen Vorteile der elektronischen Prüfung auf. Es entstehen aber auch zusätzliche Aufgaben wie „Faculty Development“ für schriftliche Prüfungen, Entwicklung neuer Frageformen, Qualitätsaspekte berücksichtigen neben Zeit- und Kostensparnissen, überfakultäre Kooperationen etablieren und konkrete Lernzielkataloge festlegen.

## 5 Literaturverzeichnis

**Baillie S & Rhind S** (2008): A Guide to Assessment Methods in Veterinary Medicine. Royal College of Veterinary Surgeons Trust, London (UK).

**Benning M** (2009): Schummeln ist nicht. *duz-Magazin* 1-2/09, 38-39.

**Björnsson Jk** (2008): Changing Icelandic National Testing from Traditional Paper and Pencil Based Tests to Computer Based Testing: Some Background, Challenges and Problems to Overcome. In: Scheuermann F and Pereira AG (Eds.): Towards a Research Agenda on Computer-Based Assessment, EU-Report-CBA, JRC Scientific and Technical Records, 10-14.

**Briese A & Hartung J** (2008): Erstes elektronisches Examen im Tiermedizinstudium. *TiHo-Anzeiger* 3/2008, 5.

**Cabezón Ih** (2009): Die mächtige Itemanalyse. *curare – Zeitschrift des ASTA der MHH* (93), 12.

**Ehlers JP, Friker J, Liebich H-G & Stolla R** (2002): PC-Ausstattung und -nutzung von Studierenden der Tiermedizin im Vergleich zu Schülern der 12. Klasse. Med. Ausbild. 2002; 19: 124-126.

**Ehlers JP, Kaske M, Tipold A & Bollwein H** (2007): Einsatz von Feedbacksystemen in Präsenzveranstaltungen. eLearning in Niedersachsen. Hrsg.: Kompetenzzentrum eLearning Niedersachsen. 82-83.

**Fischer MR** (2008): Leitlinie für Fakultäts-interne Leistungsnachweise während des Medizinstudiums: Ein Positionspapier des GMA-Ausschusses Prüfungen und des Kompetenzzentrums Prüfungen Baden-Württemberg. GMS Z Med Ausbild 25 (1) Doc 74.

**Fischer V, Möbs D, vor dem Esche J & Haller H** (2007): A comparison of the assets and drawbacks of three different written assessment types. Abstracts of AMEE 2007, 25.-29.07.2007, Trondheim (N), 118.

**Georg W, Schubert S, Scheffner D & Burger W** (2006): Fünf Jahre Prüfungen im Reformstudiengang Medizin an der Charité - Universitätsmedizin Berlin. GMS Z Med Ausbild 23(3), Doc 48.

**Gucklenbroich C** (2004): Vergleichbar durchfallen - Mündliche Prüfungen widersprechen dem Grundsatz der Chancengleichheit. Nexus - Studentenzeitschrift der TiHo Hannover (WS 04/04).

**Jolly BC** (2007): The new Curriculum Framework and Assessment Practices: Current Challenges for Postgraduate Year 1 and 2. MJA 186/7, 33-36.

**Kerres M** (2009): Prüfungsteilnehmer akzeptieren Computer. In Prüfermagazin im Handwerk, Heft 2, 6.

**Klee W** (2007): Forderungen der TAppV erfüllt. Wirklich? Fragen und Gedanken zur tiermedizinischen Ausbildung. Deutsches Tierärzteblatt 9/2007, 1100-1101.

**Kopp V, Möltner A & Fischer MR** (2006): Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. GMS Z Med Ausbild 23 (3), Doc 50.

**Krebs R** (2004): Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung. Bern: Institut für Medizinische Lehre IML, Abteilung für Ausbildungs- und Examensforschung AAE.

**Krückeberg J, Paulmann V, Fischer V, Haller H & Matthies HK** (2008): Elektronische Testverfahren als Bestandteil von Qualitätsmanagement und Dynamisierungsprozessen in der medizinischen Ausbildung. GMS Med Inform Biom Epidemiol 2008; 4(2): Doc08.

**Lienert GA & Raatz U** (1998): Testaufbau und Testanalyse (6. Auflage). Beltz, Weinheim.

**Lutz L** (2004): Leitfaden für das Planen, Durchführen und Auswerten von Prüfungen an der ETHZ. Didaktikzentrum der ETH Zürich.

**Miller GE** (1990): The Assessment of Clinical Skills/Competence/Performance. Acad Med 65, 63-67.

**Möltner A, Schellberg D & Jünger J** (2006): Grundlegende quantitative Analysen medizinischer Prüfungen. GMS Z Med Ausbild. 23 (3), Doc53.

**Nikendei C & Jünger J** (2006): OSCE - praktische Tipps zur Implementierung einer klinisch-praktischen Prüfung. GMS Z Med Ausbild 2006, 23(3), Doc 47.

**Pospischil A, Earle M, Erni D & Schneider F** (2007): Planung und Durchführung einer elektronischen Prüfung im Fach Allgemeine Pathologie am Institut für Veterinärpathologie der Universität Zürich (IVPZ) - ein Erfahrungsbericht. In: BRAHM T & SEUFERT S (Hrsg.): "Ne(x)t Generation Learning": E-Assessment und E-Portfolio: halten sie, was sie versprechen? SCIL-Arbeitsbericht 13, St. Gallen (CH).

**Rodriguez MC** (2005): Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. Educational Measurement: Issues and Practice 4/2, 3-13.

**Rotthoff T, Baehring T, Dicken H-D, Fahron U, Fischer Mr, Adler M, Hege I & Scherbaum WA** (2006): Fallbasierte Online-Prüfungen im Medizinstudium – Zielsetzung, Implementierung und praktische Erfahrungen. GMS Med Inform Biol Epidemiol 2 (3), Doc 11.

**TAppV** (2007): Verordnung zur Approbation von Tierärztinnen und Tierärzten vom 27.07.2006 (BGBl. I, 1827), geändert durch Artikel 37 des Gesetzes vom 2.12.2007 (BGBl. I, 2686).

**Wagels R, Feige K & Tipold A** (2008): Einführung und Evaluierung des praktischen Jahres an der Tierärztlichen Hochschule Hannover. GMS Z Med Ausbild 2008;25(4):Doc98.

**Wolf KD** (2007): E-Assessment an Hochschulen: Organisatorische und rechtliche Rahmenbedingungen. In: Brahm T und Seufert S (Hrsg.): "Ne(x)t Generation Learning": E-Assessment und E-Portfolio: halten sie, was sie versprechen? SCIL-Arbeitsbericht 13, St. Gallen (CH).

## Korrespondierender Autor



Dr. med. vet. Jan P. EHLERS, M.A. || eLearning-Berater der  
Stiftung Tierärztliche Hochschule Hannover || Bünteweg 2, 30559  
Hannover || Tel: +49-511 953 8054, Fax: +49-511 953 82 8054  
<http://www.tiho-hannover.de/service/elearning/>

[Jan.Ehlers@tiho-hannover.de](mailto:Jan.Ehlers@tiho-hannover.de)