

Bernhard SPANGL<sup>1</sup>, Dóra KÉRTESZ & Christian F. FREISLEBEN-TEUTSCHER (Wien/St. Pölten)

# Online Two-stage Exams als Prüfungsmethode in Statistik-Lehrveranstaltungen

## Zusammenfassung

Kollaborative Prüfungsszenarien sorgen während der Prüfung durch zusätzliches Feedback für eine weitere Lernerfahrung. Two-stage Exams im Besonderen kombinieren die individuelle Einzelleistung mit Gruppendiskussionen und erlauben es den Studierenden bei Unsicherheit in mehreren Versuchen die richtige Lösung zu erarbeiten. Anhand zweier Statistik-Lehrveranstaltungen wird die Überführung von papierbasierten Einzelprüfungen in Online Two-stage Exams während der Covid-19 Pandemie beschrieben und die Ergebnisse analysiert. Es konnte gezeigt werden, dass die Studierenden in mehrfacher Hinsicht von diesem Format profitieren und noch während der Prüfung ein Lernzuwachs erfolgt.

## Schlüsselwörter

Online Two-stage Exam, kollaborative Prüfungsmethode, Moodle, Videokonferenz, Mint-Fächer

---

<sup>1</sup> E-Mail: [bernhard.spangl@boku.ac.at](mailto:bernhard.spangl@boku.ac.at)



## Online two-stage exams as an examination method in statistics courses

### Abstract

Collaborative examination scenarios provide additional learning experience during an exam by providing additional feedback. Two-stage exams, in particular, combine the individual performance of students with group discussions and allow students to work out a correct solution using several attempts, if they are uncertain. This paper presents the transition from individual, paper-based exams to online two-stage exams in two statistics courses during the Covid-19 pandemic and investigates the examination results. The results show that students benefit from this method in multiple ways and that learning takes place during the examination.

### Keywords

online two-stage exam, collaborative examination method, Moodle, video conferencing, STEM disciplines

## 1 Einleitung

Constructive Alignment beschreibt das Zusammenspiel zwischen den Lernergebnissen, Lehrformaten und Prüfungsmodalitäten einer Lehrveranstaltung in wechselseitiger Abhängigkeit voneinander (BIGGS & TANG, 2011). Innerhalb dieser Triade des Modells bewirkt jegliche Änderung eines der drei Elemente in der Regel auch eine Überprüfung und Anpassung der beiden anderen. Aufgrund ihres hohen Stellenwerts spielen Prüfungen als formale Leistungsüberprüfungen eine Schlüsselrolle im studentischen Lernverhalten (BIGGS & TANG, 2011), weshalb sie geplant als Mittel zur Förderung des studentischen Lernens eingesetzt werden können (EFU, 2019).

Für anwendungsorientierte Lehrveranstaltungen der sogenannten MINT-Fächer aus Mathematik, Ingenieurwesen, Naturwissenschaften und Technik erweisen sich aktive und darunter kollaborative Lehr-Lernformate als förderlich für die Problemlösungskompetenzen der Studierenden (HELLER, KEITH & ANDERSON, 1992;

LINDSLEY et al., 2016). Eine Anpassung der Prüfungsmethoden im Sinne des Constructive Alignment bewirkt durch die Integration von kollaborativen Elementen in das Prüfungsdesign eine konsequente Anbindung an die Lehr-Lernmethoden der Lehrveranstaltung (RIEGER & HEINER, 2014).

Die während der Covid-19-Pandemie erfolgte Umstellung der Hochschullehre auf Online-Lehre ab dem Sommersemester 2020 bewirkte auch einen Umbruch im Prüfungsbetrieb durch den Wegfall der Möglichkeit Prüfungen in physischer Anwesenheit und mit Prüfungsaufsicht vor Ort abzuhalten. Für eine sinnvolle Übertragung der Prüfungen bestehender Statistik-Lehrveranstaltungen auf dislozierte Onlineprüfungen mussten diese nach der Neugestaltung zur Vermeidung einer Schlechterstellung der Studierenden eine Reihe von Anforderungen erfüllen, nämlich (a) die konsequente Fortführung des aktiven und kollaborativen Ansatzes der jeweiligen Lehrveranstaltung, (b) die Beschränkung der Überwachung der Studierenden auf ein notwendiges und noch akzeptables Maß zur Wahrung einer ordnungsgemäßen Durchführung, (c) die Förderung von Lernprozessen auch während der Prüfung und (d) die automatisierte Auswertbarkeit der Prüfungen in und mithilfe der Lernplattform Moodle.

Der Einsatz von Moodle Online-Tests im Rahmen eines kollaborativen Settings sollte ein Erreichen der gesetzten Lernergebnisse sicherstellen und gleichzeitig die Aktivierung der Studierenden auch während der Prüfungen ermöglichen, sodass diese nicht nur, sondern auch durch die und während der Prüfungen lernen. Es war naheliegend, die bereits vorhandenen Prüfungsfragen auch online unter Gewährung erlaubter Hilfsmittel einzusetzen.

## **2 Online Two-stage Exams als kollaboratives Prüfungsformat**

Two-stage Exams sind eine Variante kollaborativer Prüfungsformate, bei denen die Studierenden zunächst ihre Prüfung als Einzelleistung ablegen, um unmittelbar danach in Kleingruppen die gleichen oder überwiegend die gleichen Prüfungsfragen gemeinsam zu beantworten. Die Prüfungsnote wird entweder vorab als fixer, gewichteter Durchschnitt der Einzel- und Gruppenleistung festgelegt (CHEN & KINNIBURGH, 2019; MILLER & JAMES, 2019; NEWTON et al., 2019) oder

in Abhängigkeit der Teilleistungen variabel gestaltet (ZIPP, 2007; LEIGHT et al., 2012). Bei fixer Gewichtung macht die Einzelleistung dabei zwischen 75 bis 90 Prozent der Prüfungsnote aus, die restlichen Prozent des Notenanteils entfallen auf die Gruppenleistung (RIEGER & HEINER, 2014; CHEN & KINNIBURGH, 2019; MILLER & JAMES, 2019; NEWTON et al., 2019).

Während der Gruppenphase können die Studierenden zum einen ihre zuvor erbrachten Einzelleistungen nachträglich auf ihre Richtigkeit überprüfen, zum anderen müssen sie ihre Antworten begründen und vor ihren Peers argumentieren, um innerhalb einer Gruppe Konsens über die zu wählende Lösung zu erzielen, da nur eine Antwort pro Gruppe gewertet wird. Diese Auseinandersetzung mit den unterschiedlichen Standpunkten innerhalb der Gruppe, um zu einer gemeinsamen, einstimmigen Lösung zu kommen, und insbesondere das Feedback der anderen Gruppenmitglieder auf die vorgebrachten Sichtweisen und Argumente bewirkt eine erneute Auseinandersetzung mit dem Lernstoff, die mit dessen tiefergehenden Verarbeitung einhergeht, sodass ein Lernen noch während der Prüfung erfolgt (GILLEY & CLARKSTON, 2014). Die von der Gruppe gefundene Lösung kann unmittelbar auf ihre Richtigkeit geprüft und bei Bedarf unter Punkteabzug so oft abgeändert werden, bis die Frage korrekt beantwortet ist oder keine Punkte mehr für die Frage erzielt werden können. Der Mehrwert dieses unmittelbaren, automatisierten Feedbacks liegt in der Aufklärung von Missverständnissen und etwaigen Fehlkonzepten noch unmittelbar während der Prüfung (EFU, 2019).

Die Zuteilung der Studierenden zu den einzelnen Gruppen erfolgt vorzugsweise auf Basis der erzielten Leistungen (Hausübungen, Wissensüberprüfungen etc.) während des Semesters, sodass die Mitglieder einer Gruppe in etwa das selbe Leistungsniveau aufweisen. Damit soll vermieden werden, dass sehr gute Studierende eher schwache Studierende mittragen und so das tatsächliche Leistungsniveau verzerren (ZIPP, 2007).

Durch unterschiedliche Testfragen in den einzelnen Gruppen, bei denen eventuell nur die Angabe und/oder die Antwortmöglichkeiten leicht variieren, kann sichergestellt werden, dass sich die Studierenden nur innerhalb der Gruppe, nicht aber zwischen den Gruppen austauschen. Auch ist es nicht zwingend notwendig, sowohl für die Einzelphase als auch für die Gruppenphase dieselben Testfragen zu stellen, z. B. könnten im Individualteil zu einem Themenbereich mittels offener Fragen Zusammenhänge abgefragt und in der Gruppenphase zum selben Themenbereich mittels

Konzeptfragen mit vorgegebenen Antwortmöglichkeiten das Verständnis überprüft werden. Jedoch sollten in der Gruppenphase Fragen verwendet werden, die den Studierenden einen Rückschluss auf die Richtigkeit ihrer abgegebenen Antworten während der Einzelphase ermöglichen. Andernfalls entfällt das zeitnahe Feedback auf die Fragen dieser Phase und es kommt weder zu einer Korrektur von Missverständnissen noch zu einem Erfolgsgefühl.

### **3 Material und Methoden**

Der konkrete Einsatz der vorgestellten Lehr-Lern- und Prüfungsformate erfolgte im Studienjahr 2020/21 im Rahmen der Bachelor-Einführungslehveranstaltung „Angewandte Statistik“ und der Master-Vertiefungslehveranstaltung „Statistische Versuchsplanung“. Beide Lehrveranstaltungen wurden nach dem Modell „Inverted Classroom“ (LAGE, PLATT & TREGLIA, 2000) und zur Gänze online abgehalten. Die Studierenden wurden zu Beginn beider Lehrveranstaltungen über das Format „Inverted Classroom“ und den Prüfungsmodus „Two-stage Exam“ informiert.

Die wöchentliche Erarbeitung von Themen durchlief typischerweise vier Phasen: Vorbereitung, Just-in-Time-Teaching (SIMKINS & MAIER, 2010), Vertiefung und Feedback. In der Vorbereitungsphase erarbeiteten sich die Studierenden die Theorie anhand vorgegebener Literatur individuell im Selbststudium vor der Lehrveranstaltungseinheit. Ihre Kommentare und Fragen zu den Lerninhalten wurden schriftlich an den Vortragenden übermittelt, sodass dieser bereits vor der Online-Präsenz eine erste Einschätzung zum Kenntnisstand der Studierenden treffen konnte. In der zweiten Phase, die als Just-in-Time-Teaching konzipiert war, erfolgte mittels Konzeptfragen unter Verwendung von Audio-Response-Systemen in einem „Peer Instruction“-Setting (MAZUR, 1997) die Vertiefung, Korrektur sowie die Festigung des erarbeiteten Wissens. Die Studierenden beantworteten während der Online-Präsenz wiederholt Konzeptfragen zunächst individuell mithilfe der Abstimmungs-funktion in Zoom. Bei zu wenigen richtigen Antworten galt es in der darauf folgenden Diskussion in Kleingruppen von vier bis sechs Personen die Gruppenmitglieder durch fachliche Argumente von der Richtigkeit der getroffenen Wahl zu überzeugen. Die an die Gruppendiskussion anschließende erneute Abstimmung über die ursprüngliche Frage ging schließlich mit einer Auflösung der Frage einher. „Peer

Instruction“ war somit eine direkte Vorbereitung auf die Gruppenphase des „Two-stage Exams“. Die restliche Zeit der Online-Einheit wurde zur Vertiefung der Inhalte genutzt (Phase 3). Am Ende jeder Einheit schließlich konnten die Studierenden anonym ein kurzes 3-Punkte-Feedback auf die Lehrveranstaltungseinheit geben. Die Rücklaufquoten lagen im Durchschnitt bei ca. 20 Prozent.

Die Prüfungen der beiden Lehrveranstaltungen<sup>2</sup> waren als „Online Two-stage Exams“ organisiert, die mithilfe der Lernplattform Moodle durchgeführt wurden. Diese beiden Teilprüfungen, die jeweils unterschiedliche Themengebiete behandelten, konnten entweder zu verschiedenen Terminen (in der Mitte des Semesters und am Semesterende) oder gemeinsam zu Semesterende<sup>3</sup> absolviert werden. Die Prüfungsaufsicht und die Hilfestellung bei Problemen erfolgte mithilfe der Video-Konferenz-Software Zoom.

Die Prüfungseinstellungen während der individuellen ersten Phase erlaubten weder Feedback noch eine Rückmeldung des Punktestandes an die Studierenden. Die Studierenden durchliefen die zweite Phase der Prüfung unmittelbar nach der ersten. Hierfür wurden sie in der laufenden Zoom-Sitzung in Breakout-Sessions einer Gruppe von vier bis sieben Personen zugeteilt. In jeder Gruppe wurde einem Gruppenmitglied als Gruppensprecher\*in die Möglichkeit eingeräumt, den eigenen Bildschirm für die Dauer der zweiten Phase mit der restlichen Gruppe zu teilen. Anschließend durchliefen die Gruppen parallel einen der ersten Phase ähnlichen Online-Test, wobei nur der\*die Gruppensprecher\*in die Möglichkeit hatte, die Antwort der Gruppe in Moodle einzugeben. Im Unterschied zur ersten Phase hatten die Studierenden die Möglichkeit eines (mehrmaligen) Prüfens der eingegebenen Antwort(en) mit adaptivem Punkteabzug. Als Feedback erhielten die Studierenden „richtig“ oder „falsch“. Die Prüfungsdauer betrug 20 Minuten für die Einzelphase und 40 Minuten für die Gruppenphase. Im Anschluss an alle abgehaltenen Two-stage Exams wurde von den teilnehmenden Studierenden ein Feedback zu diesem Prüfungsmodus eingeholt.

---

2 Die Gesamtnote der Lehrveranstaltung „Angewandte Statistik“ setzte sich aus den beiden Teilprüfungen und mehreren, über das gesamte Semester verteilten praktischen Projektarbeiten, auf die hier nicht näher eingegangen wird, zusammen.

3 Der Wiederholungstermin der ersten Teilprüfung zu Semesterende fand zeitlich vor dem zweiten Teilprüfungstermin statt.

### 3.1 Die Lehrveranstaltung „Angewandte Statistik“

Die Lehrveranstaltung „Angewandte Statistik“ im Ausmaß von 2 ECTS ist für Studierende des Bachelorstudiums Landschaftsplanung und Landschaftsarchitektur an der Universität für Bodenkultur Wien verpflichtend. Sie wird als integrierte Lehrveranstaltung (VU) mit immanentem Prüfungscharakter abgehalten und findet regulär laut Studienplan im 5. Semester statt. Der Lehrveranstaltungstyp „VU“ sieht mehrere Leistungsfeststellungen während des Semesters vor. In der Lehrveranstaltung „Angewandte Statistik“ erfolgen diese in Form von kurzen Tests vor jeder Einheit und mehreren, über das Semester verteilten praktischen Projektarbeiten, die es erlauben, die Studierenden aufgrund dieser Leistungen für die erste Teilprüfung in homogene Gruppen einzuteilen. Im Wintersemester 2020/21 waren 124 Studierende zu dieser Veranstaltung angemeldet. Die statistischen Kennzahlen der Prüfungsergebnisse sind in Tabelle 1 aufgelistet.

Tab. 1: Statistische Kennzahlen der Prüfungsergebnisse (Lehrveranstaltung „Angewandte Statistik“).

	Testgruppe A		1. Teilprüfung				2. Teilprüfung			
			Testgruppe B		Wiederholung		Testgruppe A		Testgruppe B	
	E	G	E	G	E	G	E	G	E	G
Anzahl <sup>4</sup>	38	7	53	9	7	2	42	8	51	8
Mögliche Punkteanzahl	13.00		13.00		12.00 <sup>5</sup>		12.00		12.00	
Min. Punkteanzahl	6.00	12.17	6.00	11.67	8.00	11.00	6.00	10.50	2.00	9.00
Mittelwert	10.67	12.65	10.49	12.49	10.43	11.14	9.74	11.43	7.57	11.02
Standardabweichung	1.73	0.27	1.53	0.53	1.51	0.24	1.74	0.48	2.13	0.96
Max. Punkteanzahl	13.00	13.00	13.00	13.00	12.00	11.50	12.00	12.00	11.00	12.00

4 In der Spalte „E“ (Einzelphase) ist die Anzahl der Studierenden, in der Spalte „G“ (Gruppenphase) die Anzahl der Breakout-Sessions angegeben.

5 Eine der 13 Testfragen der ersten Teilprüfung wurde zum Wiederholungstermin gestrichen und nicht durch eine Alternative ersetzt.

### 3.2 Die Lehrveranstaltung „Statistische Versuchsplanung“

Die Lehrveranstaltung „Statistische Versuchsplanung“ im Ausmaß von 3 ECTS ist für Studierende des Masterstudiums Nutzpflanzenwissenschaften an der Universität für Bodenkultur Wien verpflichtend. Sie wird als Vorlesung (VO) mit abschließender Prüfung abgehalten. Da dieser Lehrveranstaltungstyp keine weiteren Leistungsfeststellungen vorsieht, erfolgte die Gruppenzuteilung der Studierenden für die erste Teilprüfung im Unterschied zur Lehrveranstaltung „Angewandte Statistik“ zufällig. Für die zweite Teilprüfung konnte eine Zuordnung zu einer Gruppe aufgrund der während der ersten Teilprüfung erzielten Leistungen erfolgen. Im Sommersemester 2021 waren 78 Studierende angemeldet. Die statistischen Kennzahlen der Prüfungsergebnisse sind in Tabelle 2 aufgelistet.

Tab. 2: Statistische Kennzahlen der Prüfungsergebnisse (Lehrveranstaltung „Statistische Versuchsplanung“).

	1. Teilprüfung				2. Teilprüfung			
	Testgruppe A		Wiederholung		Testgruppe A		Testgruppe B	
	E	G	E	G	E	G	E	G
Anzahl <sup>6</sup>	26	5	8	2	18	4	13	3
Mögliche Punkteanzahl	12.00		12.00		12.00		12.00	
Min. Punkteanzahl	4.25	11.00	4.00	11.8	4.00	8.34	3.33	8.67
Mittelwert	9.22	11.78	10.44	11.9	7.13	9.84	6.89	10.60
Standardabweichung	2.39	0.42	2.70	0.11	2.05	1.16	2.04	1.34
Max. Punkteanzahl	12.00	12.00	12.00	12.00	10.75	11.42	10.00	11.50

### 3.3 Statistische Analyse

Die statistische Analyse wurde mit der Statistik-Software R (R CORE TEAM, 2021) durchgeführt. Zur Überführung der Gleichheit bzw. Unterschiedlichkeit der Testgruppen zu den einzelnen Teilprüfungsterminen wurden Äquivalenztests (TOST;

<sup>6</sup> In der Spalte „E“ (Einzelfase) ist die Anzahl der Studierenden, in der Spalte „G“ (Gruppenphase) die Anzahl der Breakout-Sessions angegeben.



R Paket TOSTER; LAKENS, 2017) bzw. ein t-Test verwendet. Für die Äquivalenztests wurde als Schranke  $d = 1.5$  gewählt.<sup>7</sup> Für den statistischen Nachweis einer Leistungssteigerung zwischen Einzel- und Gruppenphase wurden gepaarte t-Tests verwendet. P-Werte mit  $p < 0.05$  wurden als statistisch signifikant erachtet. Die Wiederholungstermine wurden aufgrund der geringen Stichprobengrößen nicht statistisch ausgewertet.

## 4 Ergebnisse & Evaluierung

Der folgende Abschnitt analysiert die Prüfungsergebnisse der Studierenden beider Lehrveranstaltungen.

### 4.1 Die Lehrveranstaltung „Angewandte Statistik“

Die durchschnittlichen erreichten Punkte in den beiden Testgruppen der 1. Teilprüfung bezogen jeweils auf die Einzel- (TOST:  $p < 0.001$ ) und Gruppenphase (TOST:  $p < 0.001$ ) sind gleich. Hingegen unterscheiden sich die durchschnittlichen Einzelleistungen in den beiden Testgruppen der 2. Teilprüfung signifikant (t-Test:  $p < 0.001$ ). Die durchschnittlichen Gruppenleistungen in den beiden Testgruppen der 2. Teilprüfung sind wieder vergleichbar (TOST:  $p < 0.001$ ).

Abbildung 1 zeigt die Ergebnisse für das Two-stage Exam im Detail. Auf der vertikalen Achse sind die erzielten Gesamtpunkte aufgetragen. Die Linien entsprechen den einzelnen Studierenden und verbinden die erzielten Gesamtpunkte in der Einzelphase (jeweils links in den Grafiken der Abbildung 1) mit den erreichten Punkten in der Gruppenphase (jeweils rechts in den Grafiken der Abbildung 1). Die Farben entsprechen den Gruppen der Gruppenphase. Die Gruppen sind entsprechend der durchschnittlichen Übungsleistung der Studierenden absteigend angeordnet.

---

<sup>7</sup> Das entspricht in etwa einem Notengrad bei einer Maximalpunktzahl von 12 Punkten und dem Notenschlüssel: „Nicht Genügend“ ( $< 50\%$ ), „Genügend“ ( $< 62.5\%$ ), „Befriedigend“ ( $< 75\%$ ), „Gut“ ( $< 87.5\%$ ), „Sehr Gut“ (sonst).

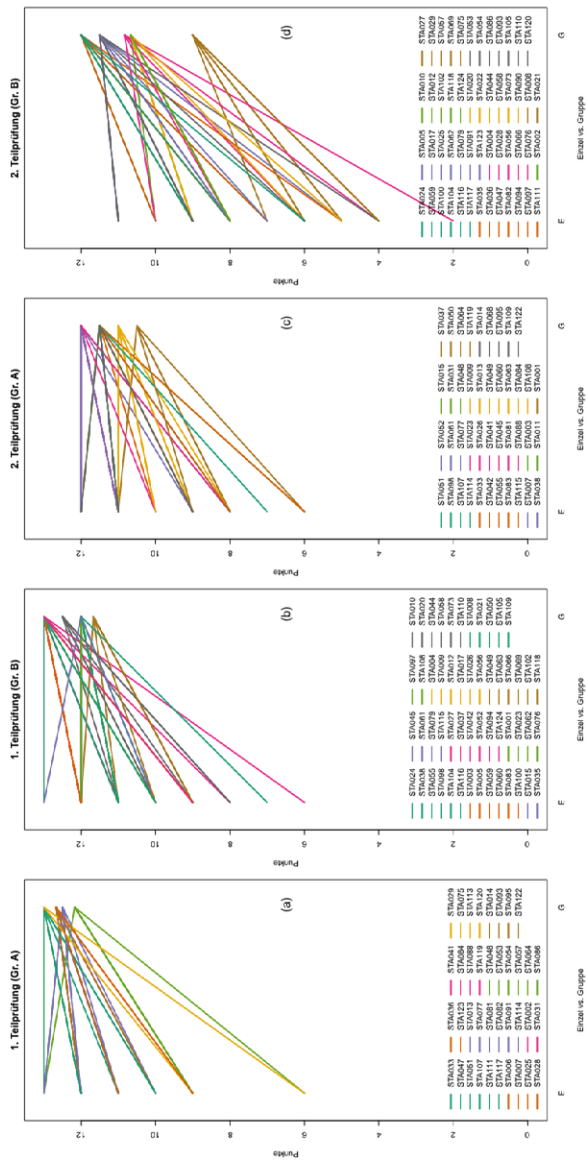


Abb. 1: Punkteverteilung des Two-stage Exams im Detail (Lehrveranstaltung „Angewandte Statistik“). Die Linien entsprechen den einzelnen Studierenden und verbinden die erzielten Gesamtpunkte in der Einzelphase E mit den erreichten Punkten in der Gruppenphase G. Die Farben entsprechen den Gruppen der Gruppenphase.

Bei allen Studierenden kam es in der Gruppenphase im Vergleich zur jeweiligen Einzelleistung zu einer statistisch signifikanten Steigerung der erzielten Punkte (gepaarte t-Tests; 1. Teilprüfung, Testgruppe A: mittlere Steigerung um 1.89 Punkte,  $p < 0.001$ , Testgruppe B: mittlere Steigerung um 2.00 Punkte,  $p < 0.001$ ; 2. Teilprüfung, Testgruppe A: mittlere Steigerung um 1.69 Punkte,  $p < 0.001$ , Testgruppe B: mittlere Steigerung um 3.45 Punkte,  $p < 0.001$ ). Die Verbesserung der Leistungen in der Gruppenphase legt nahe, dass in dieser Phase ein Lernzuwachs stattfand. Diese Annahme deckt sich mit den in der Literatur publizierten Ergebnissen (vgl. etwa GILLEY & CLARKSTON, 2014; LEIGHT et al., 2012; STEARNS, 1996) und dem Feedback der Studierenden. Die hier präsentierten Analyseergebnisse bestätigen auch Auswertungen der Prüfungsergebnisse aus einem früheren Durchgang im Wintersemester 2019/20 (vgl. SPANGL & KERTESZ, 2020).

## **4.2 Die Lehrveranstaltung „Statistische Versuchsplanung“**

Die Gleichheit der durchschnittlich erreichten Punkte in den beiden Testgruppen der 2. Teilprüfung bezogen jeweils auf die Einzel- (TOST:  $p = 0.052$ ) und Gruppenphase (TOST:  $p = 0.061$ ) kann statistisch knapp nicht verifiziert werden.

Wiederum kam es, wie in Abbildung 2 dargestellt, bei fast allen Studierenden in der Gruppenphase im Vergleich zur jeweiligen Einzelleistung zu einer statistisch signifikanten Steigerung der erzielten Punkte (gepaarte t-Tests; 1. Teilprüfung: mittlere Steigerung um 2.56 Punkte,  $p < 0.001$ ; 2. Teilprüfung, Testgruppe A: mittlere Steigerung um 2.71 Punkte,  $p < 0.001$ , Testgruppe B: mittlere Steigerung um 3.71 Punkte,  $p < 0.001$ ).

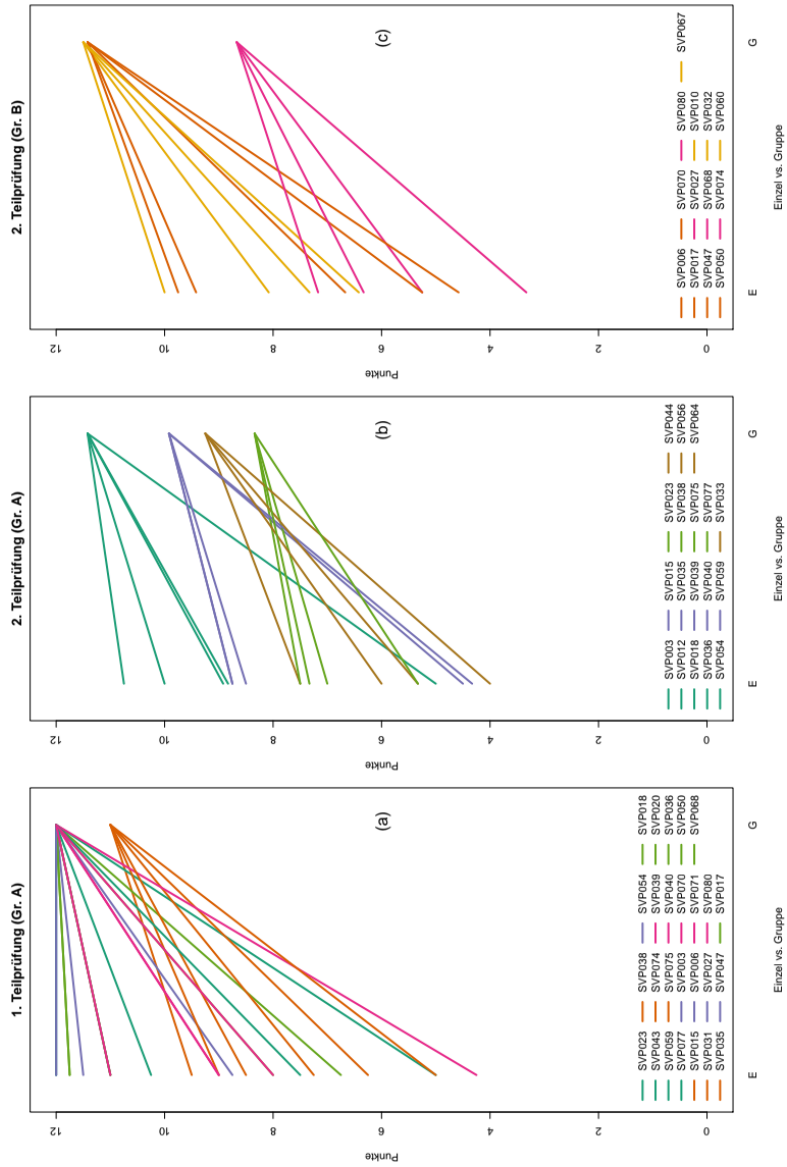


Abb. 2: Punkteverteilung des Two-stage Exams im Detail (Lehrveranstaltung „Statistische Ver- suchsplanung“). Linien und Farbcodierung sind analog zu Abbildung 1.

Da eine Zuordnung der Studierenden zu homogenen Gruppen aufgrund einer fehlenden Leistungsfeststellung nicht möglich war, wurden die Studierenden zufällig in Gruppen aufgeteilt. Diese zufällige Zuordnung der Studierenden wird in Abbildung 2a besonders deutlich. Auch heterogene Gruppen, bezogen auf die individuellen Einzelleistungen innerhalb der Gruppe, erreichten in der Gruppenphase die maximale Punkteanzahl. Leistungsstarke Studierende zogen vermutlich die schwächeren mit und verzerren so das tatsächliche Leistungsniveau. Die leistungsabhängige Zuteilung für die 2. Teilprüfung ist in Abbildungen 2b und 2c gut ersichtlich; die Einzelleistungen in den Gruppen sind deutlich homogener und die Ergebnisse der Gruppenphase entsprechen dem in der Einzelphase erbrachten Leistungsniveau der Studierenden.

## 5 Diskussion

Der Mehrwert des Two-stage Exams entstand für die Studierenden vor allem in der Gruppenphase. Die inhaltliche Auseinandersetzung innerhalb der Gruppe und die Rückmeldungen der anderen Gruppenmitglieder auf die vorgebrachten Argumente bewirken eine erneute Auseinandersetzung mit dem Lernstoff und führen zu einem Wissenserwerb noch während der Prüfung. In ihrem Feedback zum Prüfungsformat bestätigten die Studierenden in Übereinstimmung mit der Literatur eine Festigung des Gelernten und die Korrektur der im Individualteil gemachten Fehler. Die statistische Auswertung in Abschnitt 4 belegt eine Verbesserung der Ergebnisse während der Gruppenphase.

Eine Überwachung der Studierenden während der Online-Prüfung ist nur eingeschränkt möglich. Eine Kollaboration über Kanäle außerhalb der verwendeten Video-Konferenz-Software und Prüfungsplattform kann nicht unterbunden und auch nicht nachgewiesen werden.

Peer Instruction und Two-stage Exams fördern kommunikative und kooperative Kompetenzen der Studierenden (RIEGER & HEINER, 2014; KINNEAR, 2021). Damit Online Two-stage Exams erfolgreich durchgeführt werden können, müssen die Studierenden in der Lage sein, ihre Lösung in Kleingruppen auszuhandeln. Die Studierenden müssen daher zur Vorbereitung auf das Prüfungsformat ausreichend die Möglichkeit erhalten, die hierfür benötigten Kompetenzen zu erwerben bezie-

ungsweise zu vertiefen. Diese Kompetenzen werden jedoch in den Lehrveranstaltungszielen nicht gesondert ausgewiesen. Auch der Beitrag für nachfolgende Lehrveranstaltungen ist unklar.

Die Lernziele der beschriebenen Lehrveranstaltungen adressieren unterschiedliche Kompetenzen, die durch das gewählte Prüfungsformat nur unvollständig abgedeckt werden können. Hier bedarf es einer genaueren Passung der Lernziele, der Lehr-Lernmethoden und der Prüfungsmodalitäten. Nach Absolvierung der vorliegenden Lehrveranstaltungen sollten die Studierenden statistische Aufgabenstellungen rechnerisch lösen und die Ergebnisse interpretieren sowie auf Plausibilität prüfen können. Für die beiden letzteren Ziele wurden die Lehr-Lernmethoden und die Leistungsüberprüfung mithilfe von Two-stage Exams aufeinander abgestimmt. Das praktische Lösen statistischer Fragestellungen erfolgt innerhalb der Lehrveranstaltung „Angewandte Statistik“ im Rahmen von Übungsbeispielen; für die Vorlesung „Statistische Versuchsplanung“ wurde hingegen dieser Kompetenzerwerb in eine eigene Lehrveranstaltung ausgelagert. Die Wechselwirkung zwischen den Übungsbeispielen, die als Rechenaufgaben konzipiert sind, den Peer-Instruction-Einheiten beziehungsweise den Two-stage Exams wurde bisher jedoch weder adressiert noch untersucht. Es ist daher unklar, in welchem Ausmaß die Rechenaufgaben zum Konzeptverständnis beitragen und der Prüfungsvorbereitung dienen. Die Auswirkungen der Prüfungen auf die Problemlösungskompetenzen der Studierenden vor dem Hintergrund einer praktischen Anwendung des Gelernten wurde bisher ebenso wenig berücksichtigt.

In Online Two-stage Exams sind während der Gruppenphase nur geschlossene Prüfungsfragen, die eine sofortige und automatische Benotung ermöglichen, einsetzbar. Andernfalls entfällt das zeitnahe Feedback, das einen Rückschluss auf die Richtigkeit der gegebenen Antworten zulässt, und es kommt weder zu einer Korrektur von Missverständnissen noch zu einem Erfolgsgefühl der Studierenden. Einen entsprechenden Fragenpool vorausgesetzt, erfordert das Format einen Mehraufwand bei der Gruppenzuteilung der Studierenden. Zusätzlich verdreifacht sich die Prüfungsdauer aufgrund der Diskussionen in der Gruppenphase.

Ein langfristiger Lerntransfer aufgrund des Two-stage Exam konnte in der Literatur nicht eindeutig belegt werden (CHEN & KINNIBURGH, 2019). Weiterer Forschungsbedarf besteht in Hinblick auf einen nachhaltigen Lerntransfer und in welchem Ausmaß die Studierenden das erworbene Wissen für ihre Bachelor- (Landschaftsplanung) bzw. Masterarbeit (Nutzpflanzenwissenschaften) anwenden können.

## 6 Fazit

Die Überführung von papierbasierten Einzelprüfungen in Online Two-stage Exams während der Covid-19-Pandemie an der Universität für Bodenkultur Wien unterstrich das große Potenzial für das Erlangen einer gesteigerten Problemlösungskompetenz in Statistik-Lehrveranstaltungen. Das Feedback der Studierenden bestätigte den zusätzlichen Kompetenzerwerb während der Gruppenphase.

Two-stage Exams erfordern das Argumentieren der gewählten Lösung durch die Studierenden. Diese gaben in ihren Rückmeldungen zu den Lehrveranstaltungen an, dass ihnen die Diskussionen während der Gruppenphase besonders gut gefielen. Sie stimmten mehrheitlich auch für die Beibehaltung dieses Prüfungsformats.

Um den Einfluss des Prüfungsmodus Two-stage Exam auf den Lerntransfer näher zu untersuchen, sind im kommenden Abhaltungszyklus der Lehrveranstaltungen Cross-over-Studien, wie in LEIGHT et al. (2012) oder GILLEY & CLARKSTON (2014) beschrieben, geplant.

### Danksagung

Wir danken den anonymen Gutachter\*innen für die wertvollen Anmerkungen, die wesentlich zur Verbesserung des Manuskripts beigetragen haben.

## 7 Literaturverzeichnis

**Biggs, J. B., & Tang, C. S.** (2011). *Teaching for quality learning at university: What the student does* (4. ed). Maidenhead: McGraw-Hill, Society for Research into Higher Education & Open University Press.

**Chen, S., & Kinniburgh, S.** (2019). A Controlled Experiment on Two-Stage Exams in an Introductory Statistics Course. *The International Journal of Assessment and Evaluation*, 26(2), 1–12. <https://doi.org/10.18848/2327-7920/CGP/v26i02/1-12>

**Efu, S. I.** (2019). Exams as Learning Tools: A Comparison of Traditional and Collaborative Assessment in Higher Education. *College Teaching*, 67(1), 73–83. <https://doi.org/10.1080/87567555.2018.1531282>

**Gilley, B., & Clarkston, B.** (2014). Research and Teaching: Collaborative Testing: Evidence of Learning in a Controlled In-Class Study of Undergraduate Students. *Journal of College Science Teaching*, 043(03). [https://doi.org/10.2505/4/jcst14\\_043\\_03\\_83](https://doi.org/10.2505/4/jcst14_043_03_83)

**Heller, P., Keith, R., & Anderson, S.** (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, 60(7), 627–636. <https://doi.org/10.1119/1.17117>

**Kinnear, G.** (2021). Two-Stage Collaborative Exams have Little Impact on Subsequent Exam Performance in Undergraduate Mathematics. *International Journal of Research in Undergraduate Mathematics Education*, 7, 33–60. <https://doi.org/10.1007/s40753-020-00121-w>

**Lage, M. J., Platt, G. J., & Treglia, M.** (2000). Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment. *The Journal of Economic Education*, 31(1), 30–43. <https://doi.org/10.2307/1183338>

**Lakens, D.** (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362. <https://doi.org/10.1177/1948550617697177>

**Leight, H., Saunders, C., Calkins, R., & Withers, M.** (2012). Collaborative Testing Improves Performance but Not Content Retention in a Large-Enrollment Introductory Biology Class. *CBE Life Sciences Education*, 11(4), 392–401. <https://doi.org/10.1187/cbe.12-04-0048>

**Lindsley, J. E., Morton, D. A., Pippitt, K., Lamb, S., & Colbert-Getz, J. M.** (2016). The Two-Stage Examination: A Method to Assess Individual Competence and Collaborative Problem Solving in Medical Students. *Academic Medicine*, 91(10), 1384–1387. <https://doi.org/10.1097/ACM.0000000000001185>

**Mazur, E.** (1997). *Peer instruction: A user's manual*. Upper Saddle River: Prentice Hall.

**Miller, S. T., & James, C. R.** (2019). *What Is The Impact Of Collaborative Exams On Learning And Attitudes In Introductory Astronomy Classes?* 6(1), 13–17. <https://clutejournals.com/index.php/JAESE/article/view/10289>

**Newton, G., Rajakaruna, R., Kulak, V., Albabish, W., Gilley, B. H., & Ritchie, K.** (2019). Two-Stage (Collaborative) Testing in Science Teaching: Does It Improve Grades on Short-Answer Questions and Retention of Material? *Journal of College Science Teaching*, 48(4), 64–73. <https://www.jstor.org/stable/26901301>



**R Core Team** (2021). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>

**Rieger, G., & Heiner, C.** (2014). Examinations That Support Collaborative Learning: The Students' Perspective. *Journal of College Science Teaching*, 043(04). [https://doi.org/10.2505/4/jcst14\\_043\\_04\\_41](https://doi.org/10.2505/4/jcst14_043_04_41)

**Simkins, S., & Maier, M. H. (Hrsg.)**. (2010). *Just-in-time teaching: Across the disciplines, across the academy* (1st ed). Sterling: Stylus Pub.

**Spangl, B., & Kertesz, D.** (2020). Wenn Studierende bei der Statistikprüfung „Juhu“ rufen ...: Two-Stage Exams mit Moodle und Zoom in Zeiten von Corona. In J. Weißenböck, W. Gruber & Ch. Freisleben-Teutscher (Hrsg.), *Digital Learning in Zeiten von Corona – nachhaltiger Entwicklungsschub für die Hochschulen?* (S. 65–76). St. Pölten: Fachhochschule St. Pölten GmbH.

**Stearns, S. A.** (1996). Collaborative Exams as Learning Tools. *College Teaching*, 44(3), 111–112. <https://doi.org/10.1080/87567555.1996.9925564>

**Zipp, J. F.** (2007). Learning by Exams: The Impact of Two-Stage Cooperative Tests. *Teaching Sociology*, 35(1), 62–76. <https://doi.org/10.1177/0092055X0703500105>

## Autor\*in\*en



DI Dr. Bernhard Spangl || Universität für Bodenkultur Wien,  
Department für Raum, Landschaft und Infrastruktur, Institut für  
Statistik || Gregor-Mendel-Str. 33, A-1180 Wien

<https://boku.ac.at/rali/stat>

[bernhard.spangl@boku.ac.at](mailto:bernhard.spangl@boku.ac.at)



DI Dóra Kértész || FH Technikum Wien, Teaching & Learning  
Center || Höchstädtplatz 6, A-1200 Wien

<https://www.technikum-wien.at/das-teaching-and-learning-center-der-fhtw/>

[dora.kertesz@technikum-wien.at](mailto:dora.kertesz@technikum-wien.at)



Mag. Dr. Christian Freisleben-Teutscher || FH St. Pölten, SKILL  
(Service- und Kompetenzzentrum für Innovatives Lehren und  
Lernen) || Campus-Platz 1, A-3100 St. Pölten

<https://skill.fhstp.ac.at/>

[christian.freisleben-teutscher@fhstp.ac.at](mailto:christian.freisleben-teutscher@fhstp.ac.at)