

Sebastian Wirthgen, Kathrin Munt, Frauke Stenzel, Kathrin Gläser & Anika Fricke

Erfahrungen mit dem Classroom Observation Protocol for Undergraduate STEM (COPUS) in MINT und nicht-MINT Hochschulkursen

Zusammenfassung

An vielen deutschen Hochschulen finden Entwicklungsbestrebungen statt, die Hochschullehre zu verändern. Hierbei ist es ein Ziel, mehr studierendenzentrierte Lehre in die deutschen Hörsäle zu bringen. Diese Bestrebungen und Prozesse benötigen – für eine gezielte und kontinuierliche Entwicklung – Informationen darüber, was tatsächlich im Hörsaal passiert. Das Classroom Observation Protocol for Undergraduate STEM (COPUS) ist ein verlässliches Beobachtungswerkzeug, welches solche Informationen liefert und in den USA für MINT-Lehrveranstaltungen eingesetzt wird. Im Rahmen dieses Artikels wird untersucht, inwieweit sich dieser COPUS-Bogen in nicht-MINT-Lehrveranstaltungen nutzen lässt. Dazu wurden Lehrveranstaltungen aus dem MINT und nicht-MINT (nMINT) Bereich mit dem COPUS-Bogen dokumentiert und miteinander verglichen. Weiterhin wurden die Ergebnisse auf ihren Informationsgehalt über das Ausmaß an studierendenzentrierter Lehre untersucht. Die Auswertungen zeigen, dass der Bogen für MINT und nicht-MINT-Lehrveranstaltungen einsetzbar ist und reliable sowie valide Beobachtungsergebnisse darüber liefert, was in den Hörsälen passiert.

Schlüsselwörter (max. 5)

Beobachtungsprotokoll, Hospitation, Studierendenzentrierung, COPUS

1 Einleitung

Die zunehmende politische und gesellschaftliche Fokussierung auf die Weiterentwicklung der Hochschullehre ist bereits seit einigen Jahren in Deutschland zu beobachten. Durch zahlreiche Länderprojekte sowie durch das bundesweite Entwicklungsprogramm des Bundesministeriums für Bildung und Forschung (BMBF), dem Qualitätspakt Lehre (QPL), wird an so vielen deutschen Hochschulen wie noch nie gleichzeitig an der Qualitätsverbesserung der Hochschullehre gearbeitet (BMBF, 2015). Ein Aspekt hiervon sind die zahlreichen hochschuldidaktischen Weiterbildungsprogramme und Beratungsangebote für Lehrende, welche zum Ziel haben, die Lehrqualität zu verbessern.

Als ein wesentlicher Bestandteil dieser Verbesserung wird der „Shift from Teaching to Learning“ beschrieben (Berendt, 1998; Wildt, 2001), welcher einen Wandel hin zu mehr studierendenzentrierter Lehre beinhaltet. Hierbei zeichnet sich studierendenzentrierte Lehre durch einen hohen Anteil von Denk-, Arbeits- und Sprechphasen der Studierenden mit dem Fokus auf eine vertiefte Informationsverarbeitung aus (Johannes & Seidel, 2012; Seidel & Hoppert 2011). Zahlreiche Studien zeigen, dass Studierendenzentrierung zu mehr Lernerfolg im Sinne von Kompetenzerwerb führt als lehrendenzentrierte Lehre (Trigwell & Prosser, 2004; Braun & Hannover, 2011). Einhergehend fokussieren viele hochschuldidaktische Maßnahmen und lehrbezogene Entwicklungsprozesse an Hochschulen auf die Erhöhung der Studierendenzentrierung von Lehrveranstaltungen und Studiengangsstrukturen (DGHD, 2013).

Wie sich studierendenzentrierte Lehre in Lehrveranstaltungen umsetzen lässt, ist in der Literatur vielfach beschrieben (Weimer, 2002; Beichner et al., 2007; Gauci, Dantas, Williams & Kemm, 2009). Jedoch sind kaum Informationen darüber vorhanden, in welcher Art und Weise die Lehrenden und Studierenden in Lehrveranstaltungen handeln. Ein Beobachtungsbogen, welcher das Verhalten bzw. das Handeln der Lehrenden und Studierenden in der Lehrveranstaltung dokumentiert, kann Informationen darüber liefern, in welchem Ausmaß Studierendenzentrierung in den beobachteten Veranstaltungen vorliegt. Diese Informationen lassen einen Rückschluss zu, in welchem Grad hochschuldidaktische Maßnahmen wie Einzelberatung von Hochschullehrenden, Weiterbildungsangebote, Curriculumentwicklung und Entwicklungsprozesse in Fakultäten zur Realisierung von mehr Studierendenzentrierung führen (Turpen & Finkelstein, 2009).

In Deutschland gibt es eine langjährige Tradition, mit offenen oder halbstandardisierten Beobachtungsbögen Unterrichtsbeobachtungen in der Lehramtsausbildung durchzuführen. Ziel dieser Beobachtung ist die individuelle Rückmeldung zur Umsetzung bzw. Wirksamkeit didaktischer Methodik in bestimmten Lehrsettings (Havers & Toepell, 2002; Óhidy, Terhart & Zsolnai, 2007). Daneben gibt es sowohl im Schul- als auch Hochschulbereich forschungsorientierte Beobachtungsinstrumente, welche standardisiert und strukturiert sind. Beispiele hierfür sind das Reformed Teaching Observation Protocol (Piburn & Sawada, 2000), Teaching Dimensions Observational Protocol (Hora, Oleson & Ferrare, 2013), Beobachtungsbogen für Lehrveranstaltungen (Wibbecke, Wibbecke, Kahmann & Kadmon, 2016) und das Classroom Observation Protocol for Undergraduate STEM (Smith, Jones, Gilbert & Wieman, 2013). Generell lassen sich Beobachtungsinstrumente für das

Lehrverhalten an Hochschulen in zwei Gruppen unterteilen. Instrumente der einen Gruppe fokussieren die Qualität von Lehrverhalten bzw. didaktischen Handlungen. Die andere Gruppe schließt Instrumente ein, die vorrangig die Quantität dieser Handlungen in den Blick nehmen (Hora & Ferrare, 2013).

Ein strukturiertes und standardisiertes Instrument unabhängig von der fachlichen Disziplin der zu beobachteten Lehrveranstaltung zu nutzen, ist aus mehreren Gesichtspunkten sinnvoll. Zum einen ist das Konstrukt der Studierendenorientierung in seiner Beschreibung und Projektion auf das Handeln von Lehrenden und Studierenden in Lehrveranstaltungen von der Fachdisziplin und dem zu bearbeitenden Inhalt losgelöst definiert (Johannes & Seidel, 2012; Seidel & Hoppert, 2011; Lueddeke, 2003; Trigwell & Prosser, 2004). Zudem richten sich die meisten hochschuldidaktischen Maßnahmen und Angebote, die das Ziel haben, studienorientierte Lehre zu forcieren, fachunspezifisch an die Lehrenden oder Hochschulorganisationen (Berendt, Voss & Wildt, 2002). Zum anderen ist ein standardisierter Beobachtungsbogen mit wenig Arbeitsaufwand im Bereich von Beobachterschulung und Datenauswertung in einer breiten Facette von Lehrveranstaltungen einsetzbar. Zuletzt bietet die Fokussierung auf das Handeln von Lehrenden und Studierenden die Möglichkeit, andere Informationen außer Acht zu lassen und somit konkrete und belastbare Informationen zur anvisierten Fragestellung, dem Ausmaß an Studienorientierung, zu sammeln (Kromrey, 2009).

Ausgehend von der Überlegung, dass das Handeln der Lehrenden in der Lehrveranstaltung einen maßgeblichen Einfluss auf den Lernprozess der Studierenden in dieser hat, entwickelten Smith, Jones, Gilbert und Wieman (2013) das Classroom Observation Protocol for Undergraduate STEM (COPUS). Ihr Ziel war es, mit diesem Instrument auf breiter Ebene einheitliche Daten darüber zu erfassen, womit die Lehrenden und Studierenden ihre Zeit in der Lehrveranstaltung verbringen. Das Beobachtungsinstrument COPUS dokumentiert hierbei das Verhalten der Studierenden und Lehrenden anhand von 25 Kategorien (12 für Lehrende und 13 für Studierende) in kleinen Zeitintervallen von je zwei Minuten (siehe Tabelle 1).

Das COPUS wurde entwickelt für MINT-Lehrveranstaltungen, deckt aber ansonsten viele der oben beschriebenen Aspekte an ein von uns gefordertes Beobachtungsinstrument ab.

Tab. 1: Beschreibung der COPUS-Kategorien für das Lehrenden- und Studierendenhandeln (Smith et al., 2013, S. 620)

Students are doing	
L	Listening to instructor/taking notes, etc.
Ind	Individual thinking/problem solving. Only mark when an instructor explicitly asks students to think about a clicker question or another question/problem on their own.
CG	Discuss clicker question in groups of 2 or more students
WG	Working in groups on worksheet activity

Tabelle wird fortgesetzt

Students are doing	
OG	Other assigned group activity, such as responding to instructor question
AnQ	Student answering a question posed by the instructor with rest of class listening
SQ	Student asks question
WC	Engaged in whole class discussion by offering explanations, opinion, judgment, etc. to whole class, often facilitated by instructor
Prd	Making a prediction about the outcome of demo or experiment
SP	Presentation by student(s)
TQ	Test or quiz
W	Waiting (instructor late, working on fixing AV problems, instructor otherwise occupied, etc.)
O	Other – explain in comments
Instructor is doing	
Lec	Lecturing (presenting content, deriving mathematical results, presenting a problem solution, etc.)
RtW	Real-time writing on board, doc. projector, etc. (often checked off along with Lec)
FUp	Followup/feedback on clicker question or activity to entire class
PQ	Posing nonclicker question to students (non-rhetorical)
CQ	Asking a clicker question (mark the entire time the instructor is using a clicker question, not just when first asked)
AnQ	Listening to and answering student questions with entire class listening
MG	Moving through class guiding ongoing student work during active learning task
1o1	One-on-one extended discussion with one or a few individuals, not paying attention to the rest of the class (can be along with MG or AnQ)
D/V	Showing or conducting a demo, experiment, simulation, video, or animation
Adm	Administration (assign homework, return tests, etc.)
W	Waiting when there is an opportunity for an instructor to be interacting with or observing/listening to student or group activities and the instructor is not doing so
O	Other – explain in comments

Um zu prüfen, ob das COPUS ein in unserem Sinne geeignetes Beobachtungsinstrument darstellt, widmet sich dieser Artikel folgenden Fragestellungen:

- A) Ist das Beobachtungsinstrument ausreichend reliabel im Rahmen von veränderten Beobachterkonstellationen?
- B) Ist das COPUS auch bei der Hospitation in nicht-MINT (nMINT) Lehrveranstaltungen einsetzbar? Wie viele nMINT spezifische Interaktionsmuster existieren, welche mit dem COPUS nicht dokumentierbar sind?
- C) Liefert die Analyse der Beobachtungsdaten einer Lehrveranstaltung Informationen über das Ausmaß von studierendenzentrierter Lehre?

Im Rahmen der weiteren Betrachtung verstehen wir nMINT Veranstaltungen als Hochschulkurse der Geisteswissenschaften, wie beispielsweise Entwicklungspsychologie, sowie Veranstaltungen der Sozialwissenschaften, wie beispielsweise Betriebswirtschaftslehre.

2 Methodisches Vorgehen

Das COPUS wurde als Beobachtungsbogen für Lehrveranstaltungen in den MINT-Disziplinen konzipiert. Zur Untersuchung der Verwendbarkeit des COPUS in nMINT-Lehrveranstaltungen wurden über zwei Semester von vier Beobachtenden in Hospitationen insgesamt 13 MINT-Lehrveranstaltungen 32-mal und 11 nMINT Lehrveranstaltungen 25-mal mit dem Bogen protokolliert. Die Lehrveranstaltungen umfassten Pflichtveranstaltungen unterschiedlichen Charakters wie Vorlesung, interaktive Vorlesung, Seminar oder Übung. Die Dauer der Lehrveranstaltungen variierte zwischen 60 und 180 Minuten. Es wurden insgesamt 2680 Zwei-Minutenintervalle (1096 in nMINT und 1584 in MINT) protokolliert. Die Anzahl der teilnehmenden Studierenden pro Lehrveranstaltung lag zwischen sechs und 90.

Die Gruppe der Beobachtenden bestand aus vier Personen, wobei drei aus MINT-Disziplinen und eine Person aus einer nMINT-Disziplin stammen. Zwei der MINT-Beobachtenden waren eher unerfahren im Hospitieren und Beobachten von Lehrsituationen. Die anderen beiden galten zum Beginn des Erhebungszeitraums bereits als erfahren. Alle Beobachtenden durchliefen eine Vorbereitung ähnlich der, wie sie im Trainingguide von Smith et al. (2013) beschrieben ist. Aufgrund von Interpretationsunterschieden zwischen den Beobachtenden unterschiedlicher Disziplinen und der Kategorien (Codes) in englischer Sprache erweiterten wir diese Vorbereitung um Diskussionen zur Bedeutung, Wahl und Interpretation einzelner Kategorien.

Zur Prüfung und Verbesserung der Interrater-Reliabilität haben wir den im Folgenden beschriebenen Prozess durchlaufen. Das COPUS wurde in sieben Lehrveranstaltungen von Observationsteams zu je zwei oder drei Personen eingesetzt und die ermittelten Daten miteinander verglichen. In der Wahl der Handlungskategorien pro Zeitintervall zeigten sich zunächst deutliche Unterschiede, wenn die Beobachtenden verschiedenen Fachdisziplinen zugehörig sind. Zur Klärung der Beobachtungsunterschiede und zur Verbesserung der Übereinstimmungsgüte wurde innerhalb der Observationsteams das Beobachtungsergebnis nach dem in Tabelle 2 beschriebenen Leitfaden diskutiert.

Tab. 2: Leitfaden der fragengeleiteten Interpretationsergründung

Nr.	Arbeitsschritte
1.	Feststellung einer unterschiedlichen Interpretation
2.	Fragengeleitete Interpretationsergründung <ol style="list-style-type: none"> a) Wieso wurde diese Kategorie ausgewählt? b) An welchen Punkten der Kategorieerklärung spiegelt sich für den Beobachtenden das tatsächlich Geschehene wider? c) Welche Argumente sprechen gegen die Kategorisierung des jeweils anderen Beobachters? d) Welche Schlüsse lassen die bisherigen Antworten zu? e) Welche Zuordnungen ist demnach die tragfähigere? f) Welche weiteren Indikatoren können hilfreich sein, um die Kategorisierung bei zukünftigen Beobachtungen möglichst ähnlich vorzunehmen?
3.	Konsolidierung der Interpretationen

Die Kategorien des COPUS wurden von Smith, Vinson, Smith, Lewin und Stetzer im Jahr 2014 mit dem Ziel einer Klassifizierung von Lehre zusammengefasst. Zwei wesentliche Klassen/Typen beschrieben Smith et al. 2014 durch die häufige Nutzung von aktivierenden Methoden auf der einen Seite sowie durch die hauptsächliche Präsentation von Inhalten durch den Lehrenden auf der anderen Seite. Zur Prüfung, ob die Beobachtung mit dem COPUS auf Grundlage dieser zusammengefassten Kategorien eine Aussage über ein Ausmaß an Studierendenorientierung in einer Lehrveranstaltung zulässt, wurden diese zusammengefassten Kategorien in Tabelle 3 herangezogen (Smith et al., 2014).

Tab. 3: Zusammengefasste Kategorien des COPUS in Bezug auf Studierendenorientierung (nach Smith et al., 2014)

	Handlungen, die mit Studierendenorientierung assoziiert werden		Handlungen, die nicht mit Studierendenorientierung assoziiert werden	
	COPUS Kategorien	Zusammengefasste Kategorien	COPUS Kategorien	Zusammengefasste Kategorien
Lehrendenhandeln	FUp, PQ, CQ, AnQ, MG, 1o1	Guiding (G)	Lec, RtW, D/V Adm W, O	Presenting (P) Administration (A) Other (OI)
Studierendenhandeln	AnQ, SQ, WC, SP Ind, CG, WG, OG, Prd, TQ	Talking to class (STC) Working (SW)	L W, O	Receiving (R) Other (OS)

3 Ergebnisse

Abbildung 1 zeigt in Gegenüberstellung von nMINT- und MINT-Lehrveranstaltungen die Häufigkeit der einzelnen Handlungskategorien bzgl. der beobachteten Veranstaltungen, sowie in Abbildung 2 den Anteil dieser Handlungen an der Veranstaltungszeit.

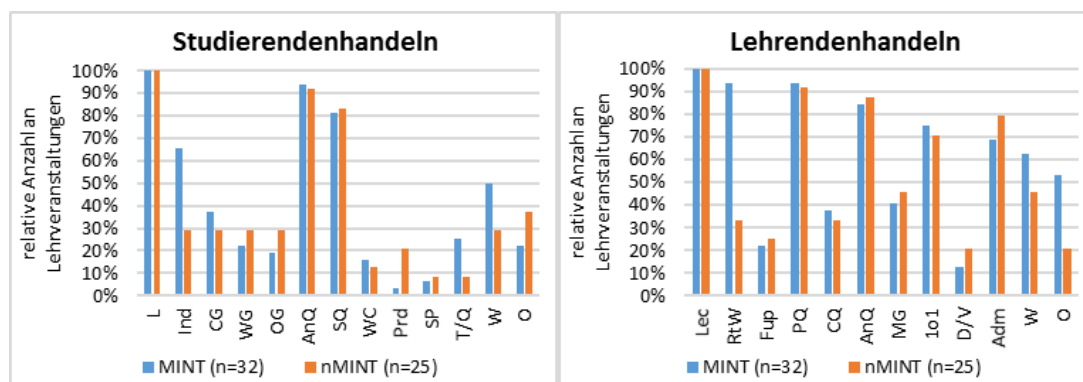


Abb. 1: Relative Anzahl der Lehrveranstaltungen, in denen die jeweiligen Kategorien (vgl. Tabelle 1) beobachtet wurden (32 MINT-Lehrveranstaltungen, 25 nMINT-Lehrveranstaltungen).

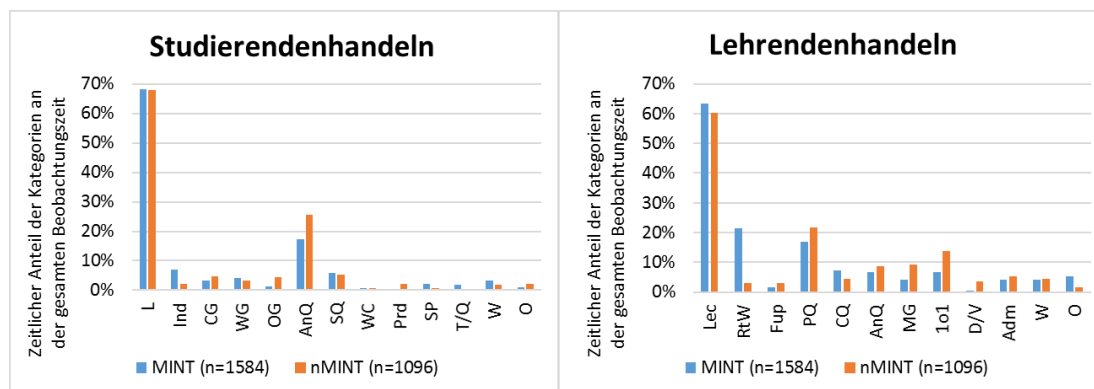


Abb. 2: Zeitlicher Anteil der Handlungskategorien (vgl. Tabelle 1) an der gesamten Beobachtungszeit ($n_{(2min, MINT)} = 1584$ und $n_{(2min, nMINT)} = 1096$).

Aus Abbildung 1 und 2 wird deutlich, dass fast alle Handlungskategorien des Studierenden- und Lehrendenverhalten in nMINT- und MINT-Veranstaltungen ähnlich häufig sowohl auf die Veranstaltungsanzahl als auch auf die zeitliche Präsenz bezogen dokumentiert wurden. Unterschiede gibt es in der Kategorie RtW – real-time writing, welche in 94 % aller beobachteten MINT-Veranstaltungen, aber nur 38 % aller nMINT-Veranstaltungen protokolliert wurde. Hierbei nimmt das Beschreiben von Tafel oder Folien auf dem Overheadprojektor bzw. elektronischen Geräten in den MINT-Disziplinen 21 % der gesamten Veranstaltungszeit ein, wobei in der Regel parallel verbal Inhalte vermittelt werden. In den nMINT-Lehrveranstaltungen sind es nur 3 % der Zeit. Das Studierendenhandeln Ind – individual thinking/problem solving wurde in 29 % der nMINT- und 66 % der MINT-Veranstaltungen beobachtet und nahm dabei in nMINT 2 % und in MINT 7 % der gesamten beobachteten Veranstaltungszeit ein. Weitere Unterschiede zwischen nMINT und MINT werden in der Häufigkeit der Kategorien Prd – making a prediction und T/Q – test or quiz sichtbar, allerdings ist der Anteil an der Veranstaltungszeit in diesen Kategorien bei beiden Disziplinen sehr gering.

Sowohl in den nMINT als auch in den MINT-Wissenschaften wurde die Kategorie O – Other (Sonstiges) selten (für nMINT $< 2\%$ bzw. für MINT $\leq 5\%$) im Rahmen der Protokollierung von den Beobachtenden gewählt (vgl. Abbildung 2). Die häufigsten Handlungen, die in dieser Kategorie dokumentiert wurden, sind das Einlegen einer geplanten Pause innerhalb der regulären Veranstaltungszeit, Clickertest bzw. Referate, in denen der Lehrende zuhört bzw. sich Notizen macht oder wartet, die Behebung technischer Probleme und das Tafelwischen (ausschließlich in MINT-Lehrveranstaltungen).

Für die Beurteilung der Reliabilität wurden zwei Größen herangezogen: der Jaccard-Koeffizient (T) und das Cohens Kappa (κ). Der Jaccard-Koeffizient dient als Maß für die Ähnlichkeit zweier Beobachtendenurteile. Hierbei wird gemessen, wie häufig die zwei Beobachtenden in ihrem Urteil übereinstimmen (Jaccard, 1901). In Anlehnung an Smith et al. (2013) wurde der Jaccard-Koeffizient für die einzelnen Kategorien bestimmt. Basis hierfür sind vier MINT- und drei nMINT-Lehrveranstaltungen, die von zwei oder drei Beobachtenden gleichzeitig protokolliert wurden. Die Ergebnisse sind in Tabelle 4 dargestellt. Insgesamt liegen mehr als 72 % aller Werte über $T = 0,9$ und der niedrigste Wert bei $0,74$ ($T_{nMINT} = 0,77$; $T_{MINT} = 0,71$), was auf eine relativ hohe Ähnlichkeit in den Beobachtungen der einzelnen Kategorien hinweist.

Tab. 4: Jaccard-Koeffizienten für die Wahl der Handlungskategorien (vgl. Tabelle 1) bei zwei oder drei Beobachtenden

		Studierendenhandeln												
		L	Ind	CG	WG	OG	AnQ	SQ	WC	Prd	SP	T/Q	W	O
T_{MINT}		0,71	0,86	0,96	0,93	0,92	0,72	0,93	0,85	0,96	1,00	1,00	0,98	0,94
T_{nMINT}		0,78	0,95	0,96	0,98	1,00	0,85	0,95	0,99	0,96	1,00	0,98	1,00	0,97
		Lehrendenhandeln												
		Lec	RtW	FUp	PQ	CQ	AnQ	MG	1o1	D/V	Adm	W	O	
T_{MINT}		0,72	0,77	0,87	0,74	0,91	0,84	0,94	0,95	0,89	0,98	0,90	0,86	
T_{nMINT}		0,81	0,97	0,88	0,77	0,93	0,93	0,96	0,93	0,96	0,91	0,99	0,99	

Zur Bestimmung des Interrater-Reliabilitätsmaßes wurde Cohens Kappa berechnet. Cohens Kappa liefert eine Aussage über die Güte von Übereinstimmungen der Einschätzung verschiedener Beobachtenden. Genauer gesagt, gibt dieser Wert Informationen darüber, wie sehr die vorliegenden Übereinstimmungen zweier Beobachtenden zufällig zustande gekommen sein könnten. Ein Kappa-Wert von $\kappa = 0$ bedeutet, dass die Übereinstimmungen zufällig zustande gekommen sind. Werte größer 0,60 gelten als Kennzeichen für eine hohe und größer 0,80 als sehr hohe Übereinstimmungsgüte (Landis & Koch, 1977). Da die Beobachtenden unterschiedlichen Fachdisziplinen entstammen und wir dadurch einen Einfluss auf die Beobachtung vermuten, wurden die Kappa-Werte über die sieben verschiedenen Lehrveranstaltungen, in denen zwei oder drei Beobachtende hospitierten, ermittelt. Es wurden alle unterschiedlichen Konstellationen von Beobachtenden hierbei berücksichtigt. In Tabelle 5 sind die Werte für zwei verschiedene Zusammensetzungen von Beobachtungspaaren aufgeführt. Hierbei setzt sich ein Paar aus der Gruppe MINT vs. MINT so zusammen, dass beide Beobachtenden MINT-Disziplinen angehören. Bei einem Paar aus der Gruppe nMINT vs. MINT gehört eine Person einer nMINT-Disziplin an.

Tab. 5: Cohens Kappa Werte in Abhängigkeit der verschiedenen Kombinationen von Beobachtenden.

Zusammensetzung der Beobachtenden	Beobachtungskombinationen	κ für alle Kategorien	κ für Studierenden-Kategorien	κ für Lehrenden-Kategorien
nMINT vs. MINT	7	0,42	0,46	0,37
MINT vs. MINT	4	0,69	0,71	0,67

Acht verschiedene Lehrveranstaltungen (vier nMINT und vier MINT) wurden im Vorfeld der Hospitationen von den Autoren und Autorinnen auf der Basis von Selbsteinschätzungen der Lehrenden und Beschreibungen der didaktischen Konzepte für die Lehrveranstaltungen in studierendenzentriert bzw. nicht studierendenzentriert eingeordnet. Veranstaltungskonzepte, die einen größeren Anteil an Kleingruppenarbeit (KGA) oder Peer-Instruction (PI) mit und ohne Classroom Response Systems (CRS) beinhalten, wurden als studierendenzentriert definiert. Konzepte, die in ihrer Beschreibung viele vortragende Elemente des Dozenten beinhalten, wurden als nicht studierendenzentriert angesehen. In Abbildung 3 sind die deutlichen Unterschiede in den Beobachtungsergebnissen dieser Lehrveranstaltungen anhand der in Tabelle 3 beschriebenen zusammengefassten Kategorien von Smith et al., 2014 zu erkennen.

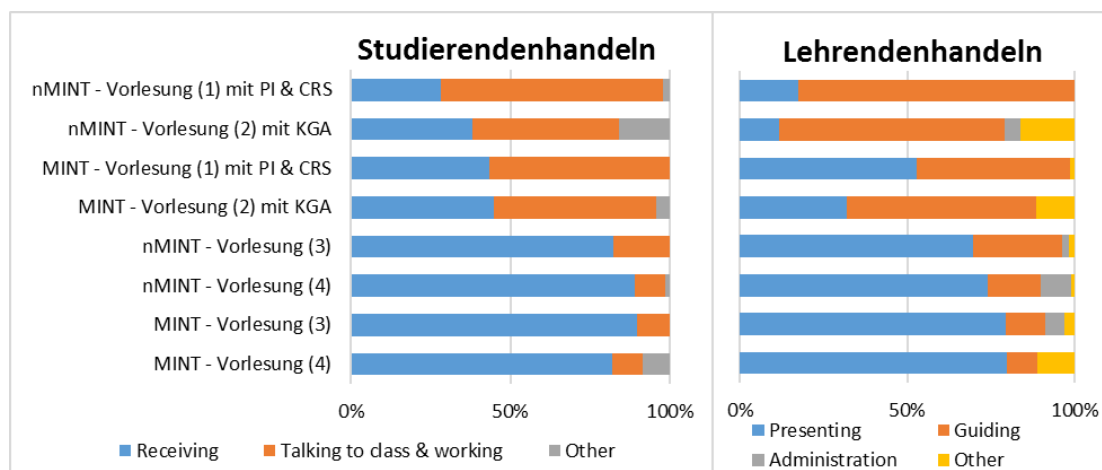


Abb. 3: Verteilung der zusammengefassten Kategorien in acht verschiedenen Lehrveranstaltungen mit unterschiedlichem Ausmaß an Studierendenzentrierung

4 Diskussion

Die Ergebnisse in Abbildung 1 und 2 zeigen nur geringe Unterschiede im Handeln der Studierenden und Lehrenden zwischen nMINT- und MINT-Lehrveranstaltungen. In einzelnen Kategorien wie beispielsweise RtW – real-time writing ist ein Unterschied jedoch auffällig und messbar und passt zur vorherrschenden Annahme, dass in den MINT-Wissenschaften die Kommunikation über die Tafel eine lange Tradition hat. Die Ergebnisse zeigen ebenfalls, dass die Studierendenkategorie Ind – individual thinking/problem solving in MINT-Lehrveranstaltungen mehr als doppelt so häufig wie in nMINT beobachtet wurde. Auch wenn der zeitliche Umfang dieser Handlung in beiden Disziplinen wiederum gering ist, ist die unterschiedliche Häufung auffällig. Dieser Unterschied begründet sich durch die in den MINT-Lehrveranstaltungen eher verbreitete Umsetzung der Methode Peer-Instruction, einhergehend mit der Nutzung von Classroom Response Systems (CRS) wie Clicker. Diese Methode sieht explizit eine Phase des individuellen Denkens der Studierenden vor (Mazur, 1997).

Alle nicht kodierbaren Handlungen sollten durch die Kategorie O – Other (Sonstiges) erfasst werden. Dies betrifft auch im Speziellen nMINT-spezifische Handlungen von Studierenden und Lehrenden in nMINT-Lehrveranstaltungen. Es konnten in den 25 Hospitationen von 11 nMINT-Lehrveranstaltungen mit einem zeitlichen Gesamtvolumen von 2192 Minuten keine solchen spezifischen Handlungen identifiziert werden. Da keine deutlichen Unterschiede zwischen den Handlungen in nMINT- sowie MINT-Lehrveranstaltungen nachgewiesen und keine nMINT-spezifischen Handlungskategorien identifiziert werden konnten, gehen wir davon aus, dass das COPUS in beiden Disziplingruppen anwendbar ist.

Die Analyse der Beobachtungsübereinstimmung in Tabelle 4 zeigt, dass zwei Beobachtenden, die eine nMINT-Lehrveranstaltung mit dem COPUS protokollieren, nur geringfügig bessere Reliabilitätswerte erreichen als bei der Protokollierung einer MINT-Lehrveranstaltung. Im Vergleich zu den von Smith et al. (2013) publizierten Jaccard-Koeffizienten konnten in mehr als der Hälfte aller Handlungskategorien ähnlich gute Wer-

te erreicht werden. Die davon abweichenden, schlechteren Werte in den Kategorien Ind, L, AnQ, WC, Lec, RtW, FUp und PQ könnten in der heterogenen Zusammensetzung bzgl. Beobachtungserfahrung und Fachdisziplin innerhalb der Gruppe der Beobachtenden begründet liegen. Hierzu wurden die Werte für Cohens Kappa als Interrater-Reliabilität zwischen Beobachtenden aus verschiedenen bzw. gleichen Fachdisziplinen ermittelt. Die Ergebnisse in Tabelle 5 zeigen, dass die Beobachtungsübereinstimmung zwischen zwei Beobachtenden aus MINT-Disziplinen (MINT vs. MINT) hierbei deutlich besser ist als zwischen zwei Beobachtenden aus verschiedenen Fachdisziplinen (nMINT vs. MINT). In keinem der analysierten Fälle konnten ähnlich gute Werte wie von Smith et al. (2014) erreicht werden, deren Beobachter immer aus MINT-Disziplinen stammten. Dies könnte unserer Meinung nach an unterschiedlichem Verständnis der Handlungskategorien der verschiedenen Beobachtenden, hervorgerufen durch die fachspezifischen Prägungen von Wahrnehmungsschemata für das Beobachten von menschlichen Handlungen, liegen. Dies müsste in weiteren Untersuchungen bestätigt werden.

Aufgrund dieser Erfahrungen ist eine umfangreiche Einarbeitung in die Verwendung des COPUS, insbesondere in die Deutung der Handlungskategorien, für das Erreichen einer hohen Interraterreliabilität notwendig ist. Die von Smith et al. (2014) angegebenen 1-2 Stunden sind nicht ausreichend. Zudem könnte sich eine prozessbegleitende Reflektion des Beobachtungs- und Kodierverhaltens als empfehlenswert erweisen. Auf Grundlage der ermittelten Reliabilitätswerte für Cohens Kappa und den Jaccard-Koeffizient und der Berücksichtigung der abgeleiteten Forderung einer prozessbegleitenden Schulung lässt sich sagen, dass die Beobachtungsdaten mit dem COPUS in MINT- und nMINT-Lehrveranstaltungen ausreichend reliabel sind. Die Einflüsse anderer Faktoren, wie Studierendekohorte, der Veranstaltungsraum mit seiner Bestuhlung oder die Hochschulart, auf die Gütekriterien muss weiter untersucht werden.

Die Beobachtungsdaten von Lehrveranstaltungen, die als studierendenzentriert bzw. nicht-studierendenzentriert definiert wurden, unterscheiden sich deutlich anhand der Kategorien *talking to the class & working, guiding* sowie *receiving* und *presenting* (vgl. Abbildung 3). In den Ergebnissen ist kein Unterschied zwischen MINT- und nMINT-Lehrveranstaltungen festzustellen. Auf Grundlage dieser Ergebnisse lässt sich ableiten, dass die Beobachtungsdaten, die mit dem COPUS ermittelt werden, eine Aussage darüber erlauben, in welchem quantitativen Ausmaß studierendenzentrierte Handlungen im Sinne von Johannes & Seidel (2012) von Studierenden und Lehrenden in Lehrveranstaltungen von MINT- und nMINT-Disziplinen stattfinden. Eine Klassifizierung bzgl. studierendenzentrierter und nicht studierendenzentrierter Lehre ausschließlich anhand dieser Beobachtungsdaten und ohne weitere Indikatoren und Definitionen ist nicht möglich (Smith et al., 2014).

Das COPUS liefert wichtige und im beschriebenen Einarbeitungsrahmen verlässliche Informationen für eine mögliche Entwicklung von MINT- und nMINT-Lehrveranstaltungen hin zu mehr Studierendenzentrierung. Damit könnte die Verwendung des COPUS, beispielsweise in Form einer Pre-Post-Erhebung, einen Beitrag dazu leisten, Informationen über die Wirksamkeit hochschuldidaktischer Maßnahmen abzuleiten. Die mit Hilfe des COPUS ermittelten Daten lassen eine quantitative Aussage über das Ausmaß von Studierendenzentrierung in Lehrveranstaltungen zu. Um Informationen über qualitative Aspek-

te der Lehrveranstaltungen zu erhalten, könnte das COPUS in Kombination mit weiteren Erhebungsinstrumenten eingesetzt werden, wie das Teaching Practic Inventory (Wieman & Gilbert, 2014), das Reformed Teaching Observation Protocol (Piburn & Sawada, 2000), das UTeach Observation Protocol (Walkington et al., 2012), das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (Braun, Gusy, Leidner & Hannover, 2008) oder das Approaches to Teaching Inventory (Trigwell & Prosser, 2004). Dann wäre zu erwarten, dass durch Kombination dieser mit dem COPUS auch didaktische Handlungsmuster, wie z.B. das Stellen einer herausfordernden Aufgabe und deren Auflösung im Plenum, in einer Veranstaltungshospitation mit Hilfe von Beobachtungsbögen dokumentiert und bewertet werden können. Die so erhaltenen quantitativen und qualitativen Aussagen zum Ausmaß der Studierendenzentrierung sollten dann in Beziehung mit der beabsichtigten Sinnhaftigkeit bzw. der Zielsetzung der einzelnen Lehrveranstaltung unter Berücksichtigung der verschiedenen Rahmenbedingungen gesetzt werden.

Literatur

- Beichner, R. J., Saul, J. M., Abbott, D. S., Morse, J. J., Deardorff, D., Allain, R. J., Bonham, S. W., Dancy, M. H. & Risley, J. S. (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. *Research-based reform of university physics*, 1(1), 2-39.
- Berendt, B. (1998). How to support and to bring about the shift from teaching to learning through academic staff development programmes: Examples and perspectives. *Higher Education in Europe*, 23(3), 317-329.
- Berendt, B., Voss, H. P. & Wildt, J. (2002). *Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten* [Loseblattsammlung]. Stuttgart: Raabe.
- Braun, E. & Hannover, B. (2011). Gelegenheiten zum Kompetenzerwerb in der universitären Lehre. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43(1), 22-28.
- Braun, E., Gusy, B., Leidner, B. & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, 54(1), 30-42.
- Bundesministeriums für Bildung und Forschung (06.11.2015). *Für bessere Studienbedingungen: Zweite Förderperiode im Qualitätspakt Lehre*. Pressemitteilung, 151/2015.
- Deutsche Gesellschaft für Hochschuldidaktik (2013). *Qualitätsstandards für die Anerkennung von Leistungen in der hochschuldidaktischen Weiterbildung*. Abgerufen von http://www.dghd.de/wp-content/uploads/2015/05/Qualit%C3%A4tsstandards_Hochschuldidaktik_11.11.2013b.pdf
- Gauci, S. A., Dantas, A. M., Williams, D. A. & Kemm, R. E. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, 33(1), 60-71.
- Havers, N. & Toepell, S. (2002). Trainingsverfahren für die Lehrerausbildung im deutschen Sprachraum. *Zeitschrift für Pädagogik*, 48(2), 174-193.
- Hora, M. T. & Ferrare, J. J. (2013). *A Review of Classroom Observation Techniques in Post-Secondary Settings*. Madison: Wiscincosin Center for Education Research.

- Hora, M. T. & Ferrare, J. J. (2014). Remeasuring postsecondary teaching: How singular categories of instruction obscure the multiple dimensions of classroom practice. *Journal of College Science Teaching*, 43(3), 36-41.
- Hora, M. T., Oleson, A. & Ferrare, J. J. (2013). *Teaching dimensions observation protocol (TDOP) user's manual*. Madison: Wisconsin Center for Education Research.
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, 37, 547-579.
- Johannes, C. & Seidel, T. (2012). Professionalisierung von Hochschullehrenden. *Zeitschrift für Erziehungswissenschaft*, 15(2), 233-251.
- Kromrey, H. (2009). *Empirische Sozialforschung: Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung*. (12., überarb. und erg. Aufl.). Stuttgart: Lucius & Lucius.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.
- Lueddeke, G. R. (2003). Professionalising teaching practice in higher education: A study of disciplinary variation and 'teaching-scholarship'. *Studies in higher education*, 28(2), 213-228.
- Mazur, E. (1997). *Peer Instruction: A User's Manual*, Prentice Hall Series in Educational Innovation.
- Óhidy, A., Terhart, E. & Zsolnai, J. (2007). *Lehrerbild und Lehrerbildung. Praxis und Perspektive der Lehrerausbildung in Deutschland und Ungarn*. Wiesbaden: VS.
- Piburn, M. & Sawada, D. (2000). *Reformed Teaching Observation Protocol (RTOP) Reference Manual*. Technical Report.
- Seidel, T. & Hoppert, A. (2011). Merkmale von Lehre an der Hochschule: Ergebnisse zur Gestaltung von Hochschulseminaren mittels Videoanalysen. *Unterrichtswissenschaft*, 39(2), 154-172.
- Smith, M. K., Jones, F. H., Gilbert, S. L. & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices. *CBE-Life Sciences Education*, 12(4), 618-627.
- Smith, M. K., Vinson, E. L., Smith, J. A., Lewin, J. D. & Stetzer, M. R. (2014). A campus-wide study of STEM courses: new perspectives on teaching practices and perceptions. *CBE-Life Sciences Education*, 13(4), 624-635.
- Trigwell, K. & Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review*, 16(4), 409-424.
- Turpen, C. & Finkelstein, N. D. (2009). Not all interactive engagement is the same: variations in physics professors' implementation of peer instruction. *Physical Review Special Topics-Physics Education Research*, 5(2), 020101.
- Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L. & Marder, M. (2012). *Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation program*. Austin: University of Texas.
- Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco: Jossey-Bass.

- Wibbecke, G., Wibbecke, A. L., Kahmann, J. & Kadmon, M. (2016). Lehrenden- und studienzentrierte Lehre messen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48(4), 184-194.
- Wieman, C. & Gilbert, S. (2014). The Teaching Practices Inventory: a new tool for characterizing college and university teaching in mathematics and science. *CBE-Life Sciences Education*, 13(3), 552-569.
- Wildt, J. (2001). Ein hochschuldidaktischer Blick auf Lehren und Lernen in gestuften Studiengängen. In U. Welbers (Hrsg.): *Studienreform mit Bachelor und Master. Gestufte Studiengänge im Blick des Lehrens und Lernens an Hochschulen. Modelle für die Geistes- und Sozialwissenschaften*. Neuwied, Kriftel: Luchterhand, 25-42.

Autor/-innen

Sebastian Wirthgen M. Ed. Ostfalia Hochschule für angewandte Wissenschaften, ZeLL - Zentrum für erfolgreiches Lehren und Lernen, Wolfenbüttel, Deutschland; Email: s.wirthgen@ostfalia.de

Dipl.-Päd. Kathrin Munt. Ostfalia Hochschule für angewandte Wissenschaften, ZeLL - Zentrum für erfolgreiches Lehren und Lernen, Wolfenbüttel, Deutschland; Email: k.munt@ostfalia.de

Dr. rer. nat. Frauke Stenzel. Ostfalia Hochschule für angewandte Wissenschaften, ZeLL - Zentrum für erfolgreiches Lehren und Lernen, Wolfenbüttel, Deutschland; Email: f.stenzel@ostfalia.de

Dipl.-Math. Kathrin Gläser. Ostfalia Hochschule für angewandte Wissenschaften, ZeLL - Zentrum für erfolgreiches Lehren und Lernen, Wolfenbüttel, Deutschland; Email: k.glaeser@ostfalia.de

Dipl.-Math. Anika Fricke. Ostfalia Hochschule für angewandte Wissenschaften, ZeLL - Zentrum für erfolgreiches Lehren und Lernen, Wolfenbüttel, Deutschland; Email: anika.fricke@ostfalia.de



Zitiervorschlag: Wirthgen, S., Munt, K., Stenzel, F., Gläser, K. & Fricke, A. (2018). Erfahrungen mit dem Classroom Observation Protocol for Undergraduate STEM (COPUS) in MINT und nicht-MINT Hochschulkursen. *die hochschullehre*, Jahrgang 4/2018, online unter: www.hochschullehre.org