In: M. Klinger, A. Schüler-Meyer, L. Wessel (Hrsg.), Hanse-Kolloquium zur Hochschuldidaktik der Mathematik 2018 (pp. 63-74). Münster: WTM, 2019.

Design von Aufgaben für Peer Instruction zum Einsatz in Übungsgruppen zur Analysis

Bauer, Thomas[1]

[1]Philipps-Universität Marburg

Zusammenfassung:

Die Methode der Peer Instruction bietet einen Ansatz, um Studierende in mathematischen Übungsgruppen fokussiert zu aktivieren. Hierdurch kann der Gefahr begegnet werden, dass das „Vorrechnen" von Übungsaufgaben durch den Tutor der Hauptbestandteil der Übungen wird. Dreh- und Angelpunkt der Methode sind Aufgaben, die so beschaffen sind, dass sie die angestrebten fachbezogenen fokussierten Argumentationsprozesse bei den Studierenden anregen. Im Beitrag werden mehrere Iterationen von Design-Zyklen eines Projekts betrachtet, in dem solche Aufgaben konstruiert und eingesetzt wurden. Es wird gezeigt, dass das Projekt als Ergebnis einerseits Aufgaben für den wöchentlichen Einsatz in Übungsgruppen zur Analysis 1 und 2 (als praktischen Ertrag) erbrachte und es andererseits die Entwicklung von Design-Prinzipien für Peer-Instruction-Aufgaben (als theoretischen Ertrag) ermöglichte.

Problemlage

Da dem Üben im Mathematikstudium hohe Bedeutung zugemessen wird, sind auch die Ansprüche, die an mathematische Übungsgruppen gestellt werden, hoch: So sollen sie u.a. Gelegenheit bieten, Fragen zu klären, über aktuelle Stoffinhalte zu diskutieren und Verständnisprobleme zu bearbeiten. Diesem Anspruch wird die reale Umsetzung allerdings nicht immer gerecht: Zum einen sind Studierende oft zurückhaltend beim Stellen von Fragen, wenn es ihnen schwer fällt, Verständnisprobleme zu formulieren, oder sie sich scheuen, diese vor der Gruppe offenzulegen. Zum anderen haben Tutoren (Übungsleiter) oft Schwierigkeiten damit, Diskussionen durch geeignete Impulse in Gang zu bringen. In der Realität wird dadurch häufig das Vorstellen von Musterlösungen (im Jargon gerne „Vorrechnen" genannt) ein Hauptbestandteil von Übungen. Prinzipiell kann es effektiv sein, wenn ein Experte für Laien Problemlösestrategien expliziert (Ableitinger und Herrmann 2011, Ableitinger 2013). Weniger Wirkung kann man jedoch erwarten, wenn dabei nicht der Lösungsprozess, sondern lediglich die Lösungen (als dessen Endprodukt) präsentiert werden.

Insgesamt besteht somit die Gefahr, dass die Studierenden in mathematischen Übungsgruppen in einer zu passiven Rolle verbleiben. Die (meist studentischen) Tutoren stehen hier vor einer anspruchsvollen Aufgabe (vgl. Biehler et al. 2012). Durch Tutorenschulungen können sie zwar in methodischer Hinsicht in gewissem Umfang trainiert werden, aber die Möglichkeiten, sie darüber hinaus auch stoffbezogen mathematikdidaktisch vorzubereiten, sind begrenzt. Diese Problemlage lässt den Wunsch nach effektiven Möglichkeiten zur Aktivierung entstehen, die methodisch einfach zu handhaben sind und daher von studentischen Tutoren gut durchgeführt werden können.

In Bauer (2018) wurde gezeigt, dass die Methode der Peer Instruction einen vielversprechenden Ansatz bietet, um diesem Problem entgegenzutreten. Während dort primär der Frage nachgegangen wurde, wie Peer Instruction zur Aktivierung genutzt werden kann und wie Studierende und Tutoren die Methode aufnehmen, soll nun im vorliegenden Beitrag der *Designprozess* für Aufgaben zur Peer Instruction in Übungsgruppen zur Vorlesung Analysis genauer in den Blick genommen werden. Dazu werden die bisherigen Iterationen der Design-Zyklen eines Entwicklungsprojekts vorgestellt, um zu zeigen, dass diese sowohl einen praktischen Ertrag (in Form einer verwendbaren Intervention) als auch einen theoretischen Ertrag (in Bezug auf Design-Prinzipien für Peer-Instruction-Aufgaben) erbracht haben.

Hintergrund und Fragestellung

Peer Instruction

Eric Mazur hat Ende der 90er Jahre im Fach Physik die Methode der Peer Instruction entwickelt (Mazur 1997, Mazur 2017). Ihre Absicht ist es, die Studierenden dadurch zu aktivieren, dass sie sich, angeregt durch vom Dozenten vorgelegte Fragen, gegenseitig Erklärungen für physikalische Sachverhalte geben. Die Fragen (bei Mazur *ConcepTest* genannt) betreffen dabei nicht den mathematisch-physikalischen Kalkül, sondern sind konzeptueller Natur, d.h. sie sind auf physikalische Konzepte bezogen, also etwa auf Vorstellungen zu Begriffen wie Geschwindigkeit, Beschleunigung oder Kraft. Die Methode – bei Mazur im Rahmen eines Inverted-Classroom-Konzepts eingesetzt – besteht in praktischer Hinsicht darin, dass eine oder mehrere Runden mit folgendem Ablauf durchgeführt werden:

Phase 1: ConcepTest. Die Studierenden erhalten eine Aufgabe mit mehreren Antwortmöglichkeiten und überlegen in Einzelarbeit, welche der gegebenen Antwortalternativen richtig ist. Sie suchen Argumente dafür und Argumente gegen die anderen Alternativen. Am Ende dieser (wenige Minuten dauernden) Phase erfolgt eine Abstimmung, mit der sich jeder Teilnehmer auf eine der Antwortalternativen festlegt.

Phase 2: Peer Discussion. Unter dem Motto „Überzeuge Deinen Nachbarn" erhalten die Studierenden in Gruppen von 2–3 Teilnehmern die Aufforderung, sich gegenseitig von der Richtigkeit ihrer gewählten Antwortalternative zu überzeugen. Hierfür werden die Gruppen nach Möglichkeit aus Teilnehmern zusammengesetzt, die verschiedene Antwortalternativen gewählt haben. Am Ende dieser Phase erfolgt eine zweite Abstimmung, in der die Teilnehmer auf Basis der vorangegangenen Diskussion eine (ggf. neue) Entscheidung für eine Antwortalternative treffen.

Phase 3 (optional): Die Lehrperson kann abschließend offene Fragen klären oder weitergehende Erläuterungen zu den Antwortalternativen anbieten.

Die produktive Wirkung, die man von Peer Instruction erwartet, liegt in Phase 2 (Peer Discussion). Wenn es gelingt, dass durch die Aufgaben dort Argumentationsprozesse angestoßen werden, bei denen die Teilnehmer zum Kern der fokussierten Konzepte vordringen, dann können hierdurch Fehlvorstellungen aufgedeckt werden und das Verständnis der Konzepte vertieft werden. Der Aufgabenkonstruktion kommt bei Peer Instruction demnach die entscheidende Bedeutung zu – es ist keineswegs klar, wie Aufgaben konkret beschaffen sein müssen, um die erwünschten Wirkungen zu erzielen. Crouch et al. (2007) geben, auf das Fach Physik bezogen, einige grundlegende Kriterien an. Bewusst sprechen sie davon, dass es keine „hard-and-fast rules" gebe (a.a.O., S. 9). Die „basic criteria", zu denen sie erfahrungsbasiert gelangt sind, lassen sich folgendermaßen zusammenfassen:

(M1) Ein ConcepTest bezieht sich auf ein einzelnes Konzept (idealerweise auf eines, mit dem Studierende erfahrungsgemäß Schwierigkeiten haben).

(M2) Ein ConcepTest erfordert Nachdenken, nicht nur Einsetzen von Zahlen in Formeln.

(M3) Ein ConcepTest bietet plausible inkorrekte Antworten an. Idealerweise sind diese aus häufig beobachteten (oder wahrscheinlichen) Fehlern von Studierenden gebildet.

(M4) ConcepTests sind möglichst unmissverständlich formuliert. (Dies erfordert in der Regel ein Ausprobieren.)

(M5) ConcepTests sind weder zu leicht noch zu schwer: Günstig ist es, wenn beim ersten Voting etwa 35–70 Prozent der Teilnehmer die richtige Antwort geben. Sind es weniger, dann verstehen zu wenige Teilnehmer das Konzept, als dass eine Diskussion Nutzen bringen würde. Sind es mehr, dann besteht zu wenig Diskussionsbedarf.

Während Peer Instruction im Fach Physik recht verbreitet ist und dort Erfolge beim Einsatz berichtet werden (Crouch und Mazur 2001), wird die Methode im Fach Mathematik noch weit weniger verwendet (siehe z.B. Miller et al. 2006 zum Einsatz in einem Calculus-Kurs in den USA). In Bauer (2018) wurde erstmals über einen Einsatz in mathematischen Übungsgruppen berichtet.

Fragestellung

Das hier dargestellte Peer-Instruction-Projekt hat seinen Schwerpunkt in der Entwicklungsarbeit, orientiert sich aber insofern an der Idee des Design Research, als es auf einen zweifachen Ertrag ausgerichtet ist (*twofold yield* bei Plomb 2013), bei dem auf der einen Seite als praktischer Ertrag eine einsatzfähige Intervention für eine komplexe unterrichtsbezogene Fragestellung erstellt wird und auf der anderen Seite als theoretischer Ertrag (wiederverwendbare) Design-Prinzipien erarbeitet werden sollen (a.a.O., S. 22, Van den Akker 1999, S. 5). Davon ausgehend lässt sich das Ziel des Projekts in folgender Weise spezifizieren:

- *Praktischer Ertrag:* Als Ergebnis des Projekts soll ein praktisch einsetzbares vollständig ausgearbeitetes Konzept für den Einsatz von Peer Instruction im mathematischen Übungsbetrieb von Grundvorlesungen zur Analysis bereitstehen. Dazu gehört, dass für jeden Themenbereich der Analysis 1 und 2 Aufgaben (ConcepTests) entwickelt werden, die im wöchentlichen Turnus in Übungsgruppen eingesetzt werden können. Durch das Projekt ist insbesondere zu klären, ob dies für alle Themenbereiche gleichermaßen gelingt.

- *Theoretischer Ertrag:* Es sollen Design-Prinzipien für die Erstellung von solchen Aufgaben entwickelt werden. Genauer geht es bei diesem Ziel darum, zu den Anforderungen an ConcepTests, die Crouch et al. (2007) in allgemeiner Form angegeben haben, konkretisierende Interpretationen für das Fach Mathematik zu finden und dabei zu spezifizieren, wie diese Anforderungen umzusetzen sind.

Design-Iterationen

Die bisherige Arbeit lässt sich, angelehnt an das Vorgehen im Design Research, als ein Prozess in mehreren Iterationen beschrieben, bei dem in jeder Iteration ein Zyklus

Analyse ➔ Design ➔ Test ➔ Evaluation

durchlaufen wird (vgl. Plomb 2013, Amiel und Reeves 2008). Die Designphasen dienten nach der ersten Iteration der Verfeinerung und waren geleitet durch die Analyse der Ergebnisse der vorhergehenden Evaluation. In den Analyse-Phasen wurden auch die Design-Prinzipien weiterentwickelt. Die im Projekt bislang durchgeführten Iterationen lassen sich im Überblick auf folgende Weise darstellen:

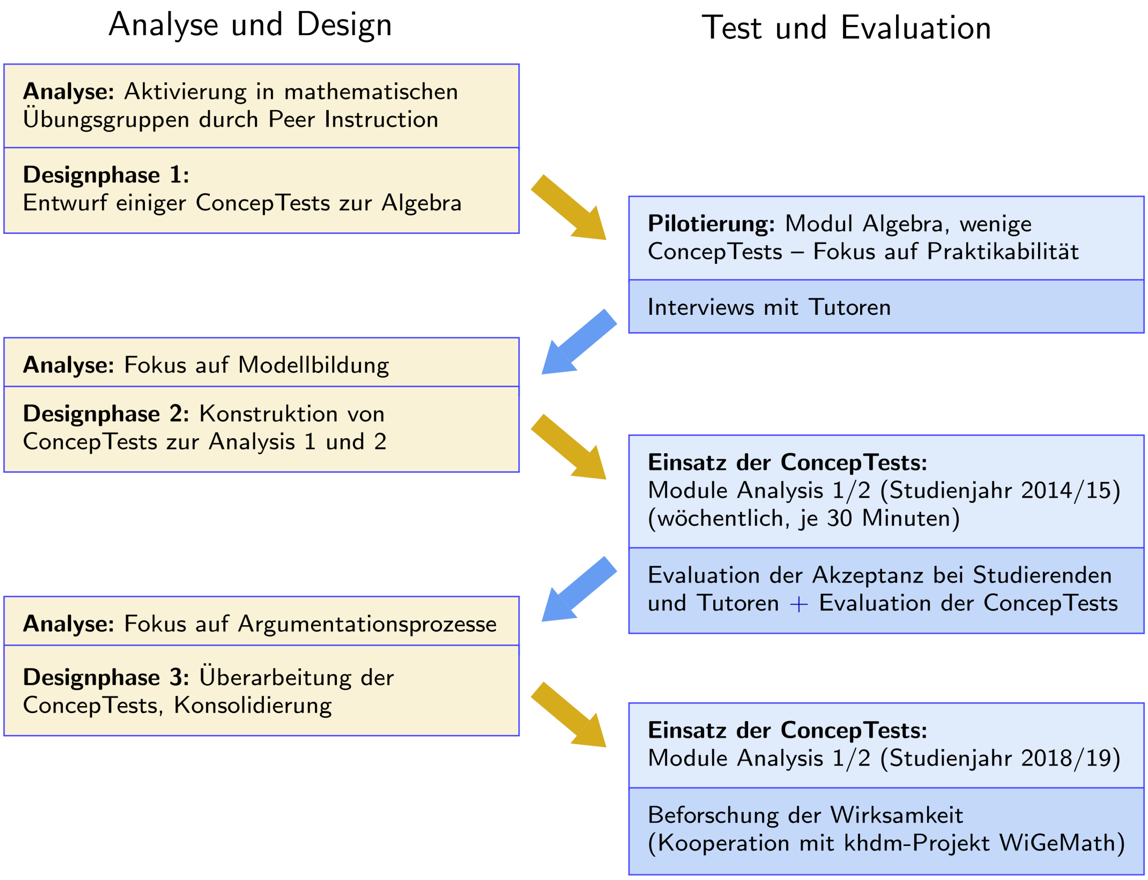


Abbildung 1: Iterationen der Designzyklen

Wir gehen nun auf die einzelnen Iterationen detaillierter ein und erläutern jeweils, welche Ergebnisse diese zu den im vorigen Abschnitt beschriebenen Zielen (im Sinne eines zweifachen Ertrags) erbracht haben.

Iteration 1. Erprobung – Fokus auf Praktikabilität

Die erste Iteration diente einer ersten Erprobung der Praktikabilität der Methode. Es sollte ermittelt werden, ob sie prinzipiell für den Einsatz in mathematischen Übungsgruppen geeignet erscheint, sowohl hinsichtlich der Konstruktion von Aufgaben als auch der Durchführung durch studentische Tutoren. Zudem sollten erste Einschätzungen zur Akzeptanz und auch zur technischen Durchführung (Präsentation der ConcepTests, Abstimmungsmethoden) gewonnen werden.

Analyse und Design: Orientiert an Beispielen von Mazur und Beispielen aus dem Good-Questions-Projekt (Miller et al. 2006) wurden einige ConcepTests konstruiert. Diese waren auf das Gebiet Algebra bezogen, da der Verfasser zu diesem Zeitpunkt eine Algebra-Vorlesung hielt und sie daher in den zugehörigen Übungen unproblematisch erproben konnte. Die Konstruktion der ConcepTests erfolgte in dieser Iteration auf Basis einer intuitiven (noch nicht weiter explizierten) Interpretation der oben formulierten Kriterien nach Crouch et al. (a.a.O.).

In dieser Iteration stellte sich auch die Design-Frage, wie viele Antwortmöglichkeiten bei einem ConcepTest jeweils angeboten werden sollen. Ergebnisse der Testtheorie zeigen, dass zu wenige Antwortmöglichkeiten (z.B. nur „Ja“ oder „Nein“) eher zum Raten ermuntern, während zu viele Antwortmöglichkeiten die Schwierigkeit eines Items erhöhen, ohne dessen Diskriminationsfähigkeit weiter zu verbessern – so empfiehlt etwa Rodriguez (2005) auf Basis einer Meta-Studie, *drei* Antwortalternativen zu verwenden. Für die Erprobung in dieser Iteration wurden daher ConcepTests mit 3 oder 4 Alternativen konstruiert.

Test und Evaluation: In einer Vorlesung zur Algebra an der Universität Marburg im Wintersemester 2013/14 wurden in drei Übungswochen in jeweils drei (von studentischen Tutoren geleiteten) Übungsgruppen ConcepTests eingesetzt. Zur Evaluation wurden anschließend kurze Interviews mit den Tutoren geführt. Hierdurch zeigte sich, dass die Tutoren der Methode prinzipiell positiv gegenüberstanden; sie berichteten ferner von positiver Akzeptanz bei den Studierenden ihrer Übungsgruppe. Die praktische Durchführung wurde als weitgehend unproblematisch geschildert; allerdings beurteilten die Tutoren, die die ConcepTests per Beamer präsentieren konnten, den Einsatz als einfacher im Vergleich zu denjenigen, die dies (aus Gründen der Raumausstattung) per Tafel tun mussten. Die Abstimmungen wurden mit Abstimmungskarten durchgeführt, was sich bei der überschaubaren Größe der Gruppen als gut praktikabel erwies.

Ergebnis: Insgesamt deuteten die Evaluationsergebnisse darauf hin, dass ein durchgängiger Einsatz der Methode prinzipiell durchführbar ist und von den Tutoren und nach deren Einschätzung auch von den Studierenden positiv aufgenommen würde.

Iteration 2. Durchgängiger wöchentlicher Einsatz – Fokus auf Modellbildung

Die zweite Iteration diente der Konstruktion von ConcepTests und ihrer Erprobung in einem durchgängigen Einsatz im Rahmen der Analysis 1 und 2. Dabei sollten die in der ersten Iteration noch auf intuitiver Basis angewandten grundlegenden Kriterien nun für das Fach Mathematik interpretiert und konkretisiert werden.

Analyse und Design: Um eine mathematikbezogene Interpretation der Kriterien (M1) und (M2) von Crouch et al. entwickeln zu können, ist es entscheidend, wie die Anforderung nach „konzeptuellen" Fragen, die darin zum Ausdruck kommt, einzulösen ist. Hierfür wurde die Idee von Concept Definition und Concept Image nach Tall und Vinner (1981) genutzt und zu einem dreistufigen Modell ausgearbeitet, das die verschiedenen Wissenselemente im Umgang mit einem Begriff hierarchisch gliedert. Da geplant war, in ConcepTests neben *Begriffen* auch *Sätze* zu thematisieren, wurden in das Modell auch Wissenselemente aufgenommen, die sich auf mathematische Sätze beziehen – diese wurden unter „Theorem Statement" und „Theorem Image" gefasst. Insgesamt lässt sich das Modell folgendermaßen darstellen (siehe Bauer 2018):

Tabelle 1: Dreistufiges Modell von Wissenselementen beim Begriffs- und Satzlernen

<table>
<tr><td>Stufe 1</td><td>Concept Definition</td><td>eine Definition des Begriffs präzise formulieren können, den Begriff gegen andere abgrenzen können</td><td>Theorem Statement</td><td>eine präzise Satzformulierung angeben können (Voraussetzungen und Konklusion herausstellen können)</td></tr>
<tr><td>Stufe 2</td><td rowspan="2">Concept Image</td><td>Sinn und Bedeutung verstehen: Begriffsinhalt, Begriffsumfang und das Begriffsnetz (Bezug zu anderen Begriffen) kennen</td><td rowspan="2">Theorem Image</td><td>Sinn und Bedeutung verstehen: die Satzaussage inhaltlich erklären können (auch graphisch oder mit Diagrammen), alternative Formulierungen angeben können</td></tr>
<tr><td>Stufe 3</td><td>den Begriff verwenden können: innerhalb des Theoriezusammenhangs, in Standardbeispielen und -anwendungen</td><td>den Satz verwenden können: innerhalb des Theoriezusammenhangs, in Standardbeispielen und -anwendungen</td></tr>
</table>

Auf Basis dieses Modells wurden ConcepTests für alle Themenbereiche der Analysis 1 und 2 semesterbegleitend erarbeitet. Dabei wurden einheitlich jeweils vier Antwortalternativen pro ConcepTest entworfen, aus denen im Single-Choice-Verfahren eine auszuwählen war. Diese Entwurfsentscheidung sollte auch die technische Umsetzung vereinheitlichen. Wir illustrieren an einem Beispiel, wie das Modell bei der Konstruktion von ConcepTests eingesetzt wurde – es konkretisiert dabei die Forderung nach konzeptuellen Fragen (M1, M2) für das Lernen mathematischer Begriffe und Sätze zu dem Design-Prinzip „Erreiche im Modell eine möglichst hohe Stufe".

Von einer differenzierbaren Funktion $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ sei bekannt, dass

$$f(0,0) = 0 \quad \text{und} \quad \operatorname{grad} f(0,0) \neq (0,0)$$

gilt. Dann muss es in jeder Umgebung von $(0,0)$

(1) Punkte geben, in denen f positiven Funktionswert hat
(2) Punkte geben, in denen f negativen Funktionswert hat
(3) beide Arten von Punkten geben
(4) Keines von diesen kann man folgern.

Hier soll einerseits das Theorem Image zum notwendigen Kriterium für Extrema („In einem lokalen Extremum verschwindet der Gradient.“) bearbeitet werden, gleichzeitig aber auch das Concept Image zu den Begriffen „Maximum“ und „Minimum“. Konkret müssen die Angaben zu $f(0,0)$, zur Positivität bzw. Negativität von Funktionswerten und zum Gradienten im Punkt (0,0) zueinander in Beziehung gesetzt werden. Der Schlüssel dafür liegt darin, die Begriffe „Maximum“ und „Minimum“ zur Formulierung von gegebenen Informationen zu verwenden (Concept Image, Stufe 3) und zu erkennen, dass in dieser Situation das notwendige Kriterium für lokale Extrema eingesetzt werden kann (Theorem Image, Stufe 3).

Um (M4) zu berücksichtigen, wurden die ConcepTests durch einen Mitarbeiter pilotiert und vor dem Einsatz bei Studierenden ggf. angepasst.

Test: Die entwickelten ConcepTests wurden im Sommersemester 2014 in der Analysis 1 (15 Übungsgruppen, 324 Teilnehmer) und im Wintersemester 2014/15 in der Analysis 2 (12 Übungsgruppen, 223 Teilnehmer) durchgängig eingesetzt. In jeder Woche wurden 30 Minuten zu Beginn jeder Übungssitzung für Peer Instruction reserviert. In dieser Zeit wurden 3 bis 5 Peer-Instruction-Runden durchgeführt (d.h. 3–5 ConcepTests verwendet). Die Entscheidung über die gewählte Anzahl lag im Ermessen der Tutoren, die auf diese Weise das Durchführungstempo und den Umfang zusätzlicher Erläuterungen in Phase 3 ihrer Lerngruppe anpassen konnten. Die Abstimmungen erfolgte überwiegend mittels Abstimmungskarten; probehalber wurden in einzelnen Sitzungen auch elektronische (internet-basierte) Abstimmungssysteme verwendet.

Evaluation: Mit Hilfe eines Rückmeldebogens erfassten die Tutoren zu jedem ConcepTest, ob er den von Crouch et al. empfohlenen Schwierigkeitsgrad hatte (ca. 35-70 Prozent richtige Antworten in der ersten Abstimmung) oder aber für die jeweilige Gruppe zu leicht oder zu schwer war. Auf diese Weise sollte ermittelt werden, in welchem Maße Kriterium (M5) bei den konstruierten ConcepTests erfüllt ist. Ferner erfassten die Tutoren zu jedem ConcepTest, als wie produktiv sie die Gruppendiskussionen zum jeweiligen ConcepTest wahrnahmen und welche weiteren Beobachtungen sie beim Einsatz des ConcepTests machten. Hierdurch sollten erste Erkenntnisse in Bezug auf (M1) und (M2) gewonnen werden.

Ergebnis:

Die Ergebnisse der Evaluation zeigen, dass die ConcepTests weitgehend im durch (M5) gegebenen Schwierigkeitsintervall von 35 bis 70 Prozent lagen, es aber oft einzelne Übungsgruppen gab, in denen ein ConcepTest deutlich außerhalb des Intervalls lag. Es gab darüber hinaus auch Fälle, in denen derselbe ConcepTest in einer Gruppe unter der 35-Prozent-Grenze und in einer anderen Gruppe über der 70-Prozent-Grenze lag.

Die Rückmeldungen der Tutoren zur Produktivität der ConcepTests erbrachten drei Ergebnisse: (1) Die ConcepTests haben in vielen Fällen dazu geführt, dass die verhandelten Begriffe und Sätze in erwünschter Tiefe diskutiert wurden und Fehlvorstellungen bearbeitet werden konnten. (2) In einzelnen Fällen kam es vor, dass sich in den Peer Discussions Gruppen zwar auf die richtige Antwortoption einigten, aber „aus den falschen Gründen“. In solchen Fällen wirkt Peer Instruction aus Sicht der Abstimmungsergebnisse effektiv, hat aber in Wirklichkeit negative Effekte. (3) Manche ConcepTests, deren Schwierigkeit außerhalb des durch (M5) gegebenen Intervalls lag, wurden von den Tutoren dennoch als sehr produktiv eingeschätzt: Einerseits entstanden auch aus manchen „zu schweren“ ConcepTests sehr fruchtbare Diskussionen (dann aber mit einer abschließenden Phase 3, die vom Tutor geleitet wurde), und andererseits wurden auch manche „zu leichte“ ConcepTests als motivational hilfreich betrachtet.

Als praktischer Ertrag dieser Iteration wurden für alle Themenbereiche der Analysis 1 und 2 ConcepTests erarbeitet, die wöchentlich eingesetzt werden können. Es gelang für alle Themenbereiche, die einschlägigen Begriffe und Sätze im gegebenen Aufgabenformat zu erfassen.

Als theoretischer Ertrag steht nun ein Modell zur Verfügung, das sich bei der Konstruktion von ConcepTests sowohl in Bezug auf das Lernen mathematischer Begriffe als auch auf das Lernen mathematischer Sätze praktisch bewährt hat. Die Beobachtung, dass in den Peer Discussions gelegentlich die richtige Antwort „aus den falschen Gründen“ gewählt wurde, zeigte, dass in dieser Hinsicht noch eine Verfeinerung des Aufgabendesigns erforderlich ist. Die berichteten Ergebnisse zu (M5) zeigen, dass in Bezug auf die Schwierigkeit der ConcepTests weitere Untersuchungen erforderlich sind.

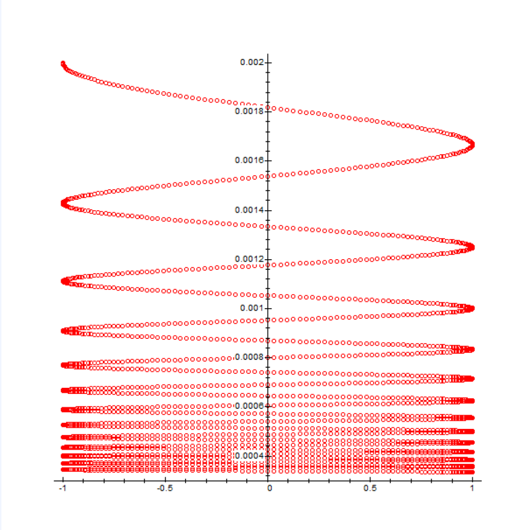
Iteration 3. Zweiter durchgängiger wöchentlicher Einsatz – Fokus auf Argumentationsprozesse

Ziel dieser Iteration ist eine Konsolidierung der bislang entworfenen Intervention und eine Verfeinerung im Hinblick auf die in der vorigen Iteration gewonnenen Ergebnisse.

Analyse und Design: Als Konsequenz von Ergebnissen der vorigen Iteration lag in dieser Iteration der genauere Fokus auf dem Argumentationsprozess, der durch einen ConcepTest angestoßen werden soll. Um die Argumentationsprozesse stärker zu steuern, wurde in den als kritisch erkannten Fällen das Design der ConcepTests dahingehend überarbeitet, dass schon durch ihre Formulierung das Abwägen von korrekten und inkorrekten Argumenten gefordert wurde. Die folgenden zwei Beispiele illustrieren dies. Im ersten Beispiel wird durch die Aufgabenstellung die Peer Discussion direkt auf das Abwägen zwischen vorgegebenen falschen und richtigen Argumentationen gelenkt. Die falschen Argumentationen sind dabei, (M3) folgend, aus bekannten Fehlschlüssen gewonnen.

Das folgende Bild zeigt die Folge $(a_n) = (\cos(\frac{n\pi}{100}), \frac{1}{n})$ in $\mathbb{R}^2$.

Sie ist



(1) konvergent, da sie beschränkt und monoton fallend ist
(2) konvergent, da die Folgenglieder a_n beliebig nahe an der x-Achse liegen, wenn n hinreichend groß ist
(3) divergent, da $(\cos(\frac{n\pi}{100}))$ divergent ist
(4) divergent, da $a_n \neq (0, 0)$ für alle $n \in \mathbb{N}$ gilt

Das folgende Beispiel geht insofern noch einen Schritt weiter, als der Argumentationsprozess hier nicht durch ein „Ausschlussverfahren" abgekürzt werden kann: Das Single-Choice-Format mit 4 Antwortmöglichkeiten wird hier letztlich zur Realisierung eines Multiple-Choice-Formats mit 2 Antwortmöglichkeiten verwendet: Bei jeder der Optionen (1) und (2) ist zu entscheiden, ob das angegebene Argument korrekt ist.

Wir betrachten die Funktion

$$f : [0, 1] \times [0, 1] \longrightarrow \mathbb{R}$$
$$(x, y) \mapsto \begin{cases} 1, & \text{falls } x > y \\ 2, & \text{sonst} \end{cases}$$

(1) f ist integrierbar, weil f stetig ist bis auf eine Jordan-Nullmenge.
(2) f ist integrierbar, weil f über eine Fallunterscheidung durch zwei stetige Funktionen definiert ist.
(3) Beides sind korrekte Argumente.
(4) Keines dieser Argumente ist korrekt.

Test: Die erstellten ConcepTests wurden (in überarbeiteter Form) in einem weiteren Analysis-Zyklus (Analysis 1 im Sommersemester 2018 und Analysis 2 im Wintersemester 2018/19) im wöchentlichen Turnus eingesetzt. Parallel wird eine Wirkungsbeforschung im Rahmen des khdm-Projekts WiGeMath durchgeführt, über deren Design und Ergebnisse an anderer Stelle berichtet werden soll. Rückmeldungen von Studierenden haben gezeigt, dass die ConcepTests den Wunsch entstehen lassen, auch über die Übungsgruppen hinaus damit zu arbeiten, um beispielsweise vertiefter zu verstehen, warum bestimmte Antwortalternativen nicht richtig sind.

Diskussion und Fazit

Als praktisches Ergebnis des Projekts kann festgehalten werden, dass es gelang, für alle Themenbereiche der Analysis 1 und 2 ConcepTests zu konstruieren, die durchgängig im wöchentlichen Übungsturnus eingesetzt werden können. Anfängliche Bedenken des Autors, dass sich das Format (und insbesondere auch das Single-Choice-Antwortformat) möglicherweise nicht für alle Themenbereiche gleichermaßen eignen würde und in manchen Bereichen zu künstlich wirkenden Aufgabenstellungen führen könnte, haben sich nicht bewahrheitet. Die Kollektion der entwickelten Aufgaben wird in Buchform (Bauer, in Vorbereitung) zur Verfügung gestellt. Als Konsequenz von bereits in der ersten Design-Iteration gemachten Erfahrungen zur Praktikabilität werden auch Foliensätze für den effizienten Einsatz in Lehrveranstaltungen erstellt. Darüber hinaus sind ausführliche Lösungen erarbeitet worden, die auch den Einsatz der Aufgaben außerhalb von Lehrveranstaltungen ermöglichen. Insgesamt wurde das Design-Ziel einer einsetzbaren Intervention hiermit erreicht.

Als theoretisches Ergebnis wurde in der zweiten Iteration das oben dargestellte dreistufige Modell zur mathematikbezogenen Realisierung von (M1) und (M2) entwickelt und in der Aufgabenkonstruktion erprobt. Es hat sich hierfür als sehr gut brauchbar erwiesen, da es konkrete Leitlinien dafür angibt, auf welche Aspekte eines Begriffs oder Satzes die Fragestellung eines ConcepTests gerichtet wird. Umgekehrt hat es sich auch zur Analyse von bereits konstruierten Aufgaben bewährt. Obwohl durch das Modell die Lernwirksamkeit einer Aufgabe freilich nicht garantiert werden kann, wird aber doch gesichert, dass sie den betrachteten Begriff oder Satz in einer Tiefe bearbeitet, durch die dessen Sinn und Bedeutung bzw. Verwendung beleuchtet wird. Das Modell unterstützt somit den Aufgabensteller dabei, die Forderung von Mazur nach „konzeptuellen Fragen“ für das Fach Mathematik zu interpretieren und einzulösen. Eine weitere Verfeinerung wurde in der dritten Design-Iteration vorgenommen. Basierend auf den Ergebnissen des Einsatzes in der zweiten Iteration wurden Aufgaben so überarbeitet, dass sie Argumentationsprozesse direkter anstoßen.

Betrachtet man die Iterationen unter dem Blickwinkel von Kriterien für die Qualität von Interventionen (Nieveen 1999), so lässt sich im Einklang mit Plomb (2013, S. 29) feststellen, dass in den verschiedenen Design-Iterationen verschiedene Kriterien im Vordergrund standen: Während zunächst die prinzipielle Praktikabilität betrachtet wurde, wurde anschließend das Augenmerk auf Konsistenz gelegt; derzeit wird im Rahmen der dritten Iteration die Wirkung untersucht.

Hinsichtlich der Evaluationen in den ersten beiden Iterationen ist einschränkend festzuhalten, dass diese nur begrenzten Umfang hatten: In der ersten Iteration wurde kurze Interviews mit den Tutoren (als Lehrende) durchgeführt und in der zweiten Iteration wurden Rückmeldungen der Tutoren mittels wöchentlicher Feedbackbögen erhoben. (Über die darüber hinaus durchgeführte semesterabschließende Evaluation zur Ermittlung der Akzeptanz der Methode bei Studierenden und Tutoren wurde in Bauer 2018 berichtet). Aus diesem Grund ist Zurückhaltung hinsichtlich der Verallgemeinerbarkeit der im Projekt erarbeiteten Design-Prinzipien angebracht. Zudem sind bei Entwicklungsprojekten Aussagen über den im Projekt betrachteten Kontext hinaus generell nur mit Vorsicht möglich (vgl. Plomb 2013, S. 34, sowie Yin 2003, S. 37). Andererseits sind die hier zugrunde gelegten Konstrukte (Concept Definition/Image, Theorem Statement/Image) nicht an die Verwendung im Gebiet Analysis gebunden, sondern für das Mathematiklernen generell von Bedeutung. Insofern scheint die Erwartung nicht unvernünftig, dass die Ergebnisse auch auf weitere Themengebiete der Hochschulmathematik grundsätzlich anwendbar sind. Unterschiede dürften sich jedoch bei der Frage zeigen, auf welche Weise Concept Image oder Theorem Image jeweils domänenspezifisch durch Aufgaben erfasst werden können.

Offene Fragen

Die im Projekt bisher gemachten Erfahrungen legen eine Reihe von Fragen zur weiteren Untersuchung nahe:

1. *In Bezug auf den Schwierigkeitsgrad von ConcepTests:* Wo liegt der „richtige“ Schwierigkeitsgrad? Crouch et al. (2007) empfehlen als Orientierung, ConcepTests mit dem Ziel zu erstellen, dass im ersten Voting etwa 35 bis 70 Prozent der Antworten richtig ausfallen. Die Forderung nach einer unteren Schranke für richtige Antworten ist insofern plausibel, als hierdurch erwartet werden kann, dass in der Teilnehmerschaft ein Mindestmaß an korrekten Argumenten vorhanden ist, die die Chance erhalten, sich in der anschließenden Diskussionsphase durchzusetzen. Vor diesem Hintergrund überrascht es, dass in unserem Projekt auch gewisse ConcepTests, die deutlich unter der 35-Prozent-Grenze lagen, als sehr produktiv eingeschätzt wurden. Dies zeigt, dass in diesen Fällen für die produktive Wirkung von ConcepTests andere Faktoren relevant waren. Es wäre wichtig aufzuklären, welche dies genau sind.

2. *In Bezug auf den Zeitpunkt des Einsatzes:* Wann ist der richtige Zeitpunkt, um Inhalte der Vorlesung durch ConcepTests in Übungen zu thematisieren? Einerseits spricht es für einen Einsatz möglichst zeitnah zur Vorlesung, dass Peer Instruction dann auf die als Hausübungen zu bearbeitenden Aufgaben vorbereiten kann – relevante Konzepte werden diskutiert, bevor sie dort benötigt werden. Andererseits wünschten sich Studierende wiederholt, Inhalte erst dann durch Peer Instruction zu thematisieren, wenn sie einschlägige Übungsaufgaben bereits bearbeitet hatten, da sie sich dann informierter an den Diskussionen beteiligen könnten. In der Tat bestätigten Tutoren, dass sich ihrem Eindruck nach die Diskussionen dann oftmals tiefergehend gestalteten. Man gibt in dieser Durchführungsvariante allerdings den vorbereitenden Brückenschlag zu den Hausübungen auf. Es erhebt sich die Frage, welcher Aspekt mehr Gewicht verdient. Anders ausgedrückt: Es wäre wünschenswert zu klären, ob sich eine der Varianten insgesamt als lernförderlicher erweist.

3. *In Bezug auf die Wirkung:* Im Projekt konnten wir sowohl auf Seiten der Studierenden als auch auf Seiten der Tutoren eine deutlich positive Akzeptanz beim Einsatz von Peer Instruction zeigen. Insbesondere schätzten die Studierenden den Einsatz mit deutlicher Mehrheit als produktiv für ihr Lernen ein (siehe Bauer 2018, Abschn. 5.2). Freilich belegt dies nicht, dass tatsächlich ein besserer Lernerfolg entsteht. Um hierzu Aufklärung zu erhalten, wird eine diesbezügliche Wirkungsbeforschung derzeit im Rahmen des khdm-Projekts WiGeMath durchgeführt.

Literaturverzeichnis

Ableitinger, C., Herrmann, A. (2011). *Lernen aus Musterlösungen zur Analysis und Linearen Algebra. Ein Arbeits-und Übungsbuch*. Wiesbaden: Vieweg+Teubner Verlag.

Ableitinger, C. (2013). Demonstrationsaufgaben im Projekt „Mathematik besser verstehen“. In: C. Ableitinger, J. Kramer, S. Prediger (Hrsg.), *Zur doppelten Diskontinuität in der Gymnasiallehrerbildung* (S. 17–38). Wiesbaden: Springer Spektrum.

Amiel, T., Reeves, T. C. (2008). Design-Based Research and Educational Technology: Rethinking Technology and the Research Agenda. *Educational Technology & Society* 11 (4), 29–40.

Bauer, Th. (2018). Peer Instruction als Instrument zur Aktivierung von Studierenden in mathematischen Übungsgruppen. *Math. Semesterberichte*, First Online: 14.06.2018

Bauer, Th. (in Vorbereitung). *Verständnisaufgaben zur Analysis 1 und 2 – für Lerngruppen, Selbststudium, Peer Instruction*. Springer.

Biehler, R., Hochmuth, R., Klemm, J., Schreiber, S., Hänze, M. (2012). Fachbezogene Qualifizierung von MathematiktutorInnen – Konzeption und erste Erfahrungen im LIMA-Projekt. In: M. Zimmermann, C. Bescherer, C. Spannagel (Hrsg.), *Mathematik lehren in der Hochschule – Didaktische Innovationen für Vorkurse, Übungen und Vorlesungen* (pp. 45–56). Hildesheim, Berlin: Franzbecker.

Crouch, C. H., Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American journal of physics*, 69(9), 970–977.

Crouch, C. H., Watkins, J., Fagen, A. P., Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-Based Reform of University Physics*, 1(1), 40–95.

Mazur, E. (1997). Peer Instruction: Getting students to think in class. In: E.F. Redish, J.S. Rigden (eds.), *The Changing Role of Physics Departments in Modern Universities* (pp. 981–988). AIP Conference Proceedings 399

Mazur, E. (2017). *Peer Instruction*. Springer Spektrum, Berlin, Heidelberg.

Miller, R. L., Santana-Vega, E., Terrell, M. S. (2006). Can good questions and peer discussion improve calculus instruction? *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, 16(3), 193–203.

Nieveen, N. (1999). Prototyping to reach product quality. In: J. van den Akker, R.M. Branch, K. Gustafson, N. Nieveen, T. Plomp (Eds), *Design approaches and tools in education and training* (pp. 125–136). Boston: Kluwer Academic.

Plomp, T. (2013). Educational Design Research: An Introduction. In: T. Plomp, N. Nieveen (Hrsg.), *Educational Design Research* (S. 10-51). Enschede: Netherlands institute for curriculum development.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3–13.

Tall, D., Vinner, S. (1981). Concept Image and Concept Definition in Mathematics with particular reference to Limits and Continuity. *Educational Studies in Mathematics* 12, 151–169.

Van den Akker, J. (1999). Principles and Methods of Development Research. In: J. van den Akker, R.M. Branch, K. Gustafson, N. Nieveen, T. Plomp (Hrsg.), *Design approaches and tools in education and training* (S. 1–14). Boston: Kluwer Academic.

Yin, R.K. (2003). *Case Study Research: Design and Methods*. Newbury Park, CA: Sage.