Erschienen in: Math. Semesterberichte 66(2), 219-241 (2019)

Peer Instruction als Instrument zur Aktivierung von Studierenden in mathematischen Übungsgruppen

Thomas Bauer

21. Mai 2018

Zusammenfassung. Üben ist ein wesentliches Element des Mathematiklernens. Vorlesungsbegleitende Übungen gehören daher in der Mathematik seit langem zum selbstverständlichen Lehrangebot. Viele Dozenten machen allerdings die Erfahrung, dass Übungen die an sie gesetzen Erwartungen nicht immer erfüllen – insbesondere, wenn man es als wichtige Ziele betrachtet, dass dort Vorlesungsinhalte diskutiert, Fragen gestellt und Verständnisprobleme geklärt werden. Einer der Gründe dafür liegt darin, dass Studierende sich in Übungen oft passiv verhalten und es sehr vom Geschick des einzelnen Übungsleiters abhängt, wie viel an Aktivierung in der Übung gelingt. Im vorliegenden Beitrag wird gezeigt, wie die Methode der Peer Instruction wirkungsvoll genutzt werden kann, um Studierende in mathematischen Übungsgruppen zu aktivieren. Der Autor hat die Methode in Übungen zu den Vorlesungen Analysis I und II erfolgreich eingesetzt. Er stellt im Beitrag die konzeptionellen Grundlagen vor, zeigt Beispiele und diskutiert Ergebnisse aus dem Pilotdurchgang.

Inhaltsverzeichnis

1 Zur Rolle des Übens im Mathematikstudium

Das Üben bildet ein selbstverständliches Element der universitären Mathematikausbildung. Es besteht weithin Einigkeit darüber, dass Mathematiklernen nicht durch bloßes Zuhören in Vorlesungen oder durch Lesen von Büchern möglich ist, sondern dass für wirksames Lernen intensive Phasen von Eigentätigkeit durch die Lernenden erforderlich sind. Viele Dozenten haben allerdings die Erfahrung gemacht, dass dies Studierenden oft nicht bewusst ist und sie sich stattdessen auf das Memorieren von Sätzen und Definitionen beschränken. Zahlreiche an Studierende gerichtete Texte von Dozenten betonen

Thomas Bauer, Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, Hans-Meerwein-Straße, 35032 Marburg, tbauer@mathematik.uni-marburg.de

daher, wie wichtig Eigenaktivität für den Erfolg im Mathematikstudium ist – oft ausgedrückt in der griffigen Formulierung, dass Mathematik kein „spectator sport" sei – und wie entscheidend das intensive Arbeiten an Übungsaufgaben für den Lernerfolg ist (zum Beispiel Lehn o. D.). Dies entspricht einer im Kern konstruktivistischen Lernauffassung: Lernen ist nicht allein durch Instruktion möglich, sondern erfordert Konstruktionsphasen, in denen das Individuum eigentätig Wissen aufbaut – man strebt daher eine angemessene Balance aus Instruktion (etwa in einer Vorlesung) und Konstruktion (durch Eigenaktivität beim Nachbereiten von Vorlesungen und besonders beim Bearbeiten von Übungsaufgaben) an (vgl. Hefendehl 2005).

An das Bearbeiten von Übungsaufgaben als wesentliche Form von Eigenaktivität beim Mathematiktreiben sind im Mathematikstudium wöchentliche Übungen gekoppelt. Eine ihrer Funktionen kann das Vorstellen von Musterlösungen sein, wenn Lernende dabei erleben, mit welchen Strategien ein Experte eine Problemlöseaufgabe angeht, wie er die gewählte Strategie durchführt und schließlich in welcher Form er die Lösung darstellt (vgl. Ableitinger und Herrmann 2011). Wäre das Vorstellen von Lösungen das einzige Ziel, so würde eine vorlesungsartige „Zentralübung" als Organisationsform ausreichen. Übungen haben jedoch weitere wichtige Ziele: Sie sollen die Gelegenheit zur Diskussion geben, in der Studierende Fragen stellen und Tutoren auf individuelle oder inter-individuelle Schwierigkeiten eingehen. Damit dies gelingt, ist es notwendig, Studierende so weit zu aktivieren, dass sie tatsächlich Fragen stellen und ihre Probleme zur Sprache bringen. Dies wiederum erfordert, dass die Studierenden sich eventueller Probleme bewusst sind und sie keine Scheu haben, diese in der Übung zu äußern.

Die – meist studentischen – Tutoren (Übungsleiter) haben in dieser Situation eine facettenreiche und anspruchsvolle Aufgabe (siehe Biehler et al. 2012). Tutorenschulungen können zwar wertvolle unterrichtspraktische Anleitungen für den Umgang mit Lerngruppen geben und die Tutoren auf diese Weise unterstützen, jedoch können naturgemäß in einer kurzen Schulung nicht umfangreiche mathematikdidaktische Fähigkeiten aufgebaut werden, die nötig wären, um etwa bei Studierendenäußerungen zu erkennen, welches Defizit zugrundeliegt (z.B. Wissenslücke oder Fehlvorstellung zu einem mathematischen Konzept) und welche Maßnahmen zur Unterstützung fachlich angemessen und didaktisch wirksam wären.

2 Aktivieren von Studierenden

2.1 Aktivieren – warum?

Obwohl sich das Grundkonzept vorlesungsbegleitender Übungen bewährt hat, haben viele Dozenten und auch Studierende den Eindruck, dass Übungen oftmals die an sie gesetzten Erwartungen nicht voll erfüllen. Es besteht insbesondere die Gefahr, dass die Studierenden in den Übungen in einer zu passiven Rolle verbleiben – allzu leicht werden das Vorstellen von Musterlösungen und instruktionale Erklärungen des Tutors (gar in Form einer „Kurzvorlesung") aus Sicht der Studierenden oder der Tutoren zur Hauptsache. Wie weit es zum Stellen von Fragen, Diskutieren und Klären von Problemen bei Vorlesungsinhalten kommt, hängt sehr vom Geschick des einzelnen Tutors ab – es besteht die Gefahr, dass diese wichtigen Elemente in den Hintergrund geraten.

Erkennt man diese Gefahren an, dann stellt sich für Dozenten die Frage, welche Gestaltungsideen und -vorgaben zu starker und lernförderlicher Aktivierung führen können.

Dabei sind die Rahmenbedingungen zu bedenken: Studentische Tutoren sind in aller Regel motiviert und engagiert, aber naturgemäß oft unerfahren. Studierende sind in aller Regel lernbereit, haben aber oftmals – gerade in der Studieneingangsphase – noch keine effektiven Lernstrategien für universitäres Lernen entwickelt. Man kann daher nicht immer erwarten, dass sie selbst die Diskussion in der Übung durch Artikulation spezifischer Probleme in Gang bringen. Die Frage ist also, wie sich die Arbeit mit engagierten Tutoren und lernbereiten Studierenden so gestalten lässt, dass eine optimale Lernwirkung in den Übungen erreicht wird. Dies bedeutet insbesondere:

- Wie lässt sich erreichen, dass Studierende in der Übung über zentrale Konzepte und Prinzipien eines Lerngegenstands nachdenken, sich ihrer Vorstellungen dazu bewusst werden und gegebenenfalls Fehlvorstellungen aufdecken?
- Wie lässt sich erreichen, dass Studierende über Verständnisprobleme sprechen?
- Wie lässt sich erreichen, dass Diskussionen in der Übung zu vertieftem Verständnis führen?

2.2 Ist jede Art von Aktivierung hilfreich?

Der Begriff „Aktivierung" kann auf den ersten Blick selbsterklärend erscheinen. Bei näherer Betrachtung stellt sich jedoch heraus, dass es wichtig ist herauszuarbeiten, was damit genau ausgedrückt werden soll und welche Art von Aktivierung man anstreben sollte, um möglichst lernwirksame Bedingungen zu schaffen.

Renkl (2011) unterscheidet aus Sicht der Lernpsychologie drei Perspektiven: die des *aktiven Tuns*, die der *aktiven Informationsverarbeitung* und die der *fokussierten Informationsverarbeitung*. Während aus der Perspektive des *aktiven Tuns* der Blick in erster Linie auf das äußerlich beobachtbare Handeln gerichtet ist, wie es etwa beim gemeinsamen Problemlösen oder im fachlichen Diskurs auftritt, berücksichtigt die Perspektive der *aktiven Informationsverarbeitung*, welche gedanklichen stoffbezogenen Aktivitäten dabei tatsächlich ausgeübt werden. Die darauf aufbauende Perspektive der *fokussierten Informationsverarbeitung* richtet den Blick genauer darauf, ob sich die Aktivitäten auf die zentralen Konzepte und Prinzipien des Lerngegenstands beziehen. Renkl (a.a.O.) argumentiert, dass für guten Lernerfolg fokussierte Informationsverarbeitung angestrebt werden sollte. Aktivität ist demnach nicht an und für sich bereits lernförderlich, sondern es kommt darauf an, in welchem Maße sie sich auf den konzeptuellen Kern des Lerngegenstands richtet. Entsprechend warnen Leuders und Holzäpfel (2011) aus mathematikdidaktischer Sicht vor einem Methodeneinsatz, der lediglich *äußere Aktivierung* erreicht, und betonen, dass es darauf ankommt, Methoden fachbezogen wirksam einzusetzen, um echte *kognitive Aktivierung* zu erreichen.

2.3 Die Rolle der Interaktion – Das Rahmenmodell passiv–aktiv–konstruktiv–interaktiv

Um bei Überlegungen zur Aktivierung auch die Rolle der Interaktion präziser zu erfassen, lässt sich ein Rahmenmodell von Chi (2009) nutzen. In diesem Modell werden Lernaktivitäten in drei Kategorien unterschieden: aktiv, konstruktiv, interaktiv. Hierbei werden Lernhandlungen *aktiv* genannt, wenn sie gegebene Informationen in einer nach außen sichtbaren Weise verarbeiten (im Gegensatz zum reinen Zuhören bei einer Vorlesung oder

passiv	*aktiv*	*konstruktiv*	*interaktiv*
einer Vorlesung zuhören, ein Video ansehen, einen Text lesen	Notizen anfertigen, eine Zusammenfassung schreiben	Selbsterklärungen geben, Fragen stellen, Diagramme zeichnen, Beispiele bilden, Aufgaben bearbeiten	Gegenseitig fragen und antworten, eigene Behauptungen begründen, die Behauptungen des Partners in Frage stellen

Tabelle 1: Kategorien von Handlungen beim Mathematiklernen (vgl. Chi 2009, Fonseca und Chi 2011)

dem Anschauen eines Videos). *Konstruktive* Handlungen gehen dabei über die gegebene Information hinaus – zum Beispiel dadurch, dass der Lernende Selbsterklärungen gibt oder Fragen formuliert. Beim Mathematiklernen kann man insbesondere das Arbeiten an Aufgaben, die die Anwendung neuen Wissens erfordern, dieser Kategorie zuordnen. *Interaktive* Lernaktivitäten gehen über aktive Handlungen dadurch hinaus, dass die Aktivität im Dialog mit einer weiteren Person (einem Experten, einem Peer) erfolgt. Allerdings ist nicht jedes Gespräch bereits eine interaktive Lernhandlung: Wenn etwa einer der Dialogpartner erklärt und der andere im wesentlichen zuhört, so liegt vielmehr eine aktive oder konstruktive Handlung des Redners vor, während der Partner passiv bleibt. Tabelle 1 stellt diese Kategorisierung dar, wobei sie wie in Fonseca und Chi (2011) durch eine *passive* Kategorie ergänzt ist.

Die beschriebene Kategorisierung ist in diesem Beitrag insofern von Interesse, als nach Chi (2009) erwartet wird, dass die Lernwirkung in Tabelle 1 von links nach rechts zunimmt. Insbesondere die Auswirkungen von Interaktion zwischen Lernenden sind vielfach untersucht (siehe z.B. Wentzel und Watkins 2011) und die Befunde deuten darauf hin, sowohl dyadisches Lernen (Zweiergruppen) als auch Lernen in größeren Gruppen positive Effekte auf die Lernwirkung haben kann. Erwartungsgemäß hängt die Stärke des Effekts dabei von der Qualität der Interaktion und von weiteren Rahmenbedingungen ab, z.B. welches Maß an persönlicher Verantwortlichkeit den Lernenden abverlangt wird (*accountability*) und welche sozialen Fähigkeiten zur kooperativen Zusammenarbeit die Beteiligten haben – Interaktion kann lernwirksam sein, ist es aber nicht an und für sich.

3 Die Methode der Peer Instruction – Idee und Durchführung

In diesem Abschnitt wird die Methode der Peer Instruction vorgestellt. Sie hat sich in der Erprobung durch den Verfasser als sehr praktikable und effektive Methode erwiesen, bei der durch Interaktion zwischen Studierenden fokussierte Informationsverarbeitung stattfindet. Wie die Ergebnisse (siehe Abschn. 5) zeigen, ist sie von Tutoren in einfacher Weise durchzuführen und wird von Studierenden sehr gut angenommen.

3.1 Die Idee der Peer Instruction

Die Methode der Peer Instruction wurde Ende der 90er Jahre von Eric Mazur entwickelt, der sie in der universitären Ausbildung im Fach Physik einsetzt. (Mazur 1997, Crouch et al. 2007). In Mazurs Konzeption wird Peer Instruction in Vorlesungen eingesetzt und unterstützt dort die Realisierung eines Inverted-Classroom-Konzepts. Da im vorliegenden Aufsatz gezeigt werden soll, wie die Methode für den Einsatz in *Übungen* genutzt werden kann, werden diejenigen Aspekte, die die ursprüngliche Vorlesungseinbettung betreffen, hier nicht ausgeführt (siehe dazu Mazur 1997). Stattdessen konzentrieren wir uns hier auf die methodische Form der Peer Instruction im engeren Sinne.

Eine Peer-Instruction-„Runde”, die 5-15 Minuten in Anspruch nehmen kann, beginnt mit einer Multiple-Choice-Frage, die ein fachliches Konzept betrifft. Im Fach Physik könnte dies der Begriff *Kraft* sein oder eine Gesetzmäßigkeit wie das *Erste Newtonsche Gesetz*. Die Frage – im Sprachgebrauch der Peer Instruction ein *ConcepTest* genannt – wird den Studierenden vorgelegt, die sich nach kurzem Überlegen auf eine Antwortoption festlegen müssen. Dies geschieht individuell und ohne vorherige Diskussion mit Mitstudierenden. Die Entscheidung wird von der Lehrperson in einer Abstimmung (*Voting*) festgehalten. Die Festlegung auf eine Antwortoption (*Commitment*) ist insofern von Bedeutung, als die Studierenden in der nächsten Phase aufgefordert werden, andere Studierende in Kleingruppen von ihrer Wahl zu überzeugen – bzw. sich von anderen überzeugen zu lassen. Nach der Diskussionsphase wird eine zweite Abstimmung durchgeführt, in der die Studierenden ihre erste Entscheidung auf Grundlage der Diskussion revidieren können. Mazur berichtet (a.a.O), dass sich die richtige Antwortalternative durch die Diskussion in der großen Mehrzahl der Fälle durchsetzt. In problematischen Fällen – oder falls generell noch Klärungsbedarf besteht – kann von der Lehrperson eine abschließende Erläuterung folgen.

In der Diskussionsphase (*Peer Discussion*) liegen der Kern der Methode und die erwünschten Wirkungen:

- Es werden nicht nur wenige hochmotivierte Studierende aktiviert, sondern im Idealfall alle Teilnehmer. Es kann echte auf den Kern des Lerngegenstands bezogene fokussierte Aktivierung erreicht werden, die sowohl für stärkere als auch für schwächere Studierende von Nutzen sein kann.
- Die Studierenden befassen sich mit dem fachlichen Konzept, überprüfen ihre Vorstellungen, ordnen Argumente und erklären ihre Überlegungen. Durch die Verbalisierung und den Diskurs können Argumente ausgeschärft, Unklarheiten und Schwierigkeiten sichtbar gemacht, Fehlvorstellungen aufgedeckt und geklärt werden. Sowohl schwächere als auch stärkere Studierende werden dabei herausgefordert – letztere insbesondere dadurch, dass die Methode es nicht erlaubt, Standpunkte als „klar”, „trivial”, etc. abzutun, sondern fordert, diese mit stichhaltigen Argumenten zu belegen.

Solche Wirkungen von der Diskussion zu erwarten steht ganz im Einklang mit interaktionistischen Sichtweisen. Bauersfeld (2000, S. 138) schreibt (in anderem Kontext):

> [. . .] fordert das Zusammenarbeiten mit dem Partner die Artikulation, Offenlegung und Begründung des eigenen Vorstellens und Denkens heraus und ermöglicht rückwirkend dessen Bereicherung in der interaktiven Auseinandersetzung.

3.2 Praktische Durchführung

Die Erprobung fand in einem Pilotdurchgang in den Modulen Analysis I und Analysis II im Sommersemester 2014 bzw. Wintersemester 2014/15 statt, in dem sie in den Übungen durchgängig eingesetzt wurden. Dies wurde in Bauer (2015, Abschn. 3) kurz berichtet. Als Rahmen wurden folgende Festlegungen getroffen:

- Es werden 30 Minuten zu Beginn jeder 90-Minuten-Übung für Peer Instruction verwendet. In dieser Zeit werden ca. 3 Peer-Instruction-Runden durchgeführt.
- Die ConcepTests wurden wöchentlich vom Dozenten vorbereitet. (Der Entwurf der ConcepTests wird in Abschn. 4 erläutert.) Für die Übungen zur Analysis I wurden insgesamt 65 ConcepTests verwendet, davon 20 aus dem GoodQuestionsProjekt von Terrel und Connelly (2005), für die Analysis II wurden 89 ConcepTests entwickelt.
- Jede Runde wurde nach dem in Abb. 2 gezeigten dreiphasigen Modell durchgeführt. Dabei folgten wir der Empfehlung von Mazur (1997), Phase 2 nur dann durchzuführen, wenn 35–70 Prozent der Teilnehmer sich bei der Abstimmung in Phase 1 für die richtige Antwortoption entschieden haben (siehe hierzu weitere Ausführungen in Abschn. 4.1).

Für die Durchführung der Votings können verschiedene Methoden eingesetzt werden: Zum einen stehen technische Lösungen zur Verfügung, die aus dem Einsatz von „Klickerfragen" in Vorlesungen bekannt sind:

- Abstimmungsgeräte (*Clicker*), die speziell hierfür angeschafft werden
- Online-Abstimmungssysteme wie PINGO (http://trypingo.com/)
- Voting-Systeme, die in universitären Lernplattformen verfügbar sind, z.B. *Live-Voting* in ILIAS

In Übungsgruppen sind dank der kleineren Teilnehmerzahlen auch ganz einfache Low-Tech-Methoden praktikabel, wie das Abstimmen mit Voting Cards oder das Abstimmen per Handzeichen.

Um die Tutoren auf den Einsatz der Methode vorzubereiten, wurde vor Semesterbeginn eine Kurzschulung durchgeführt (90 Minuten). Diese war in folgender Weise gestaltet:

1. Es wurden die Ziele von Übungen sowie die Problematik der Aktivierung von Studierenden diskutiert.
2. Das Konzept der Peer Instruction wurde den Tutoren vorgestellt und an Beispielen erläutert. Die erwünschten Wirkungen zur Aktivierung wurden diskutiert.
3. Es wurden einige Peer-Instruction-Runden mit den Tutoren als Teilnehmer durchgeführt. Auf diese Weise konnten diese die Methode praxisnah kennenlernen und ihre Wirkung selbst erleben.
4. Es bestand Gelegenheit, Fragen zur praktischen Durchführung zu besprechen, die durch das Kennenlernen im vorigen Schritt entstanden waren.

4 Design von ConcepTests zur Peer Instruction – Gute Fragen konstruieren

Der Erfolg von Peer Instruction steht und fällt mit der Qualität der ConcepTests, die man einsetzt. Mazur (1997) gibt dazu überzeugende Beispiele für das Fach Physik. Im Fach Mathematik wurden Beispiele für einen Calculus-Kurs im *GoodQuestions-Projekt* vorgestellt (Terrel und Connelly 2005, Miller et al. 2006) Auf welche Weise und nach welchen

Phase 1 (ca. 2–5 Min.)	Einzelarbeit	Die Studierenden lesen den ConcepTest. Sie überlegen, welche der Antwortalternativen richtig ist und suchen Argumente dafür und Argumente gegen die anderen Alternativen.
Voting 1		Entscheidung für eine der Antwortalternativen (Festlegung, committment)
Phase 2 (ca. 5-10 Min.)	Gruppenarbeit	„Überzeuge Deinen Nachbarn": Gruppen von 2–3 Studierenden, die nicht alle dieselbe Antwortalternative gewählt haben, versuchen sich gegenseitig zu überzeugen.
Voting 2		Neue Entscheidung für eine Antwortalternative auf Basis der Diskussion
Phase 3 (optional)	Plenum	Abschließende Klärung durch die Lehrperson

Tabelle 2: Ablauf einer Runde der Peer Instruction: Phase 2 ist die Kernphase der Methode, auf der ihre eigentliche Lernwirkung beruht. Lässt man diese Phase weg und führt stattdessen nur Phase 1 mit Voting 1 durch („Klickerfragen"), so wird man nicht dieselbe Lernwirkung erwarten können.
Crouch und Mazur (2001) schlagen für die Phasen 1 und 2 nur 1–2 Minuten bzw. 2–4 Minuten vor. Im vorliegenden Projekt haben sich allerdings die in der Tabelle angegebenen etwas längeren Zeiten als günstig erwiesen, um die Studierenden in Phase 1 mehr zum Nachdenken als zum Raten zu veranlassen und um in Phase 2 tiefergehende Diskussionen zu ermöglichen.

Prinzipien ConcepTests für das Fach Mathematik entworfen werden können, lässt sich daraus allerdings nicht generell ableiten. Wir stellen hierzu in diesem Abschnitt konzeptionelle Überlegungen vor. Diese haben sich als Grundlage für den Entwurf eines Katalogs von ConcepTests zur Analysis I und II bewährt und lassen sich auch auf weitere Themengebiete der Mathematik anwenden.

4.1 Grundsätzliche Anforderungen an Fragen

Mazur gibt aus seiner Erfahrung grundsätzliche Empfehlungen für den Entwurf von ConcepTests. In Anlehnung an Crouch et al. (2007) lassen sich diese so beschreiben:

(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 Memorieren von Fakten oder 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 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.

Schritt 1	*Schritt 2*	*Schritt 3*	*Schritt 4*
Zentrale Konzepte des Gebiets identifizieren (M1) →	Konzeptuelle Fragen in passendem Anspruchsniveau entwerfen (M2 und M5) →	Distraktoren aus bekannten Fehlern generieren (M3) →	ConcepTest pilotieren und Entwurf ggf. anpassen (M4)

Abb. 1: Arbeitsplan, der für das vorliegende Projekt entworfen wurde, um ConcepTests für Peer Instruction zu konstruieren, die den Kriterien (M1)–(M5) genügen.

Diese Empfehlungen werden gestützt durch Befunde von Miller et al. (2006): Der Nutzen der Peer Instruction rührt nicht in erster Linie daher, dass Fragen gestellt und Antworten gewählt werden, sondern liegt in der Phase der Peer Discussion. Daher ist es entscheidend, wie gut die ConcepTests zur Diskussion anregen. Wie in Abschnitt 2 dargelegt, darf man allerdings nicht von jeder Aktivität und Interaktion Lernförderliches erwarten, sondern es kommt darauf an, dass eine möglichst fokussierte Informationsverarbeitung erreicht wird, die den Kern der Konzepte und die damit typischerweise verbundenen Lernschwierigkeiten trifft.

Der Autor hat für den Entwurf von ConcepTests einen vierschrittigen Arbeitsplan aufgestellt, der sich bei der Durchführung des Projekt gut bewährt hat (siehe Abb. 1):

Schritt 1. Zunächst ist ein fachlicher Gegenstand auszuwählen, der in dem ConcepTest bearbeitet werden soll (M1). Das Ziel dabei ist, die zentralen Konzepte, die in der jeweiligen Vorlesungswoche behandelt wurden, durch ConcepTests abzudecken, die in der zugehörigen Übung bearbeitet werden.

Schritt 2. Die Herausforderung liegt dann darin, jeweils eine konzeptuelle Frage zu entwerfen (M1), die für die Zielgruppe das passende Anspruchsniveau hat (M5).
Als *konzeptuelle Fragen* werden im Folgenden solche Fragen bezeichnet, die nicht nur die Oberfläche korrekter Sprechweisen abfragen, sondern die auf das Konzept selbst und seinen Bedeutungsgehalt zielen. (Hierzu folgen in den nächsten Abschnitten weitere Ausführungen.)

Schritt 3. Die Lehrerfahrung des Dozenten wird dann als empirische Grundlage benötigt, um inkorrekte Antwortoptionen (Distraktoren) aus bekannten Fehlern und Verständnisschwierigkeiten von Studierenden zu generieren (M3).

Schritt 4. Um eindeutige Formulierungen zu erreichen (oder, realistischer formuliert, diesem Ziel jedenfalls möglichst nahe zu kommen), wird der ConcepTest pilotiert. Dies kann im einfachsten Fall dadurch geschehen, dass er vor dem Einsatz Mitarbeitern zur Beantwortung vorgelegt wird und deren Rückmeldungen für eine Überarbeitung genutzt werden.

4.2 Konstruktion von ConcepTests für das Fach Mathematik – Was macht eine gute Frage aus?

In den folgenden Teilabschnitten geht es darum, wie für das Fach Mathematik ConcepTests entworfen werden können, die den Anforderungen (M2) und (M5) genügen. Das Ziel sind demnach konzeptuelle Fragen, die zu einer Diskussion auffordern, durch die die Lernenden zum konzeptuellen Kern eines mathematischen Gegenstands vordringen und dabei nach Möglichkeit auch typische Fehlvorstellungen bearbeiten.

Wir unterscheiden dabei zwischen Aspekten, die zu berücksichtigen sind, wenn sich der ConcepTest auf das Lernen eines mathematischen *Begriffs* bezieht, und denen, die beim Lernen eines mathematischen *Satzes* von Bedeutung sind. Anschließend gehen wir darauf ein, auf welche Weise erreicht werden kann, dass ein ConcepTest genügend starke Anregung zum *Argumentieren* beinhaltet.

4.2.1 Zum Begriffslernen – Ausbilden des *Concept Image*

Betrachtet man einen mathematischen Begriff wie *Cauchy-Folge*, so kann man zunächst an die verbale Formulierung einer Definition denken:

> Eine Folge $(a_n)_{n\in\mathbb{N}}$ reeller Zahlen heißt *Cauchy-Folge*, falls gilt: Zu jedem $\varepsilon > 0$ gibt es ein $N \in \mathbb{N}$, so dass die Ungleichung $|a_n - a_m| < \varepsilon$ für alle $n, m \geq N$ gilt.

Allerdings würde man sicherlich das bloße Angeben einer solchen Definition noch nicht als „Verstehen" des Begriffs werten. So würde man etwa in einer mündlichen Prüfung an dieser Stelle sogleich feststellen wollen, ob die Angabe der Definition mit „echtem" Verständnis verbunden ist:

- Was besagt diese Definition inhaltlich?
- Wie kann man dies (z.B. mit Hilfe einer Zeichnung) erläutern?
- Worin liegt der Unterschied zum Konvergenzbegriff?
- Welche Verwendungen hat der Begriff? Wo wird er im Theorieaufbau der Analysis eingesetzt und wie geschieht dies dort konkret?

Der Unterschied zwischen dem Angeben einer Definition und weitergehendem Verstehen lässt sich mit den von Tall and Vinner (1981) eingeführten Konstrukten *Concept Definition* und *Concept Image* erfassen: Während *Concept Definition* einfach eine textuelle Formulierung der Definition meint, umfasst *Concept Image* die gesamte gedankliche Struktur, die ein Individuum dazu aufgebaut hat. Diese enthält alle Vorstellungen zum Begriff, dessen Eigenschaften und Verwendungen. Es stellt damit das Abbild des Begriffs im Denken des Lernenden dar.

Eine solche Sichtweise auf das Begriffsverstehen liegt ganz in der Tradition der Mathematikdidaktik, wonach Begriffserwerb als *Konstitutieren mentaler Objekte* aufgefasst wird (Freudenthal 1983). Dabei gehören nach Vollrath (1984) zum Begriffsverständnis insbesondere Fähigkeiten im Umgang mit Eigenschaften eines Begriffs (*Begriffsinhalt*), ein Überblick über die zugehörigen Objekte (*Begriffsumfang*), das Verständnis der Beziehungen der Begriffe zueinander (*Begriffsnetz*) sowie Fähigkeiten im Umgang mit dem Begriff beim Lösen von Problemen (siehe hierzu auch Weigand 2015). Ein solcher umfassender Blick auf das Begriffsverstehen macht deutlich, dass es beim Begriffslernen nicht damit getan ist, dass der Dozent einen Begriff „einführt", sondern dass es ein „aktiver, schöpferischer Prozess des lernenden Individuums" ist (Winter 1983, S. 186).

4.2.2 Zum Satzlernen – Ausbilden des *Theorem Image*

Analog zum Begriffslernen kann man beim Lernen mathematischer Sätze zunächst an eine verbale Formulierung denken – im Beispiel des Zwischenwertsatzes etwa an:

> *Sei $f : [a,b] \to \mathbb{R}$ eine stetige Funktion und c eine reelle Zahl zwischen $f(a)$ und $f(b)$. Dann gibt es ein $p \in [a,b]$ mit $f(p) = c$.*

Von „Verstehen" des Satzes wird man erst sprechen, wenn zur Angabe einer Satzformulierung weitere Wissenselemente hinzukommen, die *Sinn und Bedeutung* des Satzes betreffen:

- Was besagt der Satz inhaltlich?
- Wie kann man die Aussage (z.B. mit Hilfe einer Zeichnung) erläutern?
- Wird die Aussage falsch, wenn der Definitionsbereich kein kompaktes Intervall ist oder die Funktion nicht stetig ist? Welche Beispiele zeigen dies?
- Wie gehen diese Voraussetzungen in den Beweis ein?
- Welche typischen Verwendungen hat der Satz (im systematischen Aufbau der Analysis, in Anwendungen)?

Die gedankliche Struktur, die ein Individuum zu einem Satz aufbaut, soll hier in Analogie zum Concept Image als *Theorem Image* bezeichnet werden. Von echten Verstehen eines Satzes würde man nur sprechen, wenn das Theorem Image reichhaltig genug ist. Diese Sichtweise auf das Satzverstehen ist kompatibel zur Auffassung von Selden and Selden (1995), wo das Konstrukt des *Statement Image* eingeführt wird, das übergreifend für Definitionen und Sätze verwendet wird. In diesem Sinne sind Concept Image und Theorem Image zwei Spezialfälle von Statement Images.

4.2.3 Unterstützen des Begriffs- und Satzlernens durch Peer Instruction

Um das Begriffs- und Satzlernen zu fördern, wird im vorliegenden Projekt angestrebt, die Studierenden durch Peer Instruction beim Ausbilden eines reichhaltigen *Concept Image* bzw. *Theorem Image* zu den in der Vorlesung behandelten Begriffen und Sätzen zu unterstützen. Die damit gemeinten Ziele sind in Tabelle 3 in einem dreistufigen Modell dargestellt. Demnach liegt auf Stufe 1 die Kenntnis von Concept Definition bzw. Theorem Statement – allerdings bereits in etwas ausgeschärfter Form, die über das bloße Memorieren des Texts hinausgeht. Die Stufen 2 und 3 betreffen dann Concept Image bzw. Theorem Image: Stufe 2 enthält zunächst das Verstehen von Sinn und Bedeutung, während Stufe 3 die Fähigkeit zur Verwendung des Begriffs oder Satzes betrifft. In der Regel bauen die Stufen aufeinander auf: Kompetente Verwendung gelingt erst, wenn Sinn und Bedeutung verstanden sind, was wiederum die Fähigkeit zur präzisen Formulierung als Teilkomponente beinhaltet. Um eine möglichst große Verstehenstiefe zu erreichen, wurde beim Entwurf von ConcepTests im vorliegenden Projekt auf das Erreichen der Stufen 2 und 3 Wert gelegt.

Wir stellen einige Beispiele vor, anhand derer aufgezeigt werden kann, wie die bisherigen Überlegungen in die Konstruktion von ConcepTests eingehen:

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

Tabelle 3: Ziele der Peer Instruction beim Wissensaufbau zu Begriffen und Sätzen

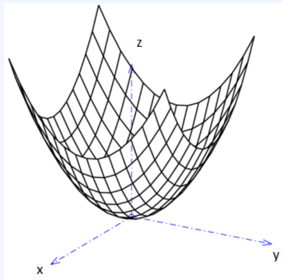
Welche Formulierung bringt zum Ausdruck, dass eine Reihe $\sum a_n$ konvergiert?

(1) Die Folgenglieder a_n sind durch eine explizite Formel gegeben, man kann sie daher explizit aufsummieren.
(2) Die Folgenglieder a_n werden für hinreichend großes n beliebig klein, man braucht daher nur endlich viele davon zu berücksichtigen, um den Summenwert zu erhalten.
(3) Es gibt einen Index n, ab dem die Summen $a_1 + \ldots + a_n$ nicht mehr größer werden.
(4) Wenn man immer längere Summen $a_1 + \ldots + a_n$ bildet, dann entsteht eine konvergente Folge von Summen.

Dieser ConcepTest zur Analysis I betrifft den Begriff *Reihe*. Er betrachtet mögliche Vorstellungen zur Konvergenz von Reihen und bearbeitet damit das zugehörige *Concept Image* (Begriffsinhalt, Stufe 2). Die Distraktoren sind aus bekannten Fehlvorstellungen gebildet (M3).

Das folgende Beispiel aus der Analysis II bearbeitet das *Concept Image* zum Begriff *Gradient*:

Das Bild zeigt den Graphen der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto x^2 + y^2$
Ihr Gradient im Nullpunkt ist $(0,0)$ und sie hat folgende Eigenschaften:



(A) Die Tangentialebene im Ursprung ist die x-y-Ebene.
(B) Die x-y-Ebene schneidet den Graphen nur im Punkt $(0,0)$.
(C) Die Funktion hat im Nullpunkt ein Extremum.

Was kann jemand schließen, der von einer differenzierbaren Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(0,0) = 0$ *nur* weiß, dass der Gradient im Nullpunkt gleich $(0,0)$ ist, aber Funktion und Bild nicht kennt?

(1) nur (A)
(2) nur (B)
(3) nur (C)
(4) (A) und (B), aber nicht (C)

Hier geht es zunächst darum, den Gradienten mit Vorstellungen zur Tangentialebene zu verbinden (Begriffsinhalt). Aber auch die Vernetzung der Begriffe *Gradient* und *Extremum* wird aktiviert (Begriffsnetz): Um den Schluss auf (B) oder (C) zu widerlegen (bekannte Fehlschlüsse), werden (den Begriffsumfang betreffende) Beispielfunktionen benötigt wie $(x,y) \mapsto x^2 - y^2$, die im Ursprung kein Extremum hat, bzw. $(x,y) \mapsto x^2$, die die x-y-Ebene nicht nur im Ursprung schneidet.

Der folgende ConcepTest aus der Analysis I bearbeitet das *Theorem Image* zum Satz über Maximum und Minimum:

Wenn eine Funktion $f : [a,b] \to \mathbb{R}$ kein Maximum hat, dann

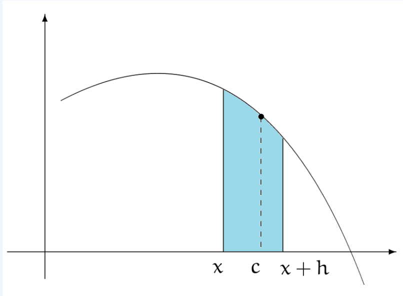
(1) muss sie unstetig sein
(2) muss sie unbeschränkt sein
(3) beides
(4) keines von beiden

Um eine Antwort zu finden, ist zu überlegen, welche Schlüsse der Satz in der gegebenen Verwendungssituation erlaubt. Zugleich wird – um Antwort (2) zu widerlegen – ein genügend reichhaltiges Bild zum Begriff *Maximum* benötigt, insbesondere Vorstellungen dazu, wie sich Funktionen verhalten können.

Solche Verknüpfungen von Begriffs- und Satzverstehen innerhalb eines ConcepTests können zum einen explizit konstruiert werden, zum anderen benötigt ein adäquates Theorem Image ohnehin die Verknüpfung mit den Concept Images der im Satz vorkommenden Begriffe – insofern liegt hier kein Konflikt mit der Anforderung (M1) vor. Der folgende ConcepTest zeigt dies, da er einerseits die Verwendung des Mittelwertsatzes der Integralrechnung (Theorem Image, Stufe 3) betrifft, aber auch Vorstellungen zum Integral (Concept Image, Stufe 2) bearbeitet.

Im folgendem Bild ist der Graph einer stetigen Funktion f zu sehen.
Wir betrachten zu festem a die Integralfunktion

$$F : x \mapsto \int_a^x f(t)\,dt$$



Welche Aussage über den Ausdruck

$$\frac{F(x+h)-F(x)}{h}$$

ist richtig?

(1) Er gibt den Flächeninhalt des farblich markierten Flächenstücks an.
(2) Er ist gleich dem Funktionswert $f(c)$ für ein geeignetes c zwischen x und $x+h$.
(3) Beides ist richtig.
(4) Keines von beiden ist richtig.

Hier werden zunächst adäquate Vorstellungen zum Integral benötigt, um $F(x+h)-F(x)$ als Inhalt des markierten Flächenstücks zu erkennen (wodurch Antwort 1 als falsch erkannt wird). Der Mittelwertsatz der Integralrechnung liefert dann das Argument dafür, dass dieser Inhalt gleich $h \cdot f(c)$ für ein geeignetes c zwischen x und $x+h$ ist (was Antwort 2 als richtig erweist).

4.2.4 Anforderung: Fragen mit inhärentem Argumentationspotential

Um die erwünschte Wirkung der Peer Instruction zu erreichen, ist es notwendig, durch die ConcepTests starke Argumentationsanlässe zu schaffen. Es werden also „argumentationshaltige” Fragen gebraucht. Wenig geeignet sind dafür Fragen der folgenden Art:

Welche der folgenden Bedingungen drückt aus, dass die Folge (a_n) eine Cauchy-Folge ist?

(1) $\forall \varepsilon > 0\; \forall N \in \mathbb{N}\; \exists n,m \geq N : |a_n - a_m| < \varepsilon$
(2) $\forall \varepsilon > 0\; \exists N \in \mathbb{N}\; \forall n,m \geq N : |a_n - a_m| < \varepsilon$
(3) $\exists \varepsilon > 0\; \forall N \in \mathbb{N}\; \exists n,m \geq N : |a_n - a_m| < \varepsilon$
(4) $\exists \varepsilon > 0\; \exists N \in \mathbb{N}\; \forall n,m \geq N : |a_n - a_m| < \varepsilon$

Zweifellos ließe sich prinzipiell eine fruchtbare Diskussion auf Grundlage dieser Frage aufbauen: Hierbei würde man etwa erörtern, wie die einzelnen Formulierungen in natürlicher Sprache ausgedrückt werden können, welche inhaltliche Bedeutung sie haben, welche Implikationen es eventuell zwischen ihnen gibt. Auch würde man Beispiele von Folgen suchen, die die jeweiligen Bedingungen erfüllen oder nicht erfüllen. All diese Überlegungen sind jedoch in der Frage nicht zwingend angelegt: Die gestellte Frage lässt sich einfach dadurch beantworten, dass man die richtige Formulierung textuell memoriert hat oder im Lehrbuch nachschlägt, ohne darüber hinaus gehende Überlegungen anzustellen. Für eine fruchtbare Diskussion wäre ein Moderator erforderlich, der darauf hinwirkt, dass die relevanten Aspekte tatsächlich herausgearbeitet werden. Das Prinzip der Peer Instruction liegt aber gerade darin, dass die Fragen zu intensiver inhaltlicher Verarbeitung anregen, *ohne* dass ein Diskussionsleiter einwirkt. Entscheidend ist daher, dass das Anregungspotential der Fragestellung *inhärent* ist. Ein besser geeigneter ConcepTest zum obigen Thema der Cauchy-Folgen ist daher:

Jemand betrachtet die durch $a_n := \sqrt{n}$ definierte Folge (a_n). Er beweist:

(i) $a_n \to \infty$
(ii) $|a_{n+1} - a_n| \to 0$

Er sagt: „Die Folge ist wegen (i) nicht konvergent und wegen (ii) eine Cauchy-Folge." und er sieht darin einen Widerspruch zu seinem Wissen aus der Analysis-Vorlesung. Wodurch wird dieser Widerspruch aufgelöst?

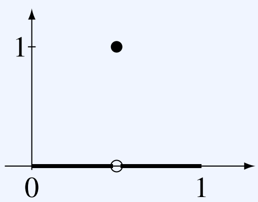
(1) Die Folge ist konvergent.
(2) Die Folge ist bestimmt divergent.
(3) Die Folge ist keine Cauchy-Folge.
(4) Es gibt Cauchy-Folgen, die nicht konvergent sind.

Hier lässt sich die richtige Antwort nicht dadurch finden, dass lediglich der Wortlaut von Definitionen und Sätzen memoriert wurde oder nachgeschlagen wird. Vielmehr ist es notwendig, diese Wissenselemente auf das konkrete Beispiel zu beziehen und zueinander in Beziehung zu setzen, um den durch die Frage provozierten kognitiven Konflikt zu beseitigen. Erst so wird sichtbar, welchen Abdruck die *Concept Definition* im *Concept Image* hinterlassen hat.

4.2.5 Argumentation als Fragegegenstand

Im Pilotdurchgang zeigte sich, dass sich die Studierenden bei manchen ConcepTests nach Phase 2 in großer Mehrheit für die richtige Antwortalternative entschieden haben, dies jedoch „aus den falschen Gründen", d.h. mit einer fehlerhaften Argumentation geschah. Solche Fehlargumentationen stellen ein Gefahrenpotential dar, da sie möglicherweise vom Tutor unbemerkt bleiben und daher schlimmstenfalls verfestigt werden. Vermeiden lässt sich dies, wenn man in diesen Fällen ConcepTests konstruiert, die eine explizite Argumentation in Form einer „weil"-Formulierung enthalten. Ein Beispiel:

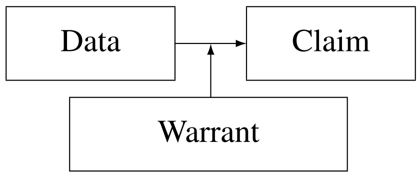
Die Funktion



(1) ist nicht integrierbar, weil sie weder stetig noch monoton ist
(2) ist integrierbar, weil alle Untersummen und alle Obersummen gleich 0 sind.
(3) ist nicht integrierbar, weil es Zerlegungen gibt, bei denen die Untersumme gleich 0 und die Obersumme gleich 1 ist.
(4) ist integrierbar, weil es für jedes $k \in \mathbb{N}$ Zerlegungen Z_k gibt, bei denen die Untersumme gleich 0 und die Obersumme gleich $1/k$ ist

Durch die Art der Fragestellung wird die Diskussion in Phase 2 der Peer Instruction direkt auf das Abwägen zwischen falschen und richtigen Argumentationen gelenkt. Dabei lässt sich Eric Mazurs Empfehlung (M3) berücksichtigen, dass Distraktoren nach Möglichkeit aus bekannten Fehlern (hier also aus bekannten Fehlschlüssen) konstruiert werden sollen.

Bei ConcepTests dieser Bauart ist zu beachten, dass Antwortmöglichkeiten der Form „Es gilt B, weil A gilt” auf den ersten Blick wie Implikationen $A \Rightarrow B$ aussehen. Wenn man sie allerdings als (materiale) Implikationen auffassen würde, dann wäre nicht nur Antwort (4) richtig, sondern alle Antworten mit Ausnahme von (1), denn $A \Rightarrow B$ ist wahr, wenn A falsch oder B wahr ist. Jedoch drückt die Konjunktion „weil” in der Fragestellung keine materiale Implikation aus, sondern es ist hier eine Interpretation im Sinne des Argumentationsmodells von Toulmin (1958) angebracht: Wie bei Weber und Alcock (2005) ausgeführt ist, geht es beim Verstehen der Schlüsse in einer Argumentation „Es gilt B, weil A gilt” nicht nur darum, ob die materiale Implikation $A \Rightarrow B$ wahr ist, sondern darum, ob eine Legitimation (ein *Garant*) dafür vorliegt, aus einer als wahr bekannten Aussage A auf eine behauptete Aussage B zu schließen. Im Toulminschen Modell (in seiner einfachsten Form) wird ein solches Argument durch ein Diagramm



dargestellt, in dem *Data* und *Claim* den Aussagen A bzw. B entsprechen und *Warrant* eine Aussage ist, die als legitimierende Brücke zwischen *Data* und *Claim* fungiert (z.B. ein Satz, der den Schluss von A auf B ermöglicht). Ein korrektes Argument im Toulmin-Modell muss von einer wahren Aussage als *Data* ausgehen und es muss einen Garanten geben, der den Schluss von *Data* zu *Claim* legitimiert. Der Garant bleibt in vielen Fällen (so auch im oben gezeigten ConcepTest) implizit. Wendet man Toulmins Modell auf den obigen ConcepTest an, so wird klar, dass nur Antwortalternative (4) eine korrekte Argumentation darstellt, während in (2) die *Data*-Aussage falsch ist und in (3) zwar die *Data*-Aussage wahr ist, es aber keinen Garanten gibt, der den Schluss von A auf B ermöglicht.

Eine Erweiterung von ConcepTests mit „weil”-Formulierung sind solche mit „weil-obwohl”-Formulierungen – auch diese können dazu eingesetzt werden, Argumentationen bewerten zu lassen:

> Sei $I := [0, \pi]$. Für die Funktion $f : I \to I$, $x \mapsto \sin(x)$ gilt
>
> $$|f'(x)| = |\cos(x)| \leq 1 \qquad \text{für alle } x \in I$$
>
> (1) f hat einen Fixpunkt, weil f eine Kontraktion ist.
> (2) f hat einen Fixpunkt, obwohl f keine Kontraktion ist.
> (3) f hat keinen Fixpunkt, weil f keine Kontraktion ist.
> (4) f hat keinen Fixpunkt, obwohl f keine Kontraktion ist.

5 Ergebnisse

Der Pilotdurchgang im Sommersemester 2014 und Wintersemester 2014/15 ermöglichte es, die folgenden Fragen zu beantworten:

1. Lassen sich zu den zentralen Konzepten in allen Themengebieten der Analysis I und

Analysis II ConcepTests konstruieren, die den in Abschnitt 4 formulierten Anforderungen genügen?
2. Wie beurteilen Studierende die Methode?
3. Wie beurteilen die Tutoren die Methode? Wie gut kommen sie bei der Durchführung der Methode zurecht?

5.1 Konstruktion von ConcepTests

Es wurden ConcepTests zu folgenden Themenbereichen eingesetzt:

- *Analysis I* (65 ConcepTests, davon 20 aus dem GoodQuestionsProjekt Terrel und Connelly (2005) und 45 neu entwickelt): Gleichungen und Ungleichungen – Maximum, Minimum, Infimum, Supremum, Vollständigkeit – Folgen und Konvergenz von Folgen, Cauchy-Folgen – Reihen und Konvergenz von Reihen – Grenzwerte und Stetigkeit – Sätze über stetige Funktionen – Differenzierbarkeit, Ableitung, Tangenten – Mittelwertsatz, Extrema – Funktionenfolgen, Supremumsnorm, Konvergenz – Satz von Taylor
- *Analysis II:* (89 ConcepTests neu entwickelt): – Obersummen, Untersummen, Riemannsummen, Integral – Stammfunktionen, Integralfunktionen, Hauptsatz, uneigentliche Integrale – metrische Räume, Grenzwerte, Stetigkeit, Vollständigkeit – Kontraktionen und Fixpunkte – Kompaktheit, zusammenhängende Mengen – parametrisierte Kurven – partielle und totale Differenzierbarkeit – Kettenregel, Richtungsableitungen, Niveaumengen – Diffeomorphismen, lokaler Umkehrsatz, Satz über implizite Funktionen – Extremwerte von Funktionen, Extrema unter Nebenbedingungen, Untermannigfaltigkeiten – Differentialgleichungen

Damit konnten alle zentralen Konzepte der beiden Vorlesungen durch ConcepTests repräsentiert werden (M1). Es traten keine prinzipiellen Schwierigkeiten auf, die notwendige konzeptuelle Eindringtiefe zu erreichen (M2) und auf Basis der bisherigen Lehrerfahrung plausible Distraktoren zu konstruieren, die sich an bereits beobachteten Fehlern orientieren (M3). Der kontinuierliche Vortest der ConcepTests durch einen Mitarbeiter erwies sich als wertvoll, um Formulierungen zu erarbeiten, die möglichst wenig Anlass zu Missverständnissen gaben (M4). Eine der Erwartungen zu Projektbeginn war, dass das Erreichen des 35–70-Prozent-Intervalls (M5) eine echte Herausforderung sein würde. Daher wurde ein wöchentliches standardisiertes Feedback durch die Tutoren eingeholt, die die ConcepTests nach der Durchführung in zwei Kategorien einschätzten:

- Zum einen wurde der *Schwierigkeitsgrad* des ConcepTests bewertet (anhand der ersten Votings in der Gruppe). Dazu wurde eine 5-stufige Likertskala verwendet: viel zu leicht – etwas zu leicht – gut passend – etwas zu schwer – viel zu schwer
- Zum anderen wurde der *Ertrag* des ConcepTests bewertet (anhand der vom Tutor beobachteten Phase der Peer Discussion). Dazu wurde eine 4-stufige Likertskala verwendet: völlig unproduktiv – eher unproduktiv – eher produktiv – sehr produktiv

Solche wöchentlichen Rückmeldungen liefern wertvolle Informationen: Zum einen geben sie kontinuierlich Auskunft über den Leistungsstand der Studierenden und ermöglichen eine generelle Einschätzung des bei den ConcepTests günstigen Anspruchsniveaus. Zum anderen – und dies war die entscheidende Beobachtung – wurde deutlich, dass es kein für alle Übungsgruppen einheitliches optimales Anspruchsniveau gibt, da sich bei den Leistungsniveau der Gruppen erstaunlich große Unterschiede zeigten: ConcepTests, die in manchen Gruppen den optimalen Schwierigkeitsgrad hatten, erwiesen sich in anderen Gruppen als deutlich zu leicht und in wieder anderen Gruppen als deutlich zu schwer (und schon aus diesem Grund in letzteren Gruppen als unproduktiv). Die Unterschiede waren größer als man bei einer zufälligen Verteilung der Studierenden naiv erwarten würde. Erklären könnte man dies dadurch, dass die Zusammensetzung der Übungsgruppen vermutlich nicht als zufällig betrachtet werden kann, da bereits vorab gebildete Lerngruppen bei der Einteilung in Gruppen oft zusammenbleiben. Als Ansatz, um mit dieser Heterogenität umzugehen, wurde die Anzahl der bereitgestellten ConcepTests im Laufe des Projekts deutlich erhöht, damit die Tutoren aus den für die jeweilige Woche inhaltlich passenden ConcepTests eine Auswahl treffen konnten, die sie für ihre Gruppe als im Schwierigkeitsgrad passend einschätzten. Man erkennt dies an der deutlich größeren Anzahl ConcepTests, die für die Analysis II konstruiert wurden (89 im Vergleich zu 69 zur Analysis I).

Aus informellen Rückmeldungen durch Tutoren und Studierende wurde ferner deutlich, wie wichtig die Konstruktion der Distraktoren ist: Immer wieder wurde rückgemeldet, dass es äußerst nützlich sei, dass die ConcepTests auf typische Fehler aufmerksam machen und die Peer Discussion zutage bringt, worin der Fehler genau besteht und welche Fehlvorstellung zugrunde liegt.

5.2 Beurteilung durch Studierende

Bei den Teilnehmern an den Übungsgruppen wurde im Pilotdurchgang im Sommersemester 2014 und im Wintersemester 2014/15 jeweils am Semesterende eine Befragung durchgeführt, um zu ermitteln, wie sie die Methode der Peer Instruction einschätzen (zum einen in affektiver Hinsicht, zum anderen hinsichtlich der wahrgenommenen Wirkung). Dazu waren die folgenden Aussagen jeweils auf einer 5-stufigen Likertskala zu bewerten („trifft nicht zu – trifft eher nicht zu – unentschieden – trifft eher zu – trifft zu”):

1. *Ich habe an den Peer-Instruction-Phasen gerne teilgenommen.*
2. *Bei den Diskussionen in der Peer-Instruction-Phase habe ich mich aktiv beteiligt.*
3. *Rückblickend auf das ganze Semester fand ich den Einsatz von Peer-Instruction für mich produktiv (hinsichtlich des Lerneffekts, Aha-Erlebnissen, . . .):*
4. *Einen Teil der Präsenzzeit im Tutorium auf Peer Instruction zu verwenden, halte ich für einen sinnvollen Einsatz der Zeit.*
5. *Ich bin dafür, dass auch im nächsten Semester Peer Instruction in den Tutorien eingesetzt wird.*
6. *Über bestimmte Fragen/Aspekte habe ich später noch einmal in meiner Arbeitsgruppe/mit meinem Zettelpartner diskutiert.*

Die Ergebnisse der Befragung aus der Analysis I sind in Abb. 2 dargestellt. (Die Ergebnisse aus der gleichlautenden Befragung in der Analysis II sind sehr ähnlich.) Der Befund ist sehr ermutigend: Er zeigt, dass die Studierenden in großer Mehrheit sehr positiv auf die Methode reagieren, sowohl hinsichtlich ihrer Beteiligung als auch hinsichtlich des Nutzens, den sie in Relation zur eingesetzten Zeit wahrnehmen. Aufschlussreich ist das Ergebnis zu Frage 6: Die Methode der Peer Instruction beinhaltet keine Aufforderung, über Aspekte der ConcepTests außerhalb der Übung weiter zu diskutieren. Und tatsächlich zeigt die Befragung, dass es zu solchen zusätzlichen Wirkungen kaum kommt (auch wenn man sich dies als Dozent wünschen könnte). Gleichzeitig zeigt sich hier, dass die

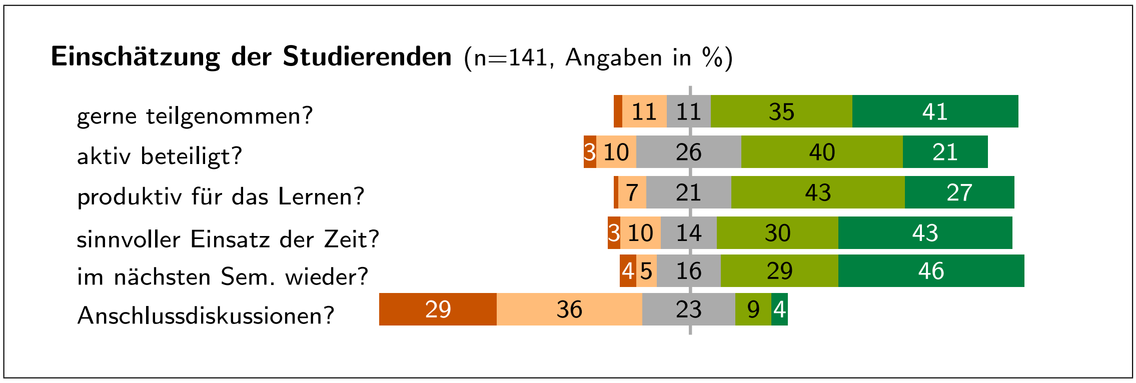


Abb. 2: Ergebnisse der Befragung der Studierenden im Pilotdurchgang in der Veranstaltung Analysis I (Sommersemester 2014). Die Anteile der Balken repräsentieren die Antworten auf der folgenden 5-stufigen Likertskala: *trifft nicht zu – trifft eher nicht zu – unentschieden – trifft eher zu – trifft zu.*

Studierenden bei der Befragung nicht nur sozial Erwünschtes antworten – insofern gibt Frage 6 den Antworten auf die Fragen 1–5 mehr Glaubwürdigkeit.

In Freitextkommentaren, die die Studierenden bei der Befragung noch anbringen konnten, wurde deutlich, aus welchen Gründen die Studierenden die Peer Instruction schätzen:

- *Die Peer-Instruction [. . .] helfen, nochmal gezielt in der VL zu blättern und zu überprüfen, ob das eigene Verständnis wichtiger Sätze richtig ist. Mir hat das gut gefallen.*
- *Ich finde diese Methode besonders als Auffrischung zu Anfang des Tutoriums sehr gut. Danach ist man wieder richtig im Stoff.*
- *Teilweise haben die Fragen die Vorstellung, die ich von dem Vorlesungsstoff hatte, verstärkt und auch mal richtig gerückt. Das war sehr gut.*
- *Insbesondere bei Fragen, die intuitiv klar waren, eine korrekte Begründung aber schwer fiel, hat mir die Methode geholfen.*
- *PI ist nützlich, um noch einmal das Verständnis d. Stoffes zu überprüfen, dabei erkennt man auch seine eigenen Probleme und Schwächen.*
- *Besonders zur Einstimmung sehr hilfreich. Allgemein sehr gewinnbringend.*
- *Die Peer-Instruction hat mein Verständnis für das VL-Skript erhöht und des Öfteren noch weitere Blickwinkel geschaffen.*

Was die Studierenden hier zum Ausdruck bringen, trifft sehr gut die Intentionen der Peer Instruction. Es gibt auf der anderen Seite auch Freitextkommentare, in denen sich zeigt, dass manche Studierende die Methode missverstanden haben:

- *Könnte man auch auf den Übungszetteln machen, bloß ohne Punktegewinn bzw. Punkteverlust. Dann hätte man im Tutorium mehr Zeit für das Besprechen der „hauptsächlichen" Übungsaufgaben.*
- *[. . .] oder dazu sollen Musterlösungen für uns geben (Musterlösungen für Peer-Instruction, damit könnten wir es nochmal schauen)*

Diese Studierenden haben nicht erkannt, dass es die *Diskussionsphase* ist, die den Kern der Methode bildet. (Selbstverständlich war dies bei der Einführung der Methode erläutert worden.) Stattdessen scheinen sie die ConcepTests irrtümlich als weiteren *Lern-* oder *Prüfungsstoff* aufzufassen, für den sie sich eine „Musterlösung" wünschen. Dass ablehnende Einschätzungen vorkommen, bestätigt Ergebnisse von Crouch und Mazur (2001),

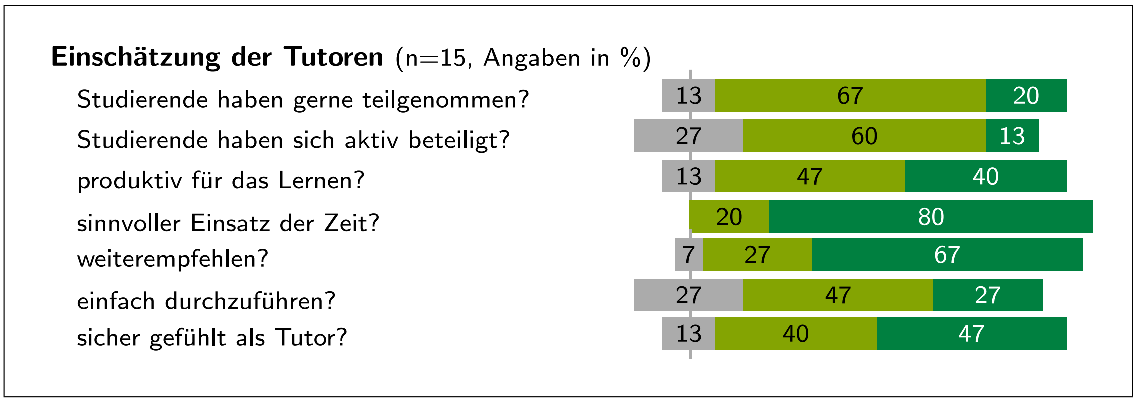


Abb. 3: Ergebnisse der Befragung der Tutoren im Pilotdurchgang in der Veranstaltung Analysis I (Sommersemester 2014). Die Anteile der Balken repräsentieren die Antworten auf der folgenden 5-stufigen Likertskala: *trifft nicht zu – trifft eher nicht zu – unentschieden – trifft eher zu – trifft zu*.

die ebenfalls davon berichten, dass in der Regel einige Studierende negativ auf die Methode reagieren. Einschränkend weisen sie darauf hin, dass man zwischen „student evaluation" und „student learning" unterscheiden müsse – wenn Studierende Vorbehalte haben, dann bedeutet das nicht, dass die Methode bei ihnen keinerlei Ertrag hat. Neben einem Missverstehen der Methode an sich könnte ein Grund für negative Einschätzungen auch darin liegen, dass introvertierte Studierende die Unterrichtsmethode als eher unangenehm empfinden, weil sie stark auf Diskurs ausgelegt ist: Insbesondere die leistungsschwachen Studierenden in dieser Gruppe mögen es als unangenehm empfinden, dass in der Peer Discussion ihr Leistungsstand für die Kommilitonen sichtbar wird.

5.3 Beurteilung durch Tutoren

Neben der Einschätzung durch Studierende ist es wichtig zu wissen, wie die Tutoren die Methode beurteilen, da sie als Lehrpersonen an einer Schlüsselstelle stehen. Dazu dienten Befragungen jeweils am Ende des Semesters im Pilotdurchgang, in der die Tutoren folgende Aussagen bewerteten:

1. *Meine Gruppe hat – meinem Eindruck nach – an den Peer-Instruction-Phasen gerne teilgenommen.*
2. *Bei den Diskussionen in der Peer-Instruction-Phase haben sich die Studierenden aktiv beteiligt.*
3. *Rückblickend auf das ganze Semester fand ich den Einsatz von Peer-Instruction produktiv (hinsichtlich des Lerneffekts, Aktivierung und Aha-Erlebnissen der Studierenden, ...).*
4. *Einen Teil der Präsenzzeit im Tutorium auf Peer-Instruction zu verwenden, halte ich für einen sinnvollen Einsatz der Zeit.*
5. *Ich würde das Verfahren weiterempfehlen:*
6. *Ich fand das Verfahren einfach durchzuführen – verglichen mit Ihren anderen Aufgaben im Tutorium (Korrekturen, Besprechung von Hausübungen, Beantworten von Fragen).*
7. *Bei der Durchführung des Verfahrens fühlte ich mich als Tutor sicher.*

Das Ergebnis (siehe Abb. 3) zeigt, dass die Tutoren den Einsatz der Methode als positiv und gewinnbringend empfunden haben. Dies ist ein wichtiges Ergebnis, da viele

Tutoren berichten, dass es für sie in traditionellen Übungen oft mühsam und frustrierend sein kann, Diskussionen in Gang zu bringen. Es ist daher durchaus von Bedeutung, Methoden zu finden, die den Tutoren ein befriedigendes Lehrerlebnis geben können und so dazu beitragen, dass diese sich auch weiterhin kraftvoll engagieren.

Weiter zeigt die Befragung, dass sich die Tutoren in der Anwendung der Methode sicher fühlen. Die Kurzschulung vor Beginn des Semesters (siehe Abschn. 3.2) erwies sich also als ausreichend, um die Tutoren in die Lage zu versetzen, die Methode erfolgreich durchzuführen. Dies bestätigt die Erwartung, dass Peer Instruction eine einfach anzuwendende Methode ist, da die konzeptionelle Arbeit in der *Konstruktion* der ConcepTests beim Dozenten liegt. Auch dies ist ein für die Praxis wichtiges Ergebnis, da der mögliche zeitliche Umfang von Tutorenschulungen – insbesondere auf bestimmte Methoden bezogen – begrenzt ist.

5.4 Fallstricke, Beachtenswertes und Perspektiven

Technische Aspekte. Für die praktische Durchführung der Methode ist zu klären, in welcher Form die ConcepTests den Studierenden vorgelegt werden. Im Pilotdurchgang konnte ein Teil der Übungsgruppen in Räumen durchgeführt werden, in denen eine Projektionsmöglichkeit zur Verfügung stand (Overheadprojektor, Beamer). Die ConcepTests konnten hier auf vom Dozenten vorbereiteten Folien gezeigt werden, was einen reibungslosen und zeitsparenden Ablauf ermöglicht. Diejenigen Tutoren, die ausschließlich Tafeln zur Verfügung hatten, waren deutlich im Nachteil, da die für das Anschreiben aufgewendete Zeit effektiver in den Gruppendiskussionen verwendet würde.

Heterogenität. In Abschn. 5.1 wurde bereits auf die Beobachtung hingewiesen, dass die Konstruktion von ConcepTests von geeignetem Schwierigkeitsgrad (M5) die Hauptherausforderung für den Dozenten ist. Als Lösungsansatz wurde im Projekt die Idee verfolgt, den Tutoren wöchentlich eine Auswahl an ConcepTests zur Verfügung zu stellen, aus denen sie jeweils nach ihrer Einschätzung passende Items entnehmen. Die konstruierten ConcepTests müssen dazu ein genügend breites Spektrum im Anspruch abdecken. Eine weitere Möglichkeit, die der Verfasser derzeit erprobt, besteht darin, kleine Serien von ConcepTests zu entwerfen, die denselben Gegenstand in aufsteigendem Schwierigkeitsgrad bearbeiten. Es kann dann jeweils der erste Teil einer Peer-Instruction-Runde (Phase 1 und Voting 1) durchgeführt werden, bis man innerhalb der Serie zum angemessenen Schwierigkeitsgrad gelangt, der eine produktive Durchführung von Phase 2 ermöglicht. Da Phase 1 und Voting 1 nicht viel Zeit in Anspruch nehmen, gelangt man hierdurch relativ rasch zu einem passenden Niveau. Einschränkend muss man sagen, dass noch nicht geklärt ist, ob die Konstruktion von ConcepTest-Serien durchgängig realisierbar ist (hinsichtlich des Aufwands, aber auch in prinzipieller Hinsicht).

Das Cramming-Problem. Cramming, d.h. die Konzentration der Lernzeiten auf die Phase kurz vor den Prüfungsterminen, ist ein bekanntes Problem im schulischen, aber auch im universitären Lernen. Dessen negative Effekte auf die Lern- und Prüfungsergebnisse sind gut belegt (siehe z.B. Schulmeister 2015) und machen sich auch beim Einsatz von Peer Instruction bemerkbar: So berichten die Tutoren von Studierenden, die unvorbereitet sind und die die Voraussetzungen (z.B. Concept Definition, Theorem Statement) nicht haben, um in der Diskussionsphase produktiv mitarbeiten zu können. Wenn Studierende ihr Lernen auf diese Weise in die Zukunft verlagern und damit Lerngelegenheiten verspielen, dann empfinden Tutoren (und in der Folge auch der Dozent) dies als frustrierend. Allerdings ist zu bedenken, dass man dies nicht der Methode der Peer Instruction anlasten sollte: In einer Übung, in der Peer Instruction nicht zum Einsatz kommt, wären diese Studierenden vermutlich ebenso passiv geblieben und hätten sich auf das Abschreiben von Musterlösungen beschränkt, weil ihnen ihr Lernstand nicht viel anderes ermöglicht – es wäre aber vermutlich weniger aufgefallen.

Peer Instruction jenseits der Studieneingangsphase? Die Erprobung von Peer Instruction, die hier vorgestellt wurde, bezieht sich auf ihren Einsatz in den Vorlesungen Analysis I und II, liegt also in der Studieneingangsphase. Es ist eine natürliche Frage, ob die Methode auch in Lehrveranstaltungen höherer Semester sinnvoll und erfolgversprechend ist. Zunächst lässt sich sagen, dass die grundsätzlichen Ziele, die auf eine tiefere Durchdringung der Vorlesungsinhalte gerichtet sind, nicht nur in der Studieneingangsphase gelten. Aus dieser Sicht spricht nichts dagegen, auch beim Einsatz in anderen Lehrveranstaltungen positive Wirkungen zu erwarten. Demgegenüber warf ein erfahrener Tutor des Pilotdurchgangs die Frage auf, ob man durch Verwendung von Peer Instruction in Veranstaltungen jenseits der Studieneingangsphase den Studierenden möglicherweise die Verantwortung für das eigene Lernen zu sehr abnimmt und dadurch deren Entwicklung behindert. Derzeit lässt sich dies nicht beantworten. Offen ist auch: Haben die in der Peer Instruction angestoßenen Prozesse metakognitive Wirkung, so dass Studierende ihr Lernen künftig selbständig effektiver gestalten und auch ohne den Einsatz von Peer Instruction aktiver lernen?

Danksagung. Ich danke meinem Mitarbeiter Carsten Bornträger für seine wertvolle Arbeit beim Vortest der ConcepTests. Die Tutoren im Pilotdurchgang haben die Methode der Peer Instruction mit großem Engagement eingesetzt; ihnen verdanke ich wertvolle Rückmeldungen, die sehr zur Weiterentwicklung beigetragen haben.

Literatur

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

Bauer, Th. (2015). Übungsgelegenheiten im Mathematikstudium – Erproben neuer Konzepte. In: Schelhowe, H., Schaumburg, M., Jasper, J. (Eds.), Teaching is Touching the Future. Academic teaching within and across disciplines (pp. 139-139). Bielefeld: Webler.

Bauersfeld, H. (2000). Radikaler Konstruktivismus, Interaktionismus und Mathematikunterricht. In: E. Begemann (Ed.), Lernen verstehen – Verstehen lernen (pp. 117-144). Frankfurt/Main: Peter Lang.

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

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. Topics in Cognitive Science, 1(1), 73-105.

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.

Fonseca, B. A., Chi, M. T. (2011). Instruction based on self-explanation. In: R.E. Mayer, P.A. Alexander (eds.), Handbook of research on learning and instruction (pp. 296-321). New York, London: Routledge.

Freudenthal, H. (1983). Didactical phenomenology of mathematical structures. Dordrecht: Reidel.

Hefendehl-Hebeker, L. (2005): Perspektiven für einen künftigen Mathematikunterricht. In: Bayrhuber, H. /Ralle, B. /Reiss, K. /Schön, H. /Vollmer, H. (Hrsg.): Konsequenzen aus PISA - Perspektiven der Fachdidaktiken. Innsbruck: Studienverlag.

Lehn, M. (o. D.). Wie bearbeitet man ein Übungsblatt? https://www.agtz.mathematik.uni-mainz.de/wie-bearbeitet-man-ein-uebungsblatt-von-prof-dr-manfred-lehn/ Zugegriffen: 19. August 2017

Leuders, T., Holzäpfel, L. (2011). Kognitive Aktivierung im Mathematikunterricht. Unterrichtswissenschaft, 39(3), 213-230.

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.

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

Renkl, A. (2011). Aktives Lernen: Von sinnvollen und weniger sinnvollen theoretischen Perspektiven zu einem schillernden Konstrukt. Unterrichtswissenschaft, 39(3), 197-212.

Schulmeister, R. (2015). Abwesenheit von Lehrveranstaltungen – Ein nur scheinbar triviales Problem. Hamburg. Veröffentlicht im Rahmen der Campus Innovation. `https://www.campus-innovation.de/fileadmin/dokumente/Schulmeister_Anwesenheit__Abwesenheit__2_.pdf`. Zugegriffen: 21.05.2018

Selden, J., Selden, A. (1995). Unpacking the logic of mathematical statements. Educational Studies in Mathematics, 29(2), 123-151.

Tall, D., Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. Educational studies in mathematics, 12(2), 151-169.

Terrell, M.S., Connelly, R. (2005). GoodQuestions Project. Department of Mathematics, Cornell University. http://www.math.cornell.edu/ GoodQuestions. Zugegriffen: 21.05.2018

Toulmin, S. E. (1958). The uses of argument. Cambridge: Cambridge university press.

Vollrath, H.-J. (1984). Methodik des Begriffslehrens. Stuttgart: Klett.

Weber, K., Alcock, L. (2005). Using warranted implications to understand and validate proofs. For the Learning of Mathematics, 25(1), 34-51.

Weigand, H. G. (2015). Begriffsbildung. In: Bruder, R., Hefendehl-Hebeker, L., Schmidt-Thieme, B., Weigand, H. G. (Eds.), Handbuch der Mathematikdidaktik (pp. 255-278). Berlin Heidelberg: Springer.

Wentzel, K. R., Watkins, D. E. (2011). Instruction based on peer interactions. In: R.E. Mayer, P.A. Alexander (eds.), Handbook of research on learning and instruction (pp. 322-343). New York, London: Routledge.

Winter, H. (1983). Entfaltung begrifflichen Denkens. Journal für Mathematik-Didaktik, 4(3), 175-204.