
BMARKS-LIT
**„Wie wirksam sind die *bettermarks-Funktionen*?
Eine systemgeleitete Literaturstudie aktueller
Forschungsbefunde“**

Ergebnisse und Empfehlungen

Dr. Benjamin Paaßen, Dr. Berit Blanc

Ziel und Aufbau der Studie

Studienziel:

Sichtung international publizierter, empirischer Evidenz für ausgewählte Funktionen und Komponenten von adaptiven Mathe-Tutoringsystemen, um ...

1. fundierte Empirie für die Wirksamkeit des *bettermarks*-Systems und -Einsatzes zusammenzustellen und
2. evidenzbasierte Empfehlungen für die weitere Ausgestaltung und den Einsatz von *bettermarks* abzuleiten.



Ziel und Aufbau der Studie

Studienaufbau:

1. Systemanalyse und Recherchekonzept: Workshop, Gespräche
2. Literaturrecherche und -analyse zu Kerneigenschaften und Rahmenbedingungen von *bettermarks*
3. a) Ausführlicher Bericht zum internen Gebrauch
b) Kurzbericht in Form eines Tagungsbeitrags

Kerneigenschaften, Rahmenbedingungen und Forschungsfragen

Forschungsfragen zu zentralen Systemeigenschaften

A) Vollständigkeit der Aufgaben & Inhalte

1. Ist eine feinere Granularität von Lernzielen und Aufgaben lernwirksam?
2. Ist eine höhere Abdeckung des Curriculums lernwirksam?

B) Intelligente Interaktionswerkzeuge

1. Sind reichhaltige Eingabewerkzeuge (z.B. Zeichenfunktionen) lernwirksam?
2. Sind variable Aufgabentypen (z.B. verschiedene Darstellungs- und Interaktionsformen) lernwirksam?

C) Mikro-Adaptivität

1. Ist fehlerbasiertes Feedback (über reines richtig/falsch-Feedback hinaus) lernwirksam?
2. Ist die Möglichkeit zur Korrektur und Reflexion (z.B. nach dem ersten Fehlversuch) lernwirksam?
3. Ist das Anzeigen von Lösungswegen (nach dem zweiten Fehlschlag) lernwirksam?
4. Sind zusätzliche Ressourcen (wie vorgerechnete Beispiele, Glossare, et al.) lernwirksam?

D) Makro-Adaptivität

1. Ist das Identifizieren von Wissenslücken sowie die gezielte Vergabe von Übungen, um diese Lücken zu schließen, lernwirksam?



Kerneigenschaften, Rahmenbedingungen und Forschungsfragen

Forschungsfragen zu Rahmenbedingungen des Einsatzes

E) Einsatz im Klassenverbund und Abdeckung des Curriculum

1. Ist die Kombination von ITS mit Lehrkräften lernwirksam (im Vergleich zu Lehrkräften ohne ITS oder ITS ohne Einbindung von Lehrkräften)?
2. (=A3.) Ist eine höhere Abdeckung des Curriculums lernwirksam?

Recherche

Zweig 1: Metastudien

13 Metastudien

Zweig 2: Wettbewerber

	SCOPUS	google scholar
ASSISTMENTS	60	356
Cognitive Tutor	110	1430
ALEKS	64	1290
IXL	25	519
Calcularis	4	110
DreamBox	0	204
area9	3	124
MatheGym	0	1
studyly	0	0
Mathe digital	0	1
Diagnose und Fördern	0	2

Nur Wirkungsstudien, Entfernen von Duplikaten

	SCOPUS	google scholar
ASSISTMENTS	15	8
Cognitive Tutor	14	6
ALEKS	10	5
IXL	1	12
Calcularis	2	7
DreamBox	0	4
area9	0	0
MatheGym	0	0
studyly	0	0
Mathe digital	0	0
Diagnose und Fördern	0	0

87 Studien

Zweig 3: SCOPUS-Literaturrecherche

1383 Treffer

Ausschlusskriterien, 50 meist zitierte Studien

50 Studien

Entfernen von Duplikaten

39 Studien



Analyse

- Ergebnisvorstellung anhand der Kerneigenschaften und Antworten auf die Forschungsfragen
- Vorbemerkung 1: Wettbewerber im Überblick
- Vorbemerkung 2: Wirkungsforschung zu digitalen Lernumgebungen



Wettbewerber-Systeme

System	Wirksamkeit	Vollständigkeit der Inhalte	Intelligente Interaktionswerkzeuge	Mikro-Adaptivität	Makro-Adaptivität	Einsatz im Klassenverbund
ASSISTMENTS	++	++	+	+	+	++
Cognitive Tutor	++	+	++	++	++	++
Calcularis	+	+	++	+	++	0 (Spezialfall)
ALEKS	0	++	0	+	++	+
Dreambox	0	+	++	0	++	+
IXL	0	++	+	0	+	0

Analyse

- Ergebnisvorstellung anhand der Kerneigenschaften und Antworten auf die Forschungsfragen
- Vorbemerkung 1: Wettbewerber im Überblick
- Vorbemerkung 2: Wirkungsforschung zu digitalen Lernumgebungen

A) Vollständigkeit der Aufgaben & Inhalte

A1. Eine feinere Auflösung von Lernzielen und Aufgaben ist lernwirksamer

- 7 Befunde
- Theorie: Tiefgang der Aufgaben ist mathematik-didaktisch wichtig
- Feedback zu einzelnen Schritten („substep-based“) ist lernwirksamer
- Aufforderung, Einzelschritte zu reflektieren und zu erklären ist lernwirksam (allerdings weniger bearbeitete Aufgaben)
- Cognitive Tutor und Calcularis modellieren feine Aufgabenschritte und sind wirksam; ASSISTMENTS ist ohne diese feine Modellierung wirksam => kein eindeutiges Bild

A) Vollständigkeit der Aufgaben & Inhalte

A2. Eine höhere Abdeckung eines Curriculums ist nicht eindeutig lernwirksam

- 4 – uneinheitliche – Befunde
- Theorie: breitere Abdeckung als ein Qualitätsmerkmal
- Zwei Metastudien sprechen für breite Abdeckung, eine eher dagegen
- Analyse der Wettbewerber uneindeutig:
 - *ASSISTMENTS*, *ALEKS*, *DreamBox*, *IXL* haben breite Abdeckung aber uneinheitliche Wirksamkeit
 - *Cognitive Tutor* und *Calcularis* haben enge Abdeckung und sind wirksam
- Praktische Erwägung: Spannungsverhältnis zwischen Breite (Abdeckung) und Tiefe (Aufgabenqualität; feine Granularität)
- Rahmenbedingung: Praktische Einsetzbarkeit im Unterricht erfordert breite Abdeckung – unabhängig von direkter Wirksamkeit

B) Intelligente Interaktionswerkzeuge

B1. Intelligente bzw. reichhaltige Eingabewerkzeuge sind als Lernformat wirksamer als „einfache Formate“.

- 10 Befunde mit eingeschränkter Übertragbarkeit auf *bettermarks*, da andere (Lern-)Werkzeuge untersucht wurden
- Empirie: Andere Eingabetypen als multiple choice haben i.d.R. deutlich höhere Lernwirksamkeit: offene Eingaben, Texte, Diagramme zeichnen und auch Dynamisch Mathe-Tools
- Korrelation: Reichhaltige Eingabewerkzeuge v.a. bei Aufgaben im Rahmen von Projektlernen, Problemlösungen und Modellierung und Visualisierung – Aufgaben mit hoher Lernwirksamkeit
- Theorie: Gefahr kognitiver Überforderung und/oder Ablenkung durch zu viele Handlungsoptionen

B) Intelligente Interaktionswerkzeuge

B2. Unterschiedliche Darstellungs- und Interaktionsformen können bei sorgsamer Gestaltung lernwirksamer sein.

- 9 empirische Befunde (wiederum eingeschränkte Vergleichbarkeit)
- Instruktion effektiver, wenn Mix aus prozeduralem und deklarativem Wissen; 1) eigene Ideen **finden**, 2) Ideen mit anderen Ideen **vergleichen** und 3) nach Kriterien **sortieren** lassen
- einförmige Mischung aus Textaufgaben und Multiple Choice eher ungünstig
- 1 Befund mit leichter Gegenevidenz: Systeme, die mehrere Arten von Fertigkeiten adressieren, hatten schwächere Effekte
- Korrelation: v.a. wirkungsevidente Systeme *Calcularis* und *Cognitive Tutor* verwenden verschiedene Darstellungs- und Interaktionsformen

B) Intelligente Interaktionswerkzeuge

B3. Zusatzfrage: Wie lernwirksam sind Gamification-Elemente?

- 4 Befunde
- Theorie: Gamification-Elementen können a) Enthusiasmus erzeugen, b) Leistungsfeedback geben, c) Wertschätzung zeigen und d) Zielsetzungsverhalten fördern; Wettbewerbselemente können aber auch zu Versagensangst und Eifersucht führen
- Empirie: gemischte Lage; einige Studien ohne positive Effekte oder sogar schlechterem Abschneiden
- Insgesamt **gemischte Evidenz zu Gamification-Elementen**: Ablenkung, wenn Spiel im Vordergrund; Motivation, wenn es der Transparenz dient

C) Mikro-Adaptivität

C1. Fehlerbasiertes Feedback ist eher lernwirksam als einfaches richtig/falsch-Feedback.

- 10 Befunde
- größere Effektstärken für ITS mit adaptivem Feedback als für reine Übungssysteme
- Empirie spricht eher für die Wirksamkeit ausführlichen Feedbacks
- Theorie: bei zu viel elaboriertem „push“-Feedback droht Überforderung („cognitive load“); nur so viel Feedback, dass eigenständige Bearbeitung möglich wird (Scaffolding-Theorie); Usability sicherstellen

C) Mikro-Adaptivität

C2. Die Möglichkeit zur Korrektur und Reflexion eigener Antworten ist lernwirksam.

- 8 Befunde
- Lernwirksamkeit der Korrekturmöglichkeit nur indirekt belegt (*ASSISTMENTS*)
- Wirksamkeitsevidenz für Reflexionsaufforderungen (regulatives Feedback, meta-kognitives Feedback)
- Korrelation: Wirksame ITS wie *ASSISTMENTS*, *Cognitive Tutor* und *Calcularis* lassen Korrektur und Reflexion zu

C) Mikro-Adaptivität

C3. Das Anzeigen von Lösungswegen ist eher lernwirksam.

- 5 Befunde
- Theorie: „Hängenbleiben“ führt zu Frustration und korreliert mit schwächerem Lernerfolg => Evidenz für *bettermarks*-Verfahren, Lernende bei wiederholten Fehlern fortfahren zu lassen
- Empirie: worked examples (vorgerechnete Beispiele mit ausführlichen Erklärungen des Lösungswegs) sind wirksam, wenn sie studiert werden
- Anzeigen richtiger Lösungen ohne ausführlicheres Feedback wirkt nicht über richtig/falsch-Feedback hinaus

C) Mikro-Adaptivität

C4. Zusätzliche Ressourcen (wie vorgerechnete Beispiele, Anleitungen und Videos) sind lernwirksam.

- 11 Befunde
- Umfangreiche Evidenz spricht **für** zusätzliche Ressourcen („pull scaffolding“): insbesondere bei schwächeren Lernenden; Websites, „Skill Tagebuch“, zusätzliche System-Anleitungen, Videos, zusätzliche Erklärungen bei *bettermarks*
- Zu beachten: „Assistance Dilemma“, Scaffolding-Theorie (eigenständige Lösungserarbeitung) und optimale Gestaltung (Textlänge und -schwierigkeit, Themenvertrautheit) sind zu berücksichtigen
- Adaptive Bereitstellung, Usability und einfache Sprache sind empfohlen

D) Makro-Adaptivität

Makro-Adaptivität hat keinen eindeutigen Effekt.

- 9 Befunde
- Pädagogische Theorie spricht eindeutig für Makro-Adaptivität: mastery learning und zone of proximal development
- Wenig direkte Wirkungsstudien; teilweise positive, teilweise keine Effekte
- Wettbewerber: Differenziertes Bild
 - *Cognitive Tutor*: Mastery Learning & wirksam
 - *Calcularis*: Bayesian knowledge graphs & wirksam
 - *ASSISTMENTS*: rudimentäres mastery learning & wirksam (Mastery Learning selbst hat vermutlich keinen Effekt)
 - *ALEKS* & *Area9*: vor allem Assessment & gemischte/gar keine Evidenz
- Wissenslücken-Ansatz von *bettermarks* noch **nicht** evaluiert
- Qualitative Evidenz spricht für Makro-Adaptivität durch Lehrkräfte

E) Rahmenbedingung: Einsatz im Klassenverbund

E1. Der Einsatz von ITS mit Beteiligung einer Lehrkraft ist im Vergleich zu traditionellem, nicht-digitalen Unterricht eher lernwirksam, wenn die Lehrkraft ausreichend geschult ist.

- 17 Befunde
- ITS i.d.R. lernwirksamer als traditioneller Unterricht und etwas weniger wirksam als menschliches 1:1-Tutoring
- I.d.R. positive Auswirkungen auf die Unterrichtsgestaltung
- Bedingung: Begleitung durch geschulte bzw. erfahrene Lehrkraft und Einbindung der Lehrkräfte bei der Einführung des ITS
- *ASSISTMENTS* und *Cognitive Tutor* mit deutlicher Einbindung der Lehrkräfte und insgesamt großer Wirksamkeit; *Calcularis*: ebenfalls wirksam, aber zusätzliches, außerunterrichtliches Förderinstrument ohne direkte Einbindung der Lehrkraft.

E) Rahmenbedingung: Abdeckung des Curriculums

E2. (= A3.) Eine höhere Abdeckung eines Curriculums korreliert nicht eindeutig mit Lernwirksamkeit ist aber als Rahmenbedingung, um im deutschen Markt bestehen zu können, unabdingbar.

- Abdeckung des Curriculums ist Systemeigenschaft und Rahmenbedingung zugleich
- 4 – uneinheitliche – Befunde
- Fazit: breitere Abdeckung des Curriculums ist grundsätzlich zu bevorzugen, damit das System im gesamten Schuljahr einsetzbar bleibt, aber hohe Qualität sollte gewahrt bleiben

Empfehlungen zum Systemdesign

1. Den Fokus auf Mikro-Adaptivität beibehalten und ausbauen (Eigenschaft C)
2. Lernfortschritt automatisch modellieren aber Makro-Adaptivität in den Händen der Lernenden und Lehrenden belassen (Eigenschaft D)
3. Hohe Qualität und Granularität der Inhalte bewahren (Eigenschaft A)
4. Interaktionswerkzeuge beibehalten (Eigenschaft B)

Empfehlungen zu Rahmenbedingungen

1. Die Einbindung der Lehrkräfte verstärken (Kerneigenschaft E)

- Lehrende in Systembedienung schulen
- Lehrende mit Informationen zum Lernfortschritt der Klasse versorgen
- Unterrichts- und Lerninhalte vorschlagen aber Entscheidung bei Lehrkräften belassen

2. Inhalte für den curricularen Bedarf ausbauen (Eigenschaft A)

- Lehrende und Lernende müssen zu wesentlichen curricularen Themen des jeweiligen Bundeslands/Schule Inhalte finden
- Ressourcen sonst eher in hohe Qualität und Tiefe investieren als in Breite

Forschungsdesiderata und -potenziale

Wirkungsstudien und Forschungsfragen für die Weiterentwicklung von *bettermarks*

- Zum Schließen von Wissenslücken (Makro-Adaptivität bislang mit gemischten empirischen Befunden)
- Systematische Untersuchung der Wirksamkeit der *bettermarks*-Interaktionswerkzeuge (kognitive Überlastung?)

Wirkungsstudien zum *bettermarks*-Einsatz im landesweiten Vergleich

- Lernerfolg bei Hausaufgaben mit *bettermarks* vs. ohne *bettermarks* unter Kontrolle des Zeitaufwands
- Wirkung unterschiedlicher Nutzungsintensitäten und Einsatzszenarien von *bettermarks* über die Resultate in landesweiten Vergleichsarbeiten



Forschungsdesiderata und -potenziale

Gemeinsame Forschungsstrategie *bettermarks* – EdTec

- Nutzung und Ausbau der *bettermarks*-Infrastruktur für die Durchführung kontrollierter Studien, um die kontinuierliche, evidenzbasierte Weiterentwicklung von *bettermarks* und den wirksamsten Einsatzszenarien zu unterstützen.
- Zusammenarbeit mit EdTec Lab
- Anknüpfung an offene Forschungsfragen aus vorherigen *bettermarks*-Studien (Scharnagl et al. 2014) sowie hier aufgezeigten Forschungslücken und -fragen (z. B. motivationale Effekte, Moderatorvariablen...)



Kontakt

DFKI | Deutsches Forschungsinstitut für Künstliche Intelligenz GmbH
Educational Technology Lab unter der Leitung von Prof. Dr. Niels Pinkwart

Dr. Benjamin Paaßen
benjamin.paassen@dfki.de

Dr. Berit Blanc
berit.blanc@dfki.de