Individuelle Förderung von Rechenfertigkeiten in der Grundschule durch ein computergestütztes Übungsprogramm

Wolfgang Schoppek & Christel Laue, Universität Bayreuth

Kolumnentitel:

Individuelle Förderung von Rechenfertigkeiten

Englischer Titel:

Individualized computer assisted training of basic arithmetic skills in elementary school

Summary:

We present our computer assisted training program „Merlin's Rechenmühle", which was developed for individualized practice of arithmetic skills in elementary school. The software connects to a database with more than 2300 problems of different types and difficulties. The size of the database and the facility of computer software to give immediate feedback permit individualization of practice in a group situation. We present the algorithms for predicting problem difficulties and for problem selection. In two studies, we showed the effectiveness of these algorithms. Results and experience from these studies will be exploited in a new version of the software, which will automize the time-consuming task of individual problem selection.

Keywords: arithmetic, skill acquisition, individualization, computer assisted instruction

Zusammenfassung:

Zur individuellen Förderung von Rechenfertigkeiten in der Grundschule entwickelten wir das computergestützte Übungsprogramm, „Merlins Rechenmühle", das auf eine Datenbank von über 2300 Aufgaben verschiedener Typen und Schwierigkeiten zurückgreift. Die Größe des Aufgabenpools und die Möglichkeiten des Computers zu unmittelbarer Rückmeldung ermöglichen die Individualisierung von Übung in einer Gruppensituation. Vorgestellt werden die Algorithmen zur Vorhersage der Aufgabenschwierigkeit und zur Aufgabenauswahl, sowie zwei empirische Studien, in denen die Wirksamkeit der Algorithmen empirisch nachgewiesen wurde. Die Erfahrungen aus diesen Studien sollen in die Entwicklung einer neuen Programm-version fließen, bei der die sehr zeitaufwändige individuelle Aufgabenauswahl automatisiert wird.

Schlüsselbegriffe: Rechnen, Fertigkeitserwerb, Individualisierung, computergestützte Instruktion

In den 1990er Jahren kam es in den USA zu einer heftigen Diskussion zwischen dem National Council of Teachers of Mathematics (NCTM), Praktikern und Wissenschaftlern, in der um die richtigen Inhalte und Lehrmethoden des Mathematik-Unterrichtes gestritten wurde („Math Wars“, Baroody, 2003). Die Extrempositionen in diesem Streit werden markiert durch den „drill & practice“-Ansatz mit dem Schwerpunkt auf dem auswendig Lernen von arithmetischen Fakten einerseits, und durch den konstruktivistischen Problemlöse-Ansatz andererseits, der die aktive Auseinandersetzung mit mathematischen Problemen in den Mittelpunkt stellt. Der konzeptuelle Ansatz mit einer Betonung der Vermittlung mathematischer Prinzipien und der entdeckende Ansatz nehmen dabei Zwischenpositionen ein. Viele Praktiker neigen dem „drill & practice“-Ansatz zu, Wissenschaftler finden sich in allen Lagern, und der NCTM vertritt die gemäßigt konstruktivistische Position des entdeckenden Ansatzes. Bemerkenswert ist, dass alle Positionen mit Ausnahme des radikalen Konstruktivismus der Übung eine wichtige Rolle zuweisen. Diese Einigkeit unterstreicht die zentrale Bedeutung von Übung beim Lernen von Mathematik.

Eine andere grundsätzliche Forderung an Unterricht ist die Individualisierung. Die übermäßig großen Streuungen der Leistungen deutscher Schüler in den PISA-Studien verleihen dieser Forderung ganz aktuell Gewicht. Nur durch individualisierten Unterricht kann verhindert werden, dass schwache Schüler den Anschluss an das Leistungsniveau einer Klasse verpassen. Ebenso sollten leistungsstarke Schüler individuell gefördert werden. Aus der Bedeutung der Übung und der Forderung nach Individualisierung lässt sich individualisierte Übung als ein wichtiger Baustein guten Mathematik-Unterrichts ableiten.

Welche inhaltliche Rolle spielt die Übung beim Rechnen Lernen in der Grundschule? Erstens ist Übung notwendig, um wiederkehrende Routineabläufe zu automatisieren. Dies sind beim elementaren Rechnen z.B. Strategien der Zahlzerlegung bei Zehnerübergängen, Strategien des Umganges mit Zehner- und Einerstellen beim Addieren zweistelliger Zahlen, oder die Arbeitsschritte beim schriftlichen Rechnen. Automatisierung von Abläufen geht mit Beschleunigung und einer Abnahme der benötigten Aufmerksamkeitsressourcen einher (Shiffrin & Schneider, 1984). Die dadurch frei werdenden Ressourcen stehen dann für komplexere Aufgabenstellungen zur Verfügung, wenn etwa beim Lösen von Sachproblemen mehrere Möglichkeiten durchgerechnet werden müssen. Zweitens dient Übung dem Aufbau eines Netzwerkes arithmetischer Fakten: Die Zerlegungen der Zahlen bis 20 und das kleine 1x1 sollen direkt aus dem Gedächtnis abgerufen werden können. Auch dies dient der Ersparnis von Aufmerksamkeit und Zeit für schwierigere Anforderungen. Wir sind keinesfalls der Ansicht, dass Übung das Verständnis der zugrunde liegenden Regeln und Gesetze ersetzen soll. Zum Erreichen von Verständnis erscheinen uns Methoden explorativen und entdeckenden Lernens, teilweise auch direkte Lehrmethoden als sinnvoll. Trotzdem lässt sich beobachten, dass sich manche mathematische Regelmäßigkeit dem Lernenden erst erschließt, wenn er eine gewisse Flüssigkeit im Rechnen erreicht hat (Briars & Siegler, 1984).

Das Rechnen-Lernen in der Grundschule kann als Erwerb kognitiver Fertigkeiten aufgefasst werden. Nach Anderson & Fincham (1994) beginnt der Aufbau von Fertigkeiten mit einer deklarativen Beschreibung des zu lernenden Ablaufs und seiner zugrunde liegenden Gesetze. Beim Addieren zweistelliger Zahlen kann man z.B. so vorgehen, dass man zuerst die Einerstellen addiert, sich das Ergebnis merkt, dann die Zehnerstellen addiert, und schließlich das Zwischenergebnis dazu addiert. Dabei kann man die Kenntnis des Kommutativgesetzes dazu benutzen, immer die kleinere von zwei Zahlen zur größeren zu addieren. Diese deklarative Beschreibung wird zunächst in einem langsamen Interpretationsprozess in Denk- und Handlungsschritte umgesetzt. Mit zunehmender Übung bilden sich als Elemente des prozeduralen Wissens Produktions-regeln heraus, die eine Beschleunigung bewirken, indem sie Wissen integrieren, das in früheren Stadien des Fertigkeitserwerbs erst im deklarativen Gedächtnis gesucht werden muss.

Beim Erwerb komplexerer Fertigkeiten gehen wir gemäß Anderson, Reder und Simon (1998, 2002) davon aus, dass diese sich aus Teilfertigkeiten zusammensetzen, die auch einzeln übbar sind. Beispielsweise wird die Fertigkeit der Addition einstelliger Zahlen in der schriftlichen Addition zur Teilfertigkeit. Die Annahme der Dekomponierbarkeit von Fertigkeiten wird vielfältig gestützt: Die „Kognitiven Tutoren“, intelligente tutorielle Systeme, die in der Arbeitsgruppe um Anderson entwickelt wurden (Anderson, Conrad, & Corbett, 1993; Koedinger, Anderson, Hadley, & Mark, 1997) nutzen dieses Prinzip und gehören zu den erfolgreichsten ihrer Art. Zweitens basieren Techniken der kognitiven Aufgabenanalyse (z.B. GOMS: Kieras & Meyer, 2000) auf der Dekomponierbarkeit von Arbeitsabläufen. Deren erfolgreiche Anwendung zur

Vorhersage von Ausführungs- und Lernzeiten spricht für die Angemessenheit der Annahme der Dekomponierbarkeit. Schließlich zeigen die Ergebnisse der Expertise-Forschung, dass zielgerichtete und umfangreiche Übung von Teilfertigkeiten (z.B. beim Spielen eines Instrumentes) Voraussetzung für das Erreichen eines Experten-Status ist (Ericsson, Krampe & Tesch-Römer, 1993).

Damit Übung effizient ist, muss sie mehrere Kriterien erfüllen. Zum einen müssen die angebotenen Aufgaben dem Fertigkeitsniveau des Übenden entsprechen, d.h. nicht zu leicht und nicht zu schwierig sein. Aus dem Potenzgesetz der Übung (s. Card, Moran & Newell, 1983) lässt sich ableiten, dass bei Aufgaben, die Übende schon gut beherrschen, weitere Übung nur noch wenig Verbesserung bringt. Zu schwierige Aufgaben würden die meisten Schüler dagegen entmutigen. Zum anderen muss die Richtigkeit der Antworten fortlaufend zurückgemeldet werden, damit sich Übende keine falschen Fakten oder Vorgehensweisen einprägen (Siegler, 1988). Dass unterschiedliche Schüler zudem unterschiedlich stark von Übungen profitieren und unterschiedlich lange brauchen, erzeugt weitere Anforderungen an die Individualisierung. Wie kann eine Lehrerin[1] 25 Schüler mit individuell zugeschnittenen Aufgaben versorgen und dabei ständig die Ergebnisse kontrollieren?

Zur Bewältigung dieser Anforderungen bietet sich der Einsatz von Computer-programmen an. Ein Programm kann im Bereich Rechnen, in dem es eindeutige Ant-worten gibt, problemlos die Richtigkeit von Antworten bzw. Fehler zurückmelden, so dass bei gegebener Geräteausstattung viele Schüler gleichzeitig üben können. Programme, bei denen die Schwierigkeit der Aufgaben von Hand angepasst werden kann, werden als adaptierbar bezeichnet (Leutner, 1995). Da eine derartige Auswahl von Aufgaben sehr zeitraubend ist, erfordert eine befriedigende Lösung des Individuali-sierungsproblems die automatische Auswahl geeigneter Aufgaben. Programme, die dies leisten, bezeichnet man als adaptiv. In diesem Sinne adaptive Programme können ohne lange Vorbereitung in Übungsstunden in der Schule verwendet werden. Außerdem können Schüler selbständig mit solchen Programmen arbeiten, ohne dass die Gefahr besteht, dass zu schwere oder zu leichte Aufgaben bearbeitet werden.

Im Bereich der Sekundarstufe werden adaptive Übungs-Programme mit tutoriellen Elementen bereits sehr erfolgreich eingesetzt. Der oben erwähnte Algebra-Tutor (Koedinger et al., 1997) ist in den USA aufgrund seiner Wirksamkeit inzwischen fester

Bestandteil des Unterrichts an vielen Schulen. Für die Primarstufe sind uns mit Ausnahme einer deutschen Version von „Accelerated Math“ (Lehmann & Seeber, 2005) adaptive Programme nicht bekannt. Sie wären aber gerade hier wünschenswert, um der bekannten Schere zwischen Schülern mit schwachen und guten Leistungen frühzeitig entgegen zu wirken.

Beim Rechnen in der Grundschule gilt das Lösen von Sachaufgaben als eine besondere Hürde (Stern, 1998), und die Grundlage für die weit verbreitete Abneigung gegen diesen Aufgabentyp wird vermutlich bereits hier gelegt. Wenn es gelänge, in der Primarstufe bei mehr Kindern sowohl Kompetenz im Umgang mit Sachaufgaben zu fördern, als auch eine zumindest neutrale Einstellung zu diesem Aufgabentyp zu erreichen, wäre viel für eine positive Weiterentwicklung schulischer Kompetenzen gewonnen. Denn in der Sekundarstufe müssen auch in Fächern wie Physik, Chemie oder Erdkunde Aufgaben gelöst werden, die eine Übersetzung von sprachlichen Situationsbeschreibungen in mathematische Modelle erfordern. Wir nehmen an, dass bestehende Aversionen gegen Sachaufgaben auf diese Aufgaben übertragen werden.

Die uns zugänglichen in Deutschland auf dem Markt erhältlichen Übungsprogramme für Mathematik der Primarstufe sind lediglich adaptierbar. Außerdem weichen die angegebenen Aufgabenschwierigkeiten zum Teil deutlich von den empirisch gefundenen Schwierigkeiten ab (Laue, 2004). Ein weiterer Schwachpunkt solcher Programme liegt in der eingeschränkten Bandbreite von Sachaufgaben. So konnte Laue (2004) mit dem Programm „Alfons Lernwelt“ bei individueller Aufgabenauswahl nach einer 8-wöchigen Übungsphase zwar signifikante Trainingsgewinne gegenüber nicht trainierten Kontrollgruppen nachweisen. Aber das Training wirkte sich nur auf Rechenaufgaben in Gleichungsform aus und nicht auf Sachaufgaben.

Die hier berichteten Untersuchungen zielen darauf ab, mit einem eigens entwickelten computergestützten Übungsprogramm – „Merlins Rechenmühle“ – die erwähnten Schwachpunkte kommerzieller Programme zu überwinden. Angestrebt wird ein adaptives Übungsprogramm mit automatisierten tutoriellen Elementen, das zur Differenzierung im Unterricht eingesetzt werden kann. „Merlins Rechenmühle“ ist nicht als Lehrprogramm konzipiert, das neue Prinzipien und Fertigkeiten vermittelt, sondern als Übungsprogramm. Durch die erwartete Effizienz individualisierter Übung kann die Lehrerin wertvolle Zeit für anderweitige instruktionale Maßnahmen gewinnen.

Theoretische Grundlage von „Merlins Rechenmühle“ ist die oben dargelegte Dekomposition von Fertigkeiten, dass also mit einfachen Aufgaben zunächst grundlegende Fertigkeiten geübt werden, die dann in schwierigeren Aufgaben als Teilfertigkeit vorkommen.

Auf dem Weg zu einem voll adaptiven Übungsprogramm muss zunächst ein Algorithmus zur Auswahl von Aufgaben gefunden und geprüft werden. Im ersten hier berichteten Experiment wurden zu diesem Zweck Regeln zur Aufgabenauswahl formuliert und angewendet. Da die Regeln Angaben über die Schwierigkeiten der auszuwählenden Aufgaben beinhalten, und es unmöglich war, die Schwierigkeiten von 2300 Aufgaben empirisch zu bestimmen, wurden die Schwierigkeiten in Studie 1 lediglich geschätzt. In Studie 2 wurden die Schwierigkeiten algorithmisch auf Grundlage von Aufgabenmerkmalen bestimmt. Ziel beider Untersuchungen war es also, die Angemessenheit der Algorithmen zur Aufgaben-Auswahl und zur Bestimmung der Aufgabenschwierigkeit zu überprüfen. Dies sollte indirekt über einen Nachweis der Wirksamkeit des individualisierten Trainings geschehen. Dabei kommt es nicht nur darauf an, dass sich die mittleren Leistungen der Übenden verbessern, sondern auch darauf, dass die Varianz der Leistungen nicht zunimmt (Atkinson, 1972). Von nicht individualisierter Übung profitieren nämlich immer nur Teilgruppen von Schülern. Anspruchsvolle Aufgaben ermöglichen leistungsstarken Schülern Lerngewinne, während schwächere Schüler überfordert sind. Die Folge sind mittlere Gewinne in Leistungsmaßen bei vergrößerter Varianz. Mit leichten Aufgaben fördert man selektiv nur die schwächeren Schüler, was zu leichten Steigerungen des Mittelwertes bei verringerter Varianz führt. Diese Forderungen wurden in der ersten Studie experimentell untersucht.

Studie 1

Studie 1 ist eine experimentelle Feldstudie mit Warte-Kontrollgruppen-Design. Das Experiment wurde im Rahmen des Dissertationsprojektes der zweiten Autorin durchgeführt und ist in Laue (2004) ausführlich beschrieben.

„Merlins Rechenmühle“ [3]

Zentraler Bestandteil des Übungsprogramms ist eine Datenbank mit über 2300 Aufgaben, die sich nach dem Aufgabentyp unterscheiden lassen. Es gibt Aufgaben zur

Orientierung im Zahlenraum, Rechenaufgaben in Form von Gleichungen, und zahlreiche unterschiedliche Sach- und Textaufgaben. Die Tabelle 1 enthält eine Liste aller Aufgabentypen.

Tabelle 1: Liste der im Lernprogramm „Merlins Rechenmühle“ verwendeten Aufgabentypen

Aufgaben zur Orientierung im Zahlenraum Nachbarzahlen finden (Einer und Zehner) Zahlen vergleichen Zahlen ordnen (auf- und abwärts) Reihen fortsetzen (in ... Schritten auf- und abwärts) Zahlen in Zehner und Einer zerlegen Zehner- und Einer-Kästchen „zählen“ Zahlen in der Hunderter-Tafel erkennen
Rechenaufgaben in Form von Gleichungen Additions-, Subtraktions-, Multiplikations- und Divisionsaufgaben mit jeweils zwei Operanden, einem Operator, einem Gleichheitszeichen; gesucht wird entweder der erste oder zweite Operand, oder das Ergebnis Additions- und Subtraktionsaufgaben mit den Einheiten Euro und Meter
Sach- und Textaufgaben Kombinationsaufgaben Austauschaufgaben Vergleichsaufgaben Multiplikationsaufgaben Divisionsaufgaben (partitiv und quotitiv) (diesen Aufgaben liegen Rechnungen zugrunde, die formal den oben beschriebenen Rechenaufgaben in Form von Gleichungen entsprechen) Komplexe Aufgaben mit mindestens zwei Operatoren
Aufgaben zur Übung des Umgangs mit Maus und Tastatur Strecken am Zahlenstrahl anzeichnen Strecken am Zahlenstrahl ablesen

Die Vorgabe von Aufgaben ist in Pakete gruppiert, die zumeist aus zehn Aufgaben bestehen. Die meisten Pakete enthalten Aufgaben eines Typs, die in ansteigender Schwierigkeit präsentiert werden, d.h. es gibt z.B. nur Austauschaufgaben mit unbekannter Austauschmenge[2], oder nur Additionsaufgaben mit dem Platzhalter an zweiter Position, wobei die Zahlen von Aufgabe zu Aufgabe größer werden. Um auch die flexible Umstellung zwischen unterschiedlichen Aufgabentypen zu üben, gibt es auch Pakete, in denen Aufgaben von verwandten Typen (z.B. verschiedene Untertypen von Austauschaufgaben, Additionsaufgaben mit Platzhaltern an unterschiedlichen Positionen) gemischt vorgegeben werden.

Jede Übungs-Sitzung beginnt mit dem Einloggen des Probanden ins Programm. Mit Hilfe des Nutzer-Namens wird Information über die zu bearbeitenden Aufgabenpakete eingelesen und Daten über die Aufgabenbearbeitung protokolliert, und zwar für jede Aufgabe die Vorgabe, die Werte von ein oder zwei Lösungsversuchen, ein Code für die Richtigkeit der Lösung, sowie Bearbeitungszeiten.

Die Fenster zur Präsentation der Aufgaben sind in kräftigen Farben, aber schlicht gestaltet. Das Programm informiert den Probanden über die Richtigkeit der Lösung mittels verschiedener Bilder, die zusammen mit den Texten „Richtig“, „Prima“, oder „Falsch“ präsentiert werden. Pro Aufgabe gibt es zwei Lösungsversuche. Fortschritt und Erfolg innerhalb eines Paketes werden am unteren Rand des Fensters von einem Fortschrittsbalken angezeigt. Nach jedem Paket, bei dem mindestens die Hälfte der Aufgaben richtig gelöst wurden, erscheint eine kurze Trickfilm-Sequenz auf dem Bildschirm. Bei dieser Gelegenheit wird auch die in allen Sitzungen gesammelte Gesamtpunktzahl angezeigt. Nach fünf Paketen wird das Ende der Sitzung angezeigt und das Programm beendet. Es ist natürlich auch möglich, dass der Versuchsleiter das Programm über eine Tastenkombination vorzeitig beenden kann.

Die Tests zur Erfassung der Mathematik-Leistung

Zur Erfassung der Mathematik-Leistung vor und nach dem Training entwarfen wir einen Test, der in drei parallelen Versionen vorliegt. Die Tests enthalten Aufgaben verschiedener Typen in gemischter Reihenfolge und in ansteigender Schwierigkeit. Sie werden mit einer Zeitbegrenzung von 45 Minuten vorgegeben. Die Aufgabentypen entsprechen weitgehend denen des Lernprogramms (s. Tabelle 1).

Stichprobe und Design

Studie 1 ist als experimentelle Feldstudie angelegt, der ein Warte-Kontrollgruppen-Design mit Messwiederholungen zugrunde liegt. Zu Beginn im September 2002 wurden aus drei dritten Klassen einer Bayreuther Grundschule 31 freiwillige Teilnehmer/innen

(16 Mädchen und 15 Jungen) rekrutiert, die per Zufall auf die beiden Versuchsgruppen verteilt wurden. Bei allen Versuchsteilnehmern wurde das Ausgangsniveau im mathematischen Wissen und Können beim Lösen von Aufgaben erhoben. Danach nahm die erste Teilgruppe (n=16) für acht Wochen an den computergestützten Übungen teil. Die zweite Teilgruppe (N=15) fungierte in dieser Phase als Kontrollgruppe. In der neunten Woche wurden wieder beide Teilgruppen mit einer Parallel-Version des Mathematik-Tests getestet. Im Januar 2003 begannen dann für acht Wochen die Übungen für die zweite Teilgruppe, die im Herbst keine Übungen durchgeführt hatte. Die Untersuchung endete mit einem anschließenden dritten Test an beiden Teilgruppen, der als Paralleltest zu den beiden anderen konstruiert worden war.

Auswahl der Aufgaben und Ablauf der Übungs-Sitzungen

Die erste Sitzung begann mit zwei Aufgabenpaketen zur Übung des Umgangs mit Maus und Tastatur, gefolgt von zwei Paketen mit Textaufgaben und einem Paket mit numerischen Aufgaben. In jeder weiteren Übungsstunde sollten die Kinder fünf Aufgabenpakete bearbeiten. Dabei gingen wir nach folgenden Regeln vor:

- Numerische Aufgabenpakete werden im Wechsel mit Sachaufgaben-Paketen ausgewählt.
- Nach jeweils vier Paketen wird ein Paket mit Aufgaben zur Orientierung im Zahlenraum ausgewählt.
- Innerhalb der numerischen bzw. Sachaufgaben wechseln sich Pakete mit Additions/Subtraktions- und Multiplikations/Divisions-Aufgaben ab.
- Mit fortschreitender Fertigkeit des Übenden, werden Pakete mit gemischten Aufgabentypen vorgegeben (z.B. bei Sachaufgaben Austausch- und Vergleichsaufgaben); ebenso komplexere Sachaufgaben mit mehr als einem Rechenschritt, sowie „Zahlenrätsel“.
- Bei erfolgreicher Bearbeitung eines Aufgabenpaketes, d.h. mindestens 80% richtiger Lösungen wird die Schwierigkeit beim nächsten Paket gesteigert.

Es gab acht Übungs-Sitzungen, die jeweils eine Stunde dauerten. Während der Übungsstunden wurden die Probanden von der Zweit-Autorin zusammen mit Lehramtstudentinnen beaufsichtigt. Bei auftretenden Schwierigkeiten wurde den Kindern möglichst unspezifisch gehaltene Hilfe angeboten, etwa die Erklärung unbekannter Begriffe, oder die Aufforderung, die Aufgabe noch einmal langsam zu lesen. In einzelnen Fällen wurden Kinder in der vorgegebenen Zeit nicht mit allen Aufgabenpaketen fertig. Die nicht bearbeiteten Pakete wurden dann in die nächste Stunde verschoben.

Ab der fünften Sitzung wurde je ein Paket, das vier Wochen zuvor besonders schlecht bzw. langsam bearbeitet worden waren, wiederholt vorgegeben. Die Kinder wurden auf diesen Umstand hingewiesen und konnten ihre aktuellen Ergebnisse mit denen der ersten Bearbeitung vergleichen. Da es überwiegend zu Verbesserungen kommt, sollte diese Maßnahme die Kinder anspornen und im Sinne einer individuellen Bezugsnorm-Orientierung motivieren (Rheinberg & Fries, 1998).

Ergebnisse

Um die Reliabilität des Vor- bzw. Nachtests zu schätzen, wurden Korrelationen der Gesamtscores zwischen den Messzeitpunkten berechnet. Diese betrugen zwischen Vortest und Nachtest 1 r=.82, und zwischen Nachtest 1 und Nachtest 2 r=.85. Da es sich bei Vor- und Nachtest um parallele Testformen handelt, entspricht dieses Vorgehen einer Kombination aus Paralleltest- und Retest-Reliabilität.

Zur Beantwortung der Frage, ob die individualisierten Übungen am Computer eine zusätzliche Steigerung der Mathematik-Leistung – im Vergleich zum alleinigen Schulunterricht – bewirken, berechneten wir Varianzanalysen mit dem zwischen Versuchpersonen variierten Faktor „Trainingszeitraum“ (Training in der ersten vs. zweiten Phase), dem Messwiederholungsfaktor „Messzeitpunkt“ (Vortest, Nachtest 1, Nachtest 2) und dem Gesamtscore im jeweiligen Test als abhängiger Variable. Da sich die trainierten Gruppen stärker verbessern sollten als die nicht trainierten Gruppen, zeigt sich die Wirksamkeit des Trainings in einer signifikanten Interaktion zwischen beiden Faktoren.

Für die Gesamtpunktzahl im Mathematik-Test ergab diese Analyse die erwartete signifikante Interaktion zwischen Trainingszeitraum und Messzeitpunkt ($F_{2,30}$ =11.27, p<.01, η^2=.45). Die Mittelwerte zeigen an, dass der Leistungsanstieg der in der ersten Phase trainierten Kinder (Gruppe 1) höher war als bei der nicht trainierten Gruppe (Gruppe 2). Im zweiten Nachtest erzielte die in der zweiten Phase trainierte Gruppe 2 einen Leistungszuwachs, der verglichen mit dem leichten Rückgang bei Gruppe 1, ebenfalls auf das Training zurückgeführt werden kann (siehe Abb. 1).

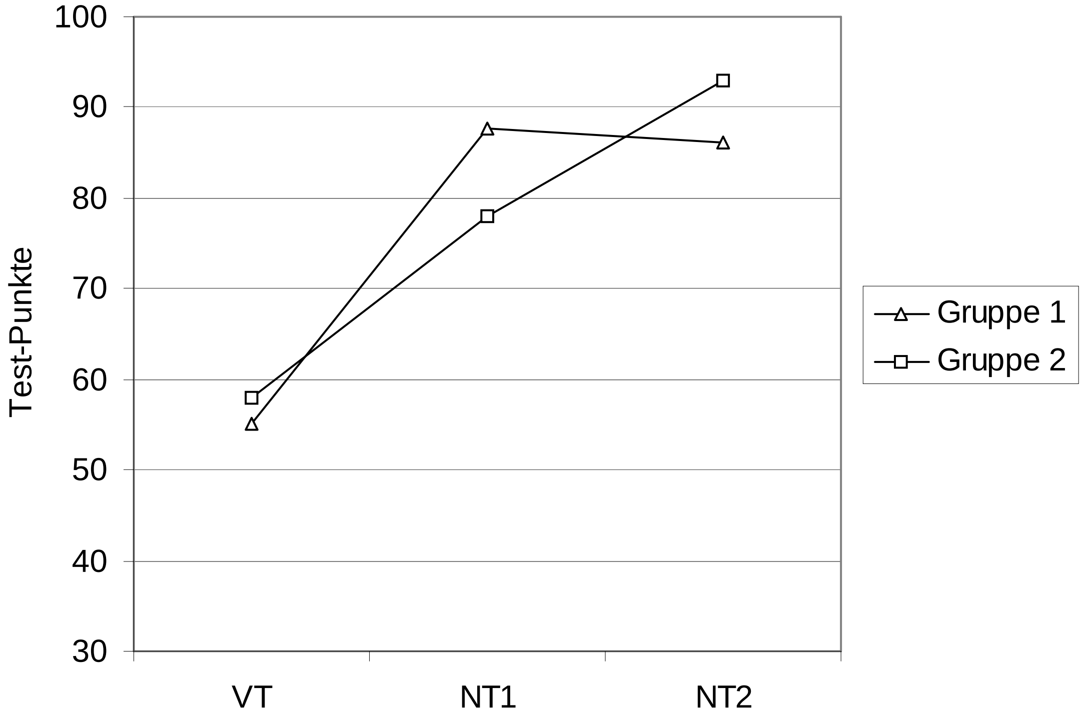

Abbildung 1: Mittlere Gesamtpunktwerte im Mathematik-Test zu drei Messzeitpunkten in den beiden Trainingsgruppen (VT: Vortest, NT: Nachtest, Gruppe 1: Training zwischen VT und NT1, Gruppe 2: Training zwischen NT1 und NT2).

Zur weiteren Absicherung der Trainingseffekte berechneten wir Kovarianzanalysen mit der Testleistung nach dem Training als abhängiger Variable, dem zwischen Versuchpersonen variierten Faktor „Training“, und der Testleistung vor der Trainingsphase als Kovariate. Für beide Trainingszeiträume ergeben sich Unterschiede zwischen den Gruppen, die sich in signifikanten Haupteffekten des Faktors „Training“ zeigen (1. Zeitraum: $F_{1,30}$ =7.68, p<.05; Zeitraum $F_{1,30}$ =25.88 , p<.01). Auch die Effekte der Kovariate „Testleistung vor der Trainingsphase“ sind in beiden Fällen signifikant (1. Zeitraum: $F_{1,30}$ =74.71 , p<.01, 2. Zeitraum $F_{1,30}$ =157.36 , p<.01). Dies bedeutet, dass die Leistung nach der Trainingsphase durch die Leistung zu Beginn vorhergesagt wird.

Da das Ausgangsniveau der beiden Versuchsgruppen im Vortest annähernd gleich war, kann der direkte Vergleich der trainierten mit der nicht trainierten Gruppe zum ersten

Nachtest (NT1) herangezogen werden, um die Effektgröße des Trainings abzuschätzen. Die gefundene Effektgröße von d=0.49 entspricht einem mittelgroßen Effekt.

Im ersten Trainingszeitraum verändern sich die Standardabweichungen aller Gruppen kaum (Trainierte Gruppe: s_0=20.4 im Vortest, s_1=19.1 im Nachtest 1; Wartekontroll-gruppe: s_0=23.2, s_1=22.3). Dies entspricht der Erwartung, dass das Training die Leistungsstreuung keinesfalls erhöhen sollte. Im zweiten Trainingszeitraum nehmen dagegen die Standardabweichungen beider Gruppen wieder zu (Gruppe 1, jetzt Kontrollgruppe: s_2=26.0 im Nachtest 2; Gruppe 2, jetzt Trainingsgruppe: s_2=26.0 im Nachtest 2). In Gruppe 1 ist dies vor allem auf die im Vergleich zu Nachtest 1 schlechteren Ergebnisse vieler Schüler zu erklären. In der Gruppe 2 – also der im zweiten Zeitraum trainierten Gruppe – kam es zu dem unerwünschten Effekt, dass die stärkeren Schüler mehr vom Training profitierten als die schwächeren.

Zusammenfassend kann festgestellt werden, dass das Übungsprogramm von nur 8 einstündigen Sitzungen im ersten Trainingszeitraum alle erwarteten Effekte erbrachte. Auch im zweiten Trainingszeitraum kam es zu deutlichen relativen Verbesserungen der Trainingsgruppe gegenüber der Kontrollgruppe, die sich verschlechterte. Allerdings nahm auch die Varianz der Leistungen zu Gunsten der stärkeren Schüler zu.

Studie 2

In Studie 2 wurden verschiedene Veränderungen des Lernprogramms und des Leistungstests erprobt. Wir änderten den Test dahingehend, dass die Aufgaben nach den drei Kategorien „Orientierung im Zahlenraum“, „Textaufgaben“ und „Rechenaufgaben“ geordnet in drei separat zeitbegrenzten Abschnitten vorgegeben wurden. So konnte für jede der drei Aufgabenkategorien ein eigener Summenscore gebildet werden. Wir versuchten außerdem, den Anteil mittelschwerer Items im Test zu erhöhen. Die verwendeten Aufgabentypen entsprechen denen der Tests aus Studie 1.

Die zweite Veränderung betraf die Schätzung der Aufgabenschwierigkeiten. Für Text- und Rechenaufgaben mit zwei Operanden und einem Operator entwickelten wir einen Algorithmus der die geschätzte Schwierigkeit der Aufgabe aus einer Linearkombination von gewichteten Aufgabenmerkmalen errechnet. Die so gewonnenen Schwierigkeiten lagen der individuellen Aufgabenauswahl von Studie 2 zugrunde.

Zur Festlegung des Einflusses von Aufgabenmerkmalen auf die Schwierigkeit stützten wir uns auf Ergebnisse von Parkman und Groen (1971), Kornmann und Wagner (1990), und Daten aus Studie 1. Danach sind relevante Merkmale von Rechenaufgaben: (1) ob sich die Aufgabe sich im Zahlenraum bis oder über 20 bewegt, (2) ob ein Zehner-übergang vorliegt oder nicht, (3) die Anzahl der Ziffern des kleineren Operanden, (4) die Position der gesuchten Zahl in der Gleichung, (5) die Gleichheit bzw. Ungleichheit der Operanden, und (6) der Operator. Bei Sachaufgaben haben wir als weiteres Merkmal das Situationsmodell nach Riley, Greeno und Heller (1983) bzw. Stern (1998) berücksichtigt. Kombinations-, Austausch-, oder Vergleichsaufgaben sind allein aufgrund ihrer Situationsstruktur unterschiedlich schwierig. Multiplikationsaufgaben werden in ihrer geschätzten Schwierigkeit den Austauschaufgaben, Divisionsaufgaben den Vergleichsaufgaben gleichgestellt. Komplexen Textaufgaben mit mehr als einer Rechenoperation ordneten wir konstant einen etwas höheren Schwierigkeitsgrad zu als den schwierigsten Vergleichs- und Divisionsaufgaben.

Tabelle 2: Aufgabenmerkmale und ihre Zahlenkodierung und Gewichte im Algorithmus zur Vorhersage der Aufgabenschwierigkeit.

Merkmal	Zahlenkodierung und Gewicht (a)
Zahlenraum	ZR bis 20 → 4, ZR über 20 → 3, a_1=3
10er Übergang in der Einer-Stelle	kein Übergang → 2, Übergang → 1, a_2=3
Anzahl Ziffern der kleinsten Zahl	eine Ziffer → 4, zwei Ziffern → 3, drei Ziffern → 2, vier oder mehr Ziffern → 1, a_3=3
Position der gesuchten Zahl (z.B. 23 + ? = 71: ges. Zahl an zweiter Position)	dritte Position → 3, zweite Position → 2, dritte Position → 1, a_4=3
Operanden	gleiche Operanden → 2, ungleiche Operanden → 1, a_5=1
Operator	Addition → 3, Subtraktion, Multiplikation → 2, Division → 1, a_6=2
Situationsmodell (nur bei Textaufgaben)	Kombination → 3, Austausch, Multiplikation → 2, Vergleich, Division → 1, a_7=2

Die Aufgabenmerkmale werden zahlenmäßig so kodiert, dass die schwierigste Variante den Wert 1 erhält, die leichteren Varianten höhere Zahlen. Die Linearkombination aus den gewichteten Werten für die Aufgabenmerkmale wird gemäß Gleichung 5 (siehe

Anhang) auf das Intervall der Lösungswahrscheinlichkeiten für das schwierigste und das leichteste Item skaliert. Tabelle 2 enthält die verwendeten Aufgabenmerkmale mit ihren Ausprägungen und Gewichten.

Die Vorhersagen des Algorithmus wurden vor dem Vortest in die Aufgabendatenbank eingearbeitet. Deshalb können die Daten aus dem Vortest zu seiner Validierung herangezogen werden. Der Algorithmus sagt die Rangfolge von Schwierigkeiten bei Additions- und Subtraktionsaufgaben gut vorher (*r*=.82). Die Regressionsgerade ist allerdings dahingehend verschoben, dass die empirischen Werte im unteren Bereich unterschätzt werden (s. Abb. 2). Da es bei der Aufgabenauswahl aber vor allem auf die Abfolge von Schwierigkeitsstufen und weniger auf deren absolute Höhe ankommt, dürften sich diese Abweichungen kaum störend auf die Aufgabenauswahl ausgewirkt haben.

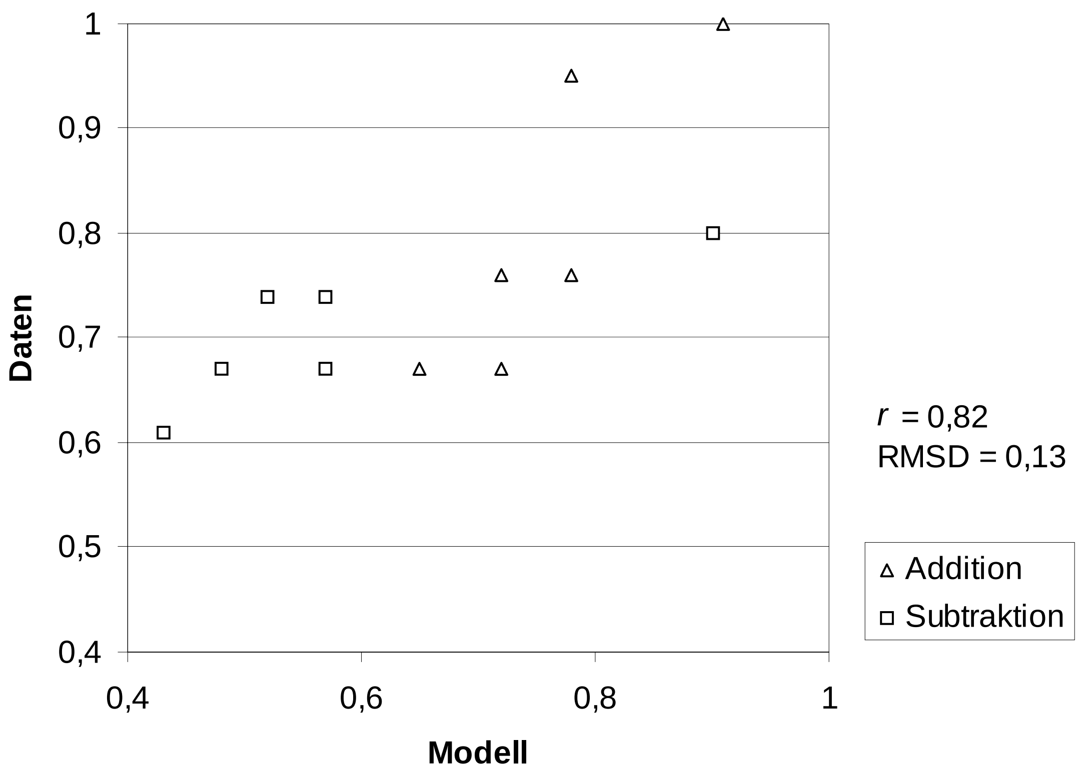

Abbildung 2: Zusammenhang zwischen algorithmisch geschätzten (Modell) und empirischen (Daten – Vortest Studie 2) Aufgabenschwierigkeiten verschiedener Additions- und Subtraktionsaufgaben. Alle Aufgaben: 2-stellige Zahlen, Zahlenraum bis 100, variabel ist der Operator, die Position des gesuchten Operanden und das vorliegen eines Zehnerüberganges.

Eine dritte Veränderung in Studie 2 gegenüber Studie 1 bestand in der erweiterten Nutzung von Datenbank-Technologie. Die Aufgaben-Datenbank wurde in eine relationale Struktur umorganisiert und durch neue Aufgaben erweitert. Die oben beschriebenen Aufgabenmerkmale wurden mit jeder Aufgabe gespeichert. Dies ermöglichte eine halbautomatische Zusammenstellung von Aufgabenpaketen, da die Menge der infrage kommenden Aufgaben mit einer einzigen Datenbankabfrage ermittelt werden kann. Erstmals wurden auch alle Nutzer-Daten in mehreren Datenbank-Tabellen abgelegt. Da nach allen Merkmalen recherchiert werden kann, kann das Programm prinzipiell feststellen, ob ein Nutzer noch Schwierigkeiten bei Aufgaben mit bestimmten Merkmalen oder Merkmalskombinationen hat.

Schließlich verwendeten wir in Studie 2 mehr gemischte Pakete, da informelle Beobachtungen ergeben hatten, dass die Kinder vor allem bei Textaufgaben schnell das zugrunde liegende Lösungsschema für ein Paket erkennen und dann anwenden, ohne jedes Mal ein Situationsmodell aufzubauen. Ein solches Vorgehen verhindert tieferes Verständnis und muss deshalb vermieden werden.

Anlage und Ablauf

Studie 2 konnte aus organisatorischen Gründen nicht als Wartekontrollgruppen-Experiment durchgeführt werden. Es fehlt eine randomisierte Kontrollgruppe. Wir hatten jedoch ein halbes Jahr später Gelegenheit, an derselben Schule eine weitere dritte Klasse im Abstand von neun Wochen zweimal zu testen (mit demselben Test wie in Studie 2). Da wir außerdem aus früheren eigenen und fremden Untersuchungen Daten über die Größenordnungen von Verbesserungen mit und ohne Training haben, versuchen wir, den Effekt des Trainings über den Vergleich von Effektgrößen abzuschätzen. Auch Veränderungen in der Leistungsvarianz können ohne direkten Vergleich mit einer Kontrollgruppe beurteilt werden. Angestrebt und erwartet wurde eine Erhöhung der Mittelwerte bei gleich bleibender Varianz als Folge des individualisierten Trainings.

20 Schülerinnen und Schüler aus zwei dritten Klassen einer Bayreuther Grundschule nahmen am Versuch teil. Die Studie begann mit dem Vortest im Oktober 2004. In den acht darauf folgenden Wochen fanden sieben Übungssitzungen statt. Mitte Dezember 2004 endete die Studie mit dem Nachtest, der als Parallel-Version zum Vortest konzipiert war. Der Ablauf der einzelnen Sitzungen war wie in Studie 1.

Ergebnisse

Zur Schätzung der Reliabilität des Vor- bzw. Nachtests wurden Korrelationen der Summenscores der einzelnen Aufgabenkategorien und des Gesamtscores zwischen erster und zweiter Messung berechnet. Für die Aufgaben zur Orientierung im Zahlenraum beträgt die Korrelation r=.69, für die Rechenaufgaben r=.68, für die Textaufgaben r=.70. Für die Reliabilität des Gesamtergebnisses ergibt sich nach diesem Verfahren r=.87. Diese Werte qualifizieren den Test als hinreichend genau, um Ergebnisse interpretieren zu können. Auch der Anteil an Aufgaben mittleren Schwierigkeitsgrades konnte – insbesondere bei den Textaufgaben – erhöht werden. Von den sieben im Vortest von allen Kindern bearbeiteten Textaufgaben liegen fünf in ihrer Schwierigkeit im erwünschten Intervall zwischen 0.2 und 0.8; zwei Aufgaben sind leichter. Im Nachtest liegen zehn von zwölf Aufgaben in diesem Intervall.

Alle 20 Kinder verbesserten sich vom Vor- zum Nachtest. Die Mittlere Gesamt-Punktzahl beträgt im Vortest 64.6 Punkte (s=17.1), im Nachtest 93.4 (s=15.1). Die Standardabweichung der Differenzen zwischen Vor- und Nachtest beträgt s=8.35. Daraus errechnet sich eine Effektgröße von d' =3.45. (s. Bortz & Döring, 2002). Die nicht trainierten Kontrollgruppen an der betroffenen Schule verbesserten sich in vergleichbaren Zeiträumen dagegen nur um d' =2.00 (Gruppe 2 in Studie 1), d' = 0,67 („nachträgliche Kontrollgruppe“ von Studie 2), oder verschlechterten sich sogar (Gruppe 1 in Studie 1). Die Verbesserungen der trainierten Gruppen in Studie 1 lagen mit d' = 2.70 (Gruppe 1) und d' =1.46 (Gruppe 2) unter der Verbesserung in Studie 2.

Weitere aktuelle Vergleichszahlen finden sich in einer groß angelegten Feldstudie zur Wirksamkeit des Programms „Accelerated Math“ (Lehmann & Seeber, 2005). Bei zwei Messungen mit dem HST im Abstand von vier Monaten fanden sich in acht vierten Klassen nur Leistungssteigerungen in Höhe von d' =1.00, und zwar sowohl in den trainierten, wie in den nicht trainierten Klassen[4]. Der Vergleich der Effektgrößen aus mehreren Untersuchungen zeigt also, dass ohne Training nur Leistungssteigerungen bis maximal d' =2.00 gefunden wurden. Es ist daher sehr unwahrscheinlich, dass die Steigerung von d' =3.45, wie sie in Studie 2 gefunden wurde, allein durch Unterrichtseinflüsse erklärbar wäre. Das Training ist mindestens so effektiv wie das aus Studie 1.

Tabelle 3: Mittelwerte und Standardabweichungen (in Klammern) der Testwerte in den drei Grundtypen von Aufgaben.

	Vortest	Nachtest	Effektgröße *d'*
Orientierung im Zahlenraum	24,0 (8,2)	31,4 (6,7)	1,24
Textaufgaben	23,4 (7,7)	34,2 (7,8)	1,79
Rechenaufgaben	17,3 (7,3)	27,9 (5,2)	1,97

Auch in den einzelnen Aufgabenkategorien wurden durchweg signifikante und große Verbesserungen erzielt (s. Tabelle 3). Die Standardabweichungen der Nachtest-Ergebnisse sind fast durchweg kleiner als die der Vortest-Ergebnisse. Zusammen mit der deutlichen Steigerung der Mittelwerte ist dies ein wünschenswertes Ergebnis, das ohne Training kaum erreicht werden kann.

Diskussion

Das Training mit „Merlins Rechenmühle“ hat sich als wirksam zur bedeutsamen Verbesserung der Rechenfertigkeiten von Drittklässlern erwiesen. In allen Untersuchungen lagen die Übungsgewinne der trainierten Gruppen über denen der Kontrollgruppen mehrerer anderer Studien. Der Erfolg der Individualisierung wird auch dadurch belegt, dass bis auf eine Ausnahme die Streuung der Testwerte wenigstens gleich blieb, sich teilweise sogar verringerte. Dies bedeutet, dass Schüler mit sehr unterschiedlichen Eingangsvoraussetzungen vergleichbar von den Übungen profitierten.

Kulik (1994) errechnete in einer Metaanalyse von Untersuchungen zur computer-gestützten Instruktion eine mittlere Effektgröße von d=0.35 zwischen trainierten und nicht trainierten Gruppen. Die in unserer Studie 1 ermittelte Effektgröße liegt mit d=0.49 deutlich darüber. Welchen Anteil die individuelle Aufgabenauswahl am Erfolg des Trainingsprogramms hat, wird im Vergleich zu Untersuchungen deutlich, in denen kommerzielle Computerprogramme mit unausgelesenen Aufgaben verwendet wurden. So gab es in einem Experiment zum Einfluss eines metakognitiven Trainings von Desoete, Roeyers und De Clercq (2003), eine solche Computertrainings-Bedingung, die keine Übungsgewinne im Vergleich zu einer Kontrollgruppe bewirkte.

Diese Ergebnisse sprechen indirekt für das Funktionieren der Algorithmen zur Bestimmung der Aufgabenschwierigkeit und zur Aufgabenauswahl. Es erscheint also Erfolg versprechend, diese Algorithmen in eine automatisch adaptive Version von Merlins Rechenmühle zu übernehmen.

Zu erwägen sind bei dem hier verwendeten Vor-Nachtest-Design mögliche Testwiederholungs-Effekte, die darin bestehen, dass die reine Wiederholung eines Tests schon zu Leistungssteigerungen führt. Nach Klauer (2001) erreichen diese Effekte nach einigen Tagen bis zu zwei Wochen einen Höhepunkt und verschwinden nach etwa zwei Monaten (Willson & Putnam, 1982). Demnach dürften diese Effekte in den vorliegenden Untersuchungen keine bedeutsame Rolle gespielt haben.

Die hier berichteten Untersuchungen lieferten wertvolle Hinweise für die Weiterentwicklung des Lernprogramms. So zeigte sich, dass manche Kinder mit zehn Aufgaben pro Paket überfordert waren, weil sie die vorgesehenen fünf Pakete nicht in der vorgegebenen Zeit schafften. Als zusätzlich zu adaptierender Parameter käme deshalb die Paketgröße in Frage. Auch die Möglichkeit des Programms zur Fehleranalyse soll ausgebaut werden. Da detaillierte Aufgabenmerkmale mit in der Datenbank gespeichert sind, wäre es möglich, gehäufte Fehler bei einzelnen Merkmalen aufzuspüren und bei der Aufgabenauswahl zu berücksichtigen.

Insgesamt erscheint uns die angestrebte Form von Individualisierung auf Ebene von Teilfertigkeiten noch so ausbaufähig, dass die Trainingserfolge gesteigert werden können, ohne den hohen Aufwand zu betreiben, auf die Ebene einzelner Produktionsregeln zurück zu gehen (Anderson et al., 1993).

Eine interessante Beobachtung betrifft den Einfluss des Schulunterrichts auf die Testleistungen. Dieser zeigte sich Studie 1 darin, dass die Beschäftigung mit dem 1x1 in der Schule zwischen Vortest und erstem Nachtest von einem sprunghaften Anstieg in der Leistung bei Multiplikations- und Divisions-Aufgaben gefolgt wurde. Dieses Leistungsniveau konnte im zweiten Nachtest von der Trainingsgruppe 2 gerade gehalten werden, während sich die Kontrollgruppe verschlechterte. Die unerwartete Verschlechterung lässt sich im Nachhinein damit erklären, dass während des zweiten Trainings im Unterricht kaum 1x1-Aufgaben geübt wurden. Individualisierte computergestützte Übungen sind also auch günstig für die Sicherung zurück liegender Lernergebnisse.

Längerfristig wäre für das Lernprogramm eine Skalierung der Aufgaben nach der Item Response Theorie sehr nützlich. Damit ist eine Schätzung der Fähigkeitsparameter mit einer individuellen Stichprobe von Fragen möglich, was die automatische Aufgabenauswahl auf einer gesicherten messtheoretischen Grundlage erleichtern würde. Bei Rechenaufgaben, möglicherweise auch bei Zahlenraumaufgaben erscheint eine Rasch-Skalierung machbar und sollte angestrebt werden. Bei Textaufgaben, deren Lösung ein ganzes Bündel an heterogenen Anforderungen stellt, ist es fraglich, ob die für eine Rasch-Skalierung geforderte lokale stochastische Unabhängigkeit sinnvoll postuliert werden kann (siehe aber Arendasy, Sommer & Glück, 2004 für eine überraschend homogene Skalierung von Austauschaufgaben).

Literaturverzeichnis

Arendasy, M., Sommer, M. & Glück, J. (2004). Dimensionalität und differenzielle Validität von Textaufgaben. *Zeitschrift für Pädagogische Psychologie, 18*, 231-243.

Anderson, J. R., Conrad, F. G. & Corbett, A. T. (1993). The LISP tutor and skill acquisition. In J. R. Anderson (Ed.), *Rules of the mind* (pp.143-164). Mahwah, NJ: Erlbaum.

Anderson, J. R. & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1322-1340.

Anderson, J. R., Reder, L. M. & Simon, H. A. (1998). Radical constructivism and cognitive psychology. In D. Ravitch (Ed.), *Brookings papers on education policy* (pp. 227-255). Washington, DC: Brookings Institution.

Anderson, J. R., Reder, L. M. & Simon, H. A. (2000). Applications and misapplications of cognitive psychology to mathematics education. *Texas Educational Review*

Atkinson, R. C. (1972). Ingredients for a theory of instruction. *American Psychologist, 27,* 921-931.

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. H. Schoenfeld (Ed.), In A. J. Baroody & A. Dowker (Ed.), *The development of arithmetic concepts and skills: Constructing adaptive expertise.* (pp. 1-33). Mahwah, NJ: Erlbaum.

Card, S. K., Moran, T. P. & Newell, A. (1983). *The Psychology of Human - Computer Interaction.* Hillsdale, NJ: Erlbaum.

Cobb, P., Yackel, E. & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education, 23,* 2-33.

Desoete, A., Roeyers, H. & De Clercq, A. (2003). Can offline metacognition enhance mathematical problem solving? *Journal of Educational Psychology, 95,* 188-200.

Ericsson, K. A., Krampe, R. T. & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363-406.

Heckhausen, H. (1989). *Motivation und Handeln.* Berlin: Springer.

Kieras, D. E. & Meyer, D. E. (2000). The role of cognitive task analysis in the application of predictive models of human performance. In R. R. Hoffman; N. J. Cooke; K. A. Ericsson; G. Klein; E. Salas; D. K. Simonton; R. J. Sternberg (Ed.), In J. M. Schraagen; S. F. Chipman; V. L. Shalin (Ed.), *Cognitive task analysis. Expertise: Research and Applications* (pp.237 -260). Mahwah, NJ: Erlbaum.

Klauer, K. J. (2001). *Handbuch kognitives Training*. Göttingen: Hogrefe.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*, 30-43.

Kornmann, R. & Wagner, H.-J. (1990). Ermittlung der Lernbasis bei einfachen Kopfrechenaufgaben im Zahlenraum 0-20. *Zeitschrift für Heilpädagogik, 41* (Beiheft 17), 211-218.

Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. L. Baker & H. F. O'Neil (Eds.), *Technology assessment in education and training* (pp. 9-33). Hillsdale, NJ: Erlbaum.

Laue, C. (2004). *Individualisierung beim Training von Rechenfertigkeiten*. Universität Bayreuth: Dissertation zur Erlangung des Doktorgrades.

Leutner, D. (1995). Adaptivität und Adaptierbarkeit multimedialer Lehr- und Informationssysteme. In L. J. Issing, & P. Kliemsa (Eds.), *Information und Lernen mit Multimedia* (pp.139 -147).Weinheim: Beltz-PVU.

Parkman, J. M. & Groen, G. J. (1971). Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology, 89,* 332-342.

Rheinberg, F. & Fries, S. (1998): Förderung der Lernmotivation: Ansatzpunkte, Strategien und Effekte. *Psychologie in Erziehung und Unterricht, 45,* 168-184.

Riley, M. S., Greeno, J. G. & Heller, J. H. (1983). Development of children's problem-solving ability in arithmetic. In: H.P. Ginsburg (Ed.). *The development of mathematical thinking* (pp. 153-196). New York: Academic Press.

Shiffrin, R. M. & Schneider, W. (1984). Automatic and controlled processing revisited. *Psychological Review, 91*, 269-276.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*, 258-275.

Stark, R, Graf, M, Renkl, A, Gruber, H & Mandl, H. (1995). Förderung von Handlungskompetenz durch geleitetes Problemlösen und multiple Lernkontexte. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 32,* 289-312.

Staub, F. C. & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, 94,* 344-355.

Stern, E. (1998). Die Entwicklung des mathematischen Verständnisses im Kindesalter. Lengerich: Pabst Science Publishers.

Willson, V. L. & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal, 19*, 249-258

Fußnoten

[1] Mit „Lehrerin“ sind hier weibliche und männliche Personen gleichermaßen gemeint, ebenso wie mit „Schüler“.

[2] Austauschaufgaben sind eine Art von Sachaufgaben, bei denen Gegenstände zwischen Personen ausgetauscht werden; bei Kombinationsaufgaben werden Mengen von Gegenständen vereiningt; in Vergleichsaufgaben werden die Mächtigkeiten von Mengen miteinander verglichen, ohne dass etwas Anschauliches dabei passiert

[3] „Merlins Rechenmühle“ wurde vom Erstautor mit dem Autorensystem „Mediator“ der Firma Matchware entwickelt. Bei der Version von Studie 2 hat Stefan Loos maßgeblich am Aufbau der Datenbank mitgewirkt, dem wir an dieser Stelle herzlich danken wollen.

[4] Die Autoren selbst geben niedrigere Werte für die Effektgrößen an, da sie die Mittelwerte der Differenzen durch die gemittelten Standardabweichungen des Vor- und Nachtests teilen und nicht, wie es richtig wäre, durch die Standardabweichung der Differenzen.