

Autor: [Bernhard Jacobs](#), Fachrichtung 5.1 Erziehungswissenschaft der Universität des Saarlandes
created: 10.1.2008

URL des Originals: <http://www.phil.uni-sb.de/~jakobs/wwwartikel/teststudy/teststudy3.html>

Führt selbst gesteuertes Testen mit Feedback zu höheren Behaltensleistungen als das Einprägen mit Hilfe einer Landkarte?

Abstract

In der Studie von [Jacobs \(2007\)](#) erzielten Studenten höhere Behaltensleistungen durch selbst gesteuertes Testen mit Feedback im Vergleich zu einer Methode, die das Einprägen auf der Basis einer Landkarte verlangte. Entgegen der Erwartung gelang es jedoch nicht, den Vorteil des Selbsttestens deutlich in der subjektiven Bewertung der Übenden zu bestätigen. Ziel der vorliegenden Studie war es, die Behaltensergebnisse zu replizieren und die Überprüfung der subjektiven Einschätzungen zu verbessern. Beide Übungsmethoden bewirkten erwartungswidrig vergleichbare Behaltenswerte. Die Studierenden bewerteten jedoch die Unterrichtsqualität von selbst gesteuertem Testen mit Feedback eindeutig höher und bevorzugten bei freier Übungsauswahl das selbst gesteuerte Testen vor dem Einprägen mit der Landkarte.

Schlagworte: Übung, Testen, Einprägen, Feedback, Drill, clickable map, study only.

Einleitung, Ausgangslage und Zielsetzung

Die bereits in den früheren Studien dargelegten theoretischen Ausführungen zur Wirksamkeit von Aufgaben mit Feedback führten zu der Schlussfolgerung, als ernsthafte und faire Vergleichsbasis für Testen mit Feedback eine Übungsmethode auszuwählen, welche die gleichen Informationen anbietet, diese aber als direkte Antworten und ohne eigene Erinnerungsbemühung. In der Zwischenzeit liegen etliche empirische Studien vor, die den Vorteil des Testens mit Feedback gegenüber dem Einprägen der korrekten Antwort [study only, read statement] belegen konnten [z.B.: [LaPorte & Voss 1975](#), [Carrier & Pashler 1992](#), [Cull 2000](#), Carpenter, Pashler & Vul 2006, Carpenter & Pashler 2007, Pavlik 2007]. Jacobs (2006) fand empirische Evidenz für das Lehrziel "Erlernen der Bundesstaaten der USA". In einem Wiederholungsdesign erzielten Studenten unter 3 verschiedenen Test- und Feedbackbedingungen stets numerisch höhere, in zwei Fällen signifikant bessere Behaltenswerte als unter einer Methode, welche lediglich erneutes Einprägen vorsah. Zwischen den einzelnen Testübungsmethoden (Multiple Choice, Short Answer, Covert Short Answer) konnten hingegen keine klaren Unterschiede gefunden werden.

In einem klassischen Randomisierungsdesign wollte [Jacobs \(2007\)](#) die Befunde replizieren und eine neue Testmethode prüfen. Das Einprägen der US-Bundesstaaten mit Hilfe einer Landkarte erwies sich dabei aber erwartungswidrig als genauso wirksam wie zwei Computer gesteuerte Test- und Feedbackvarianten. Lediglich das selbst gesteuerte Testen, dort Selbsttesten genannt, war dem Einprägen signifikant überlegen. Beim Selbsttesten entscheidet der Übende, welchen Bundesstaat er wie häufig testen will. Durch Anklicken auf das Staatsgebiet kann er den Namen des Staates als Feedback einsehen. Allerdings gelang es wegen methodischer Unzulänglichkeiten nicht, den Vorteil des Selbsttestens auch in der subjektiven Bewertung der Studenten sicher zu bestätigen, wenngleich auch eindeutige Belege für die Präferenz des Selbsttestens im Vergleich zum Einprägen vorgelegt wurden.

Vorliegende Studie verfolgt als bedeutsamstes Ziel, den bei Jacobs (2007) gefundenen Behaltensunterschied zwischen Selbsttesten und Einprägen zu replizieren. Das Experiment von Jacobs (2007) basierte auf einer ziemlich geringen Probandenanzahl. Auf der einen Seite erschwert dieses Manko den Nachweis eines tatsächlich vorhandenen experimentellen Effektes.

Man darf sich freuen, wenn er dann trotzdem gefunden wird. Auf der anderen Seite nährt die geringe Probandenzahl Zweifel an der Zuverlässigkeit eines signifikanten Unterschieds. Eine erfolgreiche Replikation würde zu einer bedeutsamen Klärung der Situation beitragen. Da jedoch auch bei dieser Replikation zu wenige Probanden für ein wünschenswertes klassisches Randomisierungsdesign ohne Vortest zur Verfügung standen, wurde trotz theoretischer Bedenken ein Vortest verwendet, um durch den Vergleich der Zuwachsraten die anfängliche Personenstreuung zu reduzieren und die statistische Effizienz zu steigern. Durch die Verwendung des Vortest erhält nun aber auch die Bedingung "Einprägen der Bundesstaaten via Landkarte" eine Testung, wodurch dann kein ausschließliches Einprägen in Reinkultur mehr vorliegt. Nur wenn die Vortestung relativ geringe Erfolgsquoten nach sich zieht, kann der Behaltenseinfluss dieser Vortestung in Grenzen gehalten werden.

Die Erhebung zur subjektiven Bewertung der Übungsmethoden bei Jacobs (2007) bezog sich auf das gesamte Übungsszenario, obgleich die experimentelle Variation nur etwa ein Drittel der gesamten Übungssitzung umfasste, was möglicherweise eine Erklärung dafür bietet, warum sich die Methodeneffekte nicht mehr klar heraus kristallisieren konnten. Aus diesem Grund sollte die Erhebung zur subjektiven Bewertung der beiden Übungsmethoden stringenter auf die Wahrnehmung der Experimentalphase gerichtet werden.

Die empirische Untersuchung

Versuchspersonen

An der Untersuchung nahmen insgesamt 47 Studierende der Universität des Saarlandes teil. Das Durchschnittsalter betrug 23 Jahre. Lehramtsstudierende, die an Seminaren des Verfassers teilnahmen, wurden im Rahmen der jeweiligen Lehrveranstaltung zur Teilnahme am Experiment motiviert und absolvierten das Lernexperiment in den Cip-Räumen der Philosophischen Fakultäten der Universität des Saarlandes. An dieser Stelle gilt mein Dank allen beteiligten Studierenden, die ca. eine Stunde ihrer Zeit unentgeltlich der Wissenschaft zur Verfügung stellten.

Versuchsablauf

Abbildung 1 : Versuchsablauf

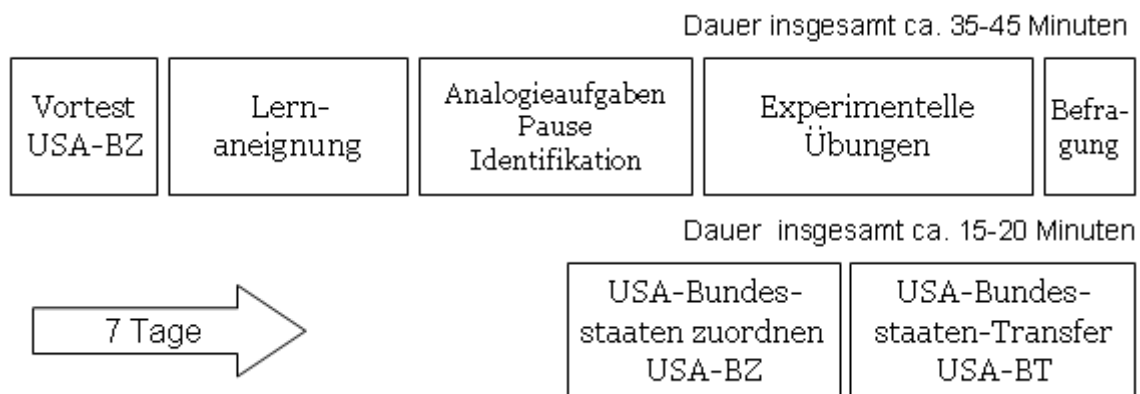


Abbildung 1 gliedert den Versuchsablauf in zwei 2 Sitzungstermine. Der erste Termin diente der Vortesterhebung, Lernaneignung und den experimentellen Übungsmethoden, der zweite Termin den Nachmessungen zum Zwecke der Überprüfung des Behaltens. Die Lernaneignung

entspricht exakt dem [Vorgehen bei Jacobs \(2007\)](#). Um die ökologische Validität der Übung zu erhöhen, stand in der Anweisung unmittelbar vor jeder experimentellen Übung analog zu Jacobs (2007), aber abweichend zu Jacobs (2006) der Hinweis, der Lerner könne sich auch Notizen machen. Der Untersuchungsleiter ermutigte die Studenten bei Unklarheiten, so zu lernen, wie sie es für sinnvoll erachteten.

Nach der experimentellen Übung folgte eine kurze Befragung zur Einschätzung des experimentellen Teils der Übung. Am Ende wurde den Studenten mitgeteilt, in einer Woche fände ein weiterer Untersuchungstermin statt, aber keinerlei Information gegeben, was dort erfasst werden sollte. Die Nachtests fanden exakt eine Woche später zur gleichen Zeit statt. In Abweichung zu Jacobs 2007 wurde die Reihenfolge der Behaltenstests verändert.

Die Studierenden wurden nach Zufall auf die experimentellen Bedingungen aufgeteilt. Tabelle 1 zeigt einen klassischen Randomisierungsplan mit Vortest.

Tabelle 1: Formalisierung des Versuchsplan

R	O	X ₁	O
R	O	X ₂	O

Der Vortest bezieht sich nur auf den USA-BZ. Bei der Auswertung der Daten werden die Posttestwerte der Gruppen sowie der Behaltenszuwachs als Differenz (Posttest-Vortest) berücksichtigt.

Die experimentellen Übungsmethoden

Tabelle 2 beschreibt die experimentellen Übungsmethoden und verweist auf die entsprechenden Programme, welche einen genauen Einblick in den Ablauf der jeweiligen Bedingung erlauben. Beide Übungsmethoden wurden von Jacobs 2007 übernommen und dort etwas ausführlicher besprochen. Die verwendete Grafik stammt aus:

<http://www.jayzeebear.com/map/usa.html> [18.5.2005]. Ich danke dem Konstrukteur der Grafik für die Erlaubnis, diese für wissenschaftliche Zwecke verwenden zu dürfen.

Tabelle 2: Beschreibung der Übungsmethoden

Übungsmethode (siehe auch Anhang)	Userkennung: Alabama Passwort: Alabama
Erneutes Studieren mit der Landkarte: Study only (SO): Auf dem Bildschirm erscheint die Karte der USA-Bundesstaaten einschließlich der Namen der Bundesstaaten in den entsprechenden Territorien. Der Lerner hat die Aufgabe, mit einer ihm sinnvoll erscheinenden Strategie die Bundesstaaten einzuprägen.	Studieren via konventioneller Landkarte
Selbsttesten mit CSA (Selbsttesten) Auf dem Bildschirm ist eine Karte mit allen Bundesstaaten der USA ohne Namen zu sehen. Der Lerner soll ein beliebiges Staatsgebiet auswählen und versuchen, sich an den Namen zu erinnern. Durch Anklicken des entsprechenden Gebietes kann er den Namen des Staates einsehen. Zusätzlich hat der Lerner die Option jederzeit in einen SO-Modus zu wechseln und auf einen Blick alle Staatsnamen in den entsprechenden Gebieten zu sehen.	Selbsttesten

*Der echte Versuch lief im Vollbildmodus ohne Menu-, Symbol-, oder Statuszeilen.

Die Programme dienen nur als Demonstration. Das JavaScript-Programm ist unverständlich und beinhaltet etliche Eigentümlichkeiten.

Diesmal fertigten besonders viele Studenten Notizen an, vornehmlich unter der Bedingung "Einprägen der Bundesstaaten via Landkarte". Die Notizen wurden am Ende des Versuchs vom Versuchsleiter eingezogen. Einige im Anhang dargestellte Notizen zeigen das Bemühen um eine aktive Lernstrategie, die keineswegs als stures Auswendiglernen klassifizierbar wäre. Im Gegensatz zu einigen Laborexperimenten, welche die Art und Vorgabe der einzuprägenden Information strikt reglementieren, wurde hier ein großer Freiheitsspielraum gewährt, die Bundesstaaten nach eigenen Vorstellungen einzuüben. Theoretisch mag es Sinn machen, eine Kontrollbedingung nach ganz bestimmten Vorgaben einschränkend zu gestalten, auch wenn sich in der Praxis kein Lerner daran halten würde. Aus praktischer Perspektive lohnen sich neue Lernmethoden nur dann, wenn sie bisher bewährte Lernmethoden an Effizienz übertreffen.

Abhängige Variablen

A) USA-Bundesstaaten-zuordnen-können (USA-BZ)

Im USA-BZ wurde der Name eines Bundesstaates vorgegeben und der Proband hatte die Aufgabe, das entsprechende Territorium auf der Landkarte mit der Maus anzuklicken. Auf diese Weise wurden stets alle Bundesstaaten der USA in zufälliger Reihenfolge getestet.

B) USA-Bundesstaaten Transfer Test (USA-BT)

Im Transfertest wurden Aspekte geprüft, die in den experimentellen Übungen niemals explizit eingeübt oder gezielt getestet wurden, sich beim Einüben aber mehr oder weniger implizit anbieten könnten oder sich aus dem Wissen der Zuordnung ableiten lassen. (genauere Angaben siehe [Jacobs 2007](#), Abhängige Variablen)

Alle Aufgaben des USA-BT waren auf einer Seite für alle Probanden in der gleichen Reihenfolge angeordnet. Alle Probanden beantworteten zunächst den BZ und anschließend den BT.

Die hier ermittelte Reliabilität des BZ (Cronbachs $\alpha = .93$), die des BT ($\alpha = .84$) sowie der Zusammenhang beider Variablen ($r = .89$) entsprechen den Ergebnissen von Jacobs (2007). Die Retestkorrelation zwischen BZ vor der Übung und BZ nach einer Woche beträgt $r = .56$, ist jedoch wegen des relativ geringen Ausgangswissens wenig aussagekräftig. Beide Behaltenswerte korrelieren signifikant positiv mit ihren Bearbeitungszeiten (BZ: $r = .41$, BT: $r = .43$). Sorgfältiges Beantworten zahlt sich offenbar aus. Möglicherweise wollten etliche Studenten die Testung möglichst schnell abschließen und haben dabei einige Fehler in Kauf genommen. Bei einer leistungsabhängigen Bezahlung verschwindet dieser Effekt vielleicht. (Jacobs 2007b)

Ergebnisse

Vergleichbarkeit der Gruppen

Beide experimentelle Gruppen erzielten hoch vergleichbare Ergebnisse im mitgeteilten Abiturnotendurchschnitt (2,4 vs. 2,4) sowie dem Alter (23,8 vs. 23,4) und bestätigen so die Effizienz der Randomisierung. Der Abiturnotendurchschnitt steht in keinem signifikanten Zusammenhang zum BZ-Vor- oder BZ-Nachtest.

Vortestergebnisse:

Bei Jacobs 2006 und 2007 wurde auf den Vortest verzichtet, um bei der Gruppe "ausschließliches Studieren [Landkarte]" jede Form der Testung zu verunmöglichen. Da für die vorliegende Studie ähnliche Probanden (z.B. Studierende aus den Seminaren des Verfassers) rekrutiert wurden, erlauben die Vortestergebnisse eine angenäherte Einschätzung des Vorwissens bei den früheren Studien. Tabelle 3 stellt die Ergebnisse für alle Probanden dar.

Tabelle 3: Mittelwert, Median und Streuung im Vortest "USA-BZ"

	N	M	Md	s
alle	45	14	10	11
ohne Ausreißer	42	12	10	8

Ohne Aussortierung der Ausreißer werden 14% aller Bundesstaaten korrekt zugeordnet. Nach Elimination der mittels Tukeys Boxplot ermittelten Ausreißer sinkt der Prozentsatz geringfügig auf 12%. Die Ergebnisse bestätigen insgesamt ein relativ geringes Vorwissen beim Zuordnen der Namen zu den entsprechenden Territorien auf der Landkarte. Wegen der Zufallszuweisung wurden gleiche Vortestwerte für beide Experimentalgruppen erwartet, was sich sowohl mit als auch ohne Ausreißer bestätigen ließ. Da die deutlich höheren Erfolgsquoten der 3 Ausreißer auf früheres Einüben hindeuten, wurden sie für den Vergleich der Behaltensdaten beider Übungsmethoden ausgeschlossen.

Lernprozessdaten der Bedingung Selbsttesten

Die Übungsmethode Selbsttesten gewährte die Option, jederzeit vom Testmodus in den reinen Studiermodus der Landkarte und wieder zurück zu wechseln. Tabelle 4 stellt Mittelwerte und Mediane einiger Prozessdaten von 23 Probanden dar. Aus ihr wird erkennbar, dass die Studenten im Durchschnitt ca. 2 bis 3 mal in den Studiermodus wechselten und dort insgesamt ca. 2 Minuten verweilten. Ca. 80% der gesamten Übungszeit der Methode Selbsttesten wurde für das direkte Selbsttesten der Bundesstaaten verwendet. In dieser Testzeit klickten die Studenten im Median 134 mal auf einen Bundesstaat, um als Feedback dann den Namen im entsprechenden Territorium einzusehen.

Tabelle 4: Prozessdaten für die Bedingung Selbsttesten (N=23)

	Anzahl der getesteten US-Bundes- staaten	Häufigkeit des Wech- sels zur Landkarte	Gesamtdauer für das Kar- tenstudium in Sekunden	Prozentualer Zeitanteil für Testen mit Feedback
Mittelwert	169	2,7	151	79 %
Median	134	2,0	100	86 %

Im Vergleich zu Jacobs (2007) fällt die Verhaltenspräferenz für das Testen gegenüber dem Kartenstudium etwas geringer aus. Jedoch kann auch hier keinerlei Zweifel aufkommen, Studierende würden das Testen mit Feedback dem Einprägen via Landkarte vorziehen, wenn ihnen diese Möglichkeit gegeben wird. Die Präferenz für das Testen hängt natürlich auch vom bisher erworbenen Wissen ab. Daten von Kornell (2007) zufolge verändern sich die selbst gewählten Anteile von Studiermodus und Testmodus mit zunehmenden Übungsdurchgängen immer mehr zugunsten des Testens. In vorliegender Studie haben die Studierenden in der Lernaneignungsphase bereits 2 Lerndurchgänge hinter sich. Sie beginnen die experimentelle

Bedingung Selbsttesten im Durchschnitt mit einer Erfolgsquote von ca. 50% korrekter Zuordnungen und ab dieser Erfolgswahrscheinlichkeit macht es durchaus Sinn, mit dem Testen zu beginnen, zumal bei Fehlern Feedback gewährt wird.

Behaltenswerte

Theoretische Erwartungen und bisherige empirische Ergebnisse ließen eindeutige Behaltensvorteile für die Methode Selbsttesten erwarten. Zunächst wird die Hypothese durch einen Vergleich der Übungsmethoden für beide Behaltensposttests auf Signifikanz geprüft. Anschließend wird beim Test "Bundesstaaten zuordnen können" die Differenz zwischen Posttest und Vortest gebildet, welche den Lernzuwachs im Zeitbereich von der Übung bis eine Woche nach der Übung erfasst und eine sensitivere Testung ermöglicht.

Tabelle 5: Prozentsatz korrekter Lösungen im Posttest sowie Leistungszuwachs im BZ

[N Landkarte = 18 ; N Selbsttesten = 22]

		M	s	t-Test und p
Bundesstaaten zuordnen BZ- Posttest	Landkarte	44.7	22.4	t(38)=.89; p=.38
	Selbsttesten	39.3	15.9	
Bundesstaaten Transfer BT-Posttest	Landkarte	52.2	18.2	t(38)=.49; p=.63
	Selbsttesten	49.5	15.7	
BZ-Zuwachs	Landkarte	33.0	17.2	t(38)=.92; p=.36
	Selbsttesten	28.3	15.1	

Wie die Tabelle 5 verdeutlicht, erzielten Studierende der Übungsmethode Einprägen mit Landkarte stets numerisch höhere Behaltenswerte als Studierende der Übungsmethode Selbsttesten. Die Überprüfung mit t-Test ergab in allen Fällen keine signifikanten Unterschiede im Behalten. Die zusätzlich berechnete Wechselwirkung zwischen experimentellen Methoden und Zeitpunkten (Vor- und Posttest) führte erwartungsgemäß zum gleichen Ergebnis ($p=.36$) wie der Vergleich des BZ-Zuwachses. Auch die Testzeiten der Übungsgruppen fallen vergleichbar aus. Alle durchgeführten Vergleiche widersprechen der aufgestellten Hypothese. Die Replikation des signifikanten Behaltensvorteils des Selbsttestens gegenüber dem ausschließlichen Einprägen auf der Grundlage einer Landkarte von Jacobs (2007) ist damit gescheitert.

6 bis 7 Wochen nach dem Posttest wurden die Probanden via Email gebeten, den Test "US-Bundesstaaten zuordnen" über das Internet von zu Hause aus zu beantworten. 24 Probanden kamen der Aufforderung nach. Die Behaltenswerte von 3 Probanden mussten ausgeschlossen werden, da ihre Follow-up-Werte über 30 Prozent höher ausfielen als im Posttest, was nur mit erneuter (unaufgeforderter) Übung erklärbar ist. Bei den verbliebenen 21 Studierenden fielen alle Follow-up-Behaltenswerte bis auf eine Ausnahme niedriger aus als die Posttestwerte, und zwar im Mittel um 17%. Die Retestreliabilität der US-BZ beträgt .83 ($N=21$). Zum Follow-up-Test war offenbar eine leicht positiv verzerrte Studententstichprobe angetreten, die im Posttest 53% und 6-7 Wochen später im Follow-up-test 36% der Bundesstaaten zutreffend identifizieren konnte. Zwischen den experimentellen Gruppen konnten auch im Follow-up-test keine signifikanten Unterschiede festgestellt werden (Landkarte: $N=10$, $M=34,8$, $s=16$ und Selbsttesten: $N=11$, $M=36,9$, $s=14,9$); $t(19) = -0.31$ $p=.76$)

Typischer Lern und Behaltensverlauf bei den Untersuchungen von Jacobs

In Tabelle 6 wurde der Versuch unternommen, den typischen Lern- und Behaltensverlauf der Untersuchungen von Jacobs zum Einprägen der US-Bundesstaaten zu skizzieren. Die verwendeten Daten vor der Übung und nach einer Woche entstammen dieser Untersuchung, während die restlichen Daten Schätzungen aus früheren Studien (Jacobs 2006,2007) repräsentieren.

Tabelle 6: Prozentsatz korrekter Zuordnungen vor, während und nach den Übungen

	% korrekt
vor der Übung	12
nach Lernaneignung und unmittelbar vor den experimentellen Übungen	50
unmittelbar nach den experimentellen Übungen	70
nach einer Woche	42

Zu Beginn der Untersuchung verfügen die Studierenden über ein relativ geringes Vorwissen von ca. 12 Prozent korrekter Zuordnungen der USA-Bundesstaaten. Mit Einbezug der Ausreißer liegt dieser Prozentsatz etwas höher. Aus den Studien von Jacobs 2006 und 2007 ließ sich die Erfolgsquote nach der Lernaneignungsphase schätzen. Die dort realisierte zweimalige Präsentation aller USA-Bundesstaaten führte am Ende zu ca. 50 % korrekter Lösungen. Erst dann begannen die experimentellen Übungen. Jacobs 2006 fand für verschiedene Übungsmethoden unmittelbar nach den experimentellen Übungen sehr konsistente Erfolgsquoten von ca. 70 Prozent. Danach setzt das Vergessen ein und nach einer Woche behalten die Studenten hier im Mittel noch ca. 42%. Einige Testübungsmethoden aus früheren Untersuchungen erreichten um die 50% korrekter Lösungen im Posttest. Die best mögliche Schätzung für ein Behalten ca. 8 Wochen nach der Übung liegt bei ca. 25 - 30%.

Hier kennen die Studierenden eine Woche nach der Übung ca. 30% mehr Bundesstaaten als zu Beginn. Die Differenz zwischen BZ-Posttest und BZ-Vortest ist hoch signifikant und entspricht einer Effektstärke von $d = 2.13$. Der Behaltenszuwachs ist im Übrigen unabhängig vom Ausgangsniveau. Die Behaltenswirksamkeit der Übungen steht damit außer Frage. Allerdings spielt es offensichtlich keine Rolle, mit welcher Methode die Staaten eingeübt wurden.

Subjektive Einschätzungen.

Im Wiederholungsexperiment von Jacobs (2006) hatten die Studierenden Erfahrung mit allen dort verwandten Übungsmethoden. Im Vergleich der einzelnen Übungsmethoden untereinander ließen sich klare Präferenzen für bestimmte Übungsvarianten nachweisen. Alle Testmethoden wurden im Hinblick auf ihre Lernförderung günstiger bewertet als wiederholtes Einprägen. Im Randomisierungsexperiment von Jacobs (2007) hingegen übte jeder Student nur mit einer ganz bestimmten Methode und musste am Ende eine Bewertung des gesamten Tutoriums abgeben. Dort konnten keine klaren Unterschiede zwischen verschiedenen Übungsvarianten gefunden werden, wofür die mangelnde Spezifizierung der Fragestellung auf den rein experimentellen Teil verantwortlich gemacht wurde. In dieser Studie sollte die subjektive Bewertung stringenter auf die experimentellen Varianten fokussiert werden in der Hoffnung, dann klare subjektive Präferenzen für Selbsttesten gegenüber Kartenlesen zu erhalten.

Im Anschluss an die Übung sollten die Studenten daher nur die letzten 12 Minuten (=Experimentalphase) auf einer 7 stufigen Skala (stimmt genau (7).... stimmt überhaupt nicht (1)) hinsichtlich etlicher Items einschätzen, die Rückschlüsse auf die Wertschätzung der Übungsmethode und die Motivation bei der Bearbeitung der experimentellen Bedingung zu-

lassen sollten. Im Gegensatz zur früheren Untersuchung, die eine Einschätzung der gesamten Übung (Lernaneignungsphase, Analogieaufgaben und experimentelle Phase) nahe legte, wurden die Studenten gleich zweimal darauf aufmerksam gemacht, ihre subjektiven Einschätzungen auf die spezielle Übungsvariante zu beziehen. Der Itembewertung lag folgender Vorspann zugrunde. Ich bewerte die letzten 12 Minuten des Übungsprogramms als.....

Tabelle 7: Vorteile von Selbsttesten gegenüber Study only via Landkarte (nach t-Test)

Item	t (44)	p (zwei seitig)
interessant	,97	,34
motivierend	1,24	,22
anregend	,71	,48
herausfordernd	-,27	,79
anstrengend (-)	2,27	,03
gute Übungsmöglichkeit	3,05	,004
geeignete Übungsmethode	2,71	,010
pädagogisch empfehlenswert	1,90	,064
behaltenswirksam	1,04	,30
langweilig (-)	1,97	,06
stupider Drill (-)	,51	,62
einfallsloses Pauken (-)	2,03	,048
ermüdend (-)	,85	,40

positiver t-Test bedeutet in richtige Richtung gepolte Vorteile für Selbsttesten.

die fettgedruckten Items sind einseitig auf dem 5% signifikant zugunsten der Methode Selbsttesten.

Tabelle 7 zeigt die erhobenen Items sowie die Ergebnisse der Mittelwertsvergleiche zwischen Selbsttesten und Landkartenstudium. Einige Items wurden auf der Basis theoretischer Überlegungen und statistischer Kennwerte zu einem Testwert zusammengefasst. Die subjektive Qualität der Unterrichtsmethode (Cronbachs $\alpha = .86$) setzt sich zusammen aus den Items "gute Übungsmöglichkeit, geeignete Übungsmethode, pädagogisch empfehlenswert und behaltenswirksam. Zur Lernmotivation/Interesse (Cronbachs $\alpha = .88$) zählen die Items interessant, motivierend, anregend, herausfordernd und langweilig. Tabelle 8 fasst die Ergebnisse des Methodenvergleichs zusammen.

**Tabelle 8: Bewertung von Übungsqualität und Übungsmotivation für beide Übungs-
methoden. [N für beide Gruppen (20-23)]**

	Variante	M	s	t-Test	Effektstärke
Eingeschätzte Qualität der Übungsmethode	Landkarte	4,13	1,26	t(42)=2,76 p=.0.008.	d = .83
	Selbsttesten	5,05	,94		
Eingeschätzte Motivierung der Übungsmethode	Landkarte	3,93	1,42	t(43)=1,15, p= .245	d = .34
	Selbsttesten	4,39	1,27		

Die subjektive Qualität der Unterrichtsmethode Selbsttesten wird eindeutig besser eingeschätzt als die des Kartenstudiums. Hinsichtlich der Lernmotivation sind keine signifikanten Unterschiede zwischen den Übungsmethoden nachzuweisen. Üben durch reines Kartenstudium (5.0) wird eindeutig anstrengender als Selbsttesten (4.1) erlebt (Effektstärke $d = .69$). Die Behaltenswerte im BZ- und BT-Nachtest stehen in keinem signifikanten Zusammenhang mit der Einschätzung der Unterrichtsqualität, der Lernmotivation oder der erlebten Anstrengung. In ihrer Gesamtheit lassen die Daten keinen Zweifel daran aufkommen, aus der Sicht der Übenden sei Selbsttesten die bessere Übungsmethode.

Zusammenfassung und Diskussion

Es ist nicht gelungen, den bei Jacobs (2007) gefundenen Behaltensvorteil des Selbsttestens gegenüber dem Einprägen zu replizieren. Beide Übungsmethoden führten objektiv zu vergleichbaren langfristigen Behaltenswerten. Selbsttesten wurde jedoch subjektiv eindeutig als die bessere Übungsmethode eingeschätzt. Die Studierenden empfanden das Einprägen mit Hilfe der Landkarte weiterhin als anstrengender, langweiliger und einem einfallsloses Pauken ähnlicher. Unter der Methode Selbsttesten konnten die Lerner zwischen Testen und Einprägen wählen. Ca. 80% der gesamten Übungszeit fielen auf das Testen mit Feedback. Diese Verhaltenspräferenz sowie die subjektive Bevorzugung des Selbsttestens sprechen ziemlich deutlich dafür, Studierende würden sich bei freier Auswahl für das Selbsttesten entscheiden und ein entsprechendes Übungsangebot begrüßen. Möglicherweise fördert das Selbsttesten bei freiwilligem Lernen eher die spontane Nutzung.

Sowohl beim Einprägen der Bundesstaates via Landkarte wie beim Selbsttesten muss der Übende bestimmen, wie er im Einzelnen vorgehen soll. Beide Lernmethoden vertrauen darauf, die eigene Lernsteuerung orientiere sich an vernünftigen Kriterien. Unter idealistischer Perspektive weiß der einzelne Lerner am besten, wie er sein Lernen organisieren soll. Bei Jacobs (2007) erzielten Studierende mit Selbsttesten numerisch bessere Behaltenswerte als Studierende unter einer speziellen, Computer gesteuerten Flashcard-Variante. Diese gab zunächst alle Staaten nacheinander zur Testung mit Feedback vor, fügte die falsch gelösten Aufgaben an das Ende der Aufgabenliste und bot sooft die falsch erinnerten Staaten zum Testen an bis sie richtig erinnert wurden. Nach der Region of proximal learning-Theorie (z.B. Kornell & Bjork 2007) werden bei beschränkter Übungszeit diejenigen Items bevorzugt eingeübt, welche man nicht sicher beherrscht, die aber durch erneute Einübung eine hohe Lernerfolgswahrscheinlichkeit versprechen. Bei diesen ist das Verhältnis von Lerneinsatz und Lernertrag am höchsten. Sehr schwere Items erfordern demgegenüber immensen Zeitaufwand bei beschränkter Erfolgswahrscheinlichkeit, was eigentlich gegen konsequentes Flashcard in der Fassung von Jacobs (2007) spricht. Bereits beherrschte Items lassen keinen weiteren Lernfort-

schritt vermuten. Möglicherweise gehen viele Studierende mehr oder weniger konsistent dieser Strategie nach und erzielen so ähnlich gute Ergebnisse wie unter einer Rechner gesteuerten Übungsvariante.

Obwohl Testen mit Feedback langfristiges Behalten besser stärken sollte als wiederholte Informationsaufnahme, konnte dieser Effekt hier gar nicht und bei den früheren Studien von Jacobs nicht durchgängig bestätigt werden. Dies lässt zum einen vermuten, dass der tatsächliche Unterschied zwischen beiden Übungsvarianten nicht sehr hoch ausfällt. Zum andern kommt es vermutlich darauf an, wie gut die Präsentation der Information vom Lerner selbst verarbeitet und im Langzeitgedächtnis abgelegt wird. Je mehr sich diese Verarbeitung von lockerem Durchlesen hin zu einer aktiven Strukturierung und Elaboration des Wissens entwickelt, desto besser kann eine Präsentation der Information auch behalten werden. Dies ist beim Einprägen der Bundesstaaten der USA aber vermutlich deutlich schwieriger als beim Erlernen eines Textes, der in der Regel viele Verknüpfungspunkte anbietet. Auch das Testen erzwingt nicht automatisch besseres Behalten. Hier spielt das ernsthafte Bemühen, die Information aus dem Langzeitgedächtnis hervorzurufen, eine wichtige Rolle. Ich war stellenweise verblüfft über die hohe Geschwindigkeit, mit der manche Probanden die Staaten anklickten. Das Feedback der korrekten Lösung bei einem Fehler muss sorgsam enkodiert werden, um später auch wieder erinnert werden zu können. Die für jene Verarbeitungsprozesse notwendigen Fähigkeiten und Strategien der einzelnen Personen bestimmen das Behalten wesentlich mehr als die verschiedenen Übungsmethoden. Je weniger der Freiheitsspielraum des Einzelnen eingeengt wird, desto schwieriger lassen sich Methodeneffekte nachweisen. Deshalb können streng kontrollierte Laborexperimente auch nicht einfach auf schulische Praxis übertragen werden (siehe dazu auch Jacobs 2008).

Rein pragmatisch legen die Ergebnisse dieser sowie der früheren Studien von Jacobs folgende pädagogischen Schlussfolgerungen nahe. Mit Hilfe einer konventionellen Landkarte lassen sich die Bezeichnungen bestimmter Areale relativ gut einprägen. Dieses sparsame Medium erfordert vom Lerner meist wiederholtes Einüben vorgegebener Information. Gelegentlich bewirkt das Testen mit Feedback jedoch einen etwas höheren Behaltenseffekt. Dieser lässt sich aber nicht einfach nachweisen. Er wird etwa beim Methodenvergleich zweier Schulklassen nicht direkt augenfällig in Erscheinung treten. Eine Computer gesteuerte Itemvorgabe sowie eine formale Testung mit klassischen Aufgabenformaten sind für den Behaltenserfolg nicht zwingend erforderlich. Der Lerner kann die Lernsteuerung beim Testen durchaus selbst übernehmen, indem er bestimmt, welche Items er wie häufig testen will. Hierzu genügt ein Testen im Geiste, wenn anschließend "Knowledge of correct response" folgt (=Covert short answer mit KCR). Unabhängig von einem möglichen objektiven Behaltensvorteil bewerten die Lerner ein Testen mit Feedback als lernwirksamer und pädagogisch wertvoller.

Literatur

- Carpenter, S. K., Pashler, H., Vul, E. (2006). What types of learning are enhanced by a cued recall test. *Psychonomic Bulletin & Review*, 13 (5), 826 - 830.
- Carpenter, S. K. & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, 14 (3), 474-478
- Carrier, M. & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 632-642
- Cull, W. L. (2000). Untangling the Benefits of Multiple Study Opportunities and Repeated Testing for Cued Recall. *Appl. Cognit. Psychol.* 14, 215-235

- Jacobs, B. (2006). Erneutes Studieren oder Testen mit Feedback beim Einüben Von Faktenwissen am Beispiel des Erlernens der Bundesstaaten der USA.
 URN: <urn:nbn:de:bsz:291-psydok-5992>
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2006/599/>
- Jacobs, B. (2007). Die Behaltenswirksamkeit wiederholten Einprägens im Vergleich zu Computer- und selbst gesteuertem Testen mit Feedback.
 URN: <urn:nbn:de:bsz:291-psydok-26913>
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2010/2691/>
- Jacobs, B. (2007b). Geld und Noten als extrinsische Motivatoren zur Verbesserung kognitiver Leistungen.
 URN: <urn:nbn:de:bsz:291-psydok-9644>
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2007/964/>
- Jacobs, B. (2008). Gezieltes Studieren gelöster Aufgaben als alternative Übungsmethode zu Testen mit Feedback.
 URN: <urn:nbn:de:bsz:291-psydok-15597>
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2008/1559/>
- Kornell, N & Bjork, R.A (2007). The promise and perils of self-regulated study. Psychonomic Bulletin & Review 14 (2), 219-224
- Kornell, N. (2007). Choosing self-testing as a study strategy. Manuscript submitted for publication ,zitiert nach Kornell, N & Bjork, R.A (2007) The promise and perils of self-regulated study. Psychonomic Bulletin & Review, 14 (2), 219-224
- LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. Journal of Educational Psychology, 67, 259-266.
- Pavlik, P. I. (2007). Understanding and applying the dynamics of test practice and study practice. Instructional Science 35 (5), 407-441.
-

Anhang:

- a) Bildschirmkopien der experimentellen Bedingungen
- b) Notizen der Studierenden während des Lernens

Erneutes Studieren mit der Landkarte: Study only (SO):



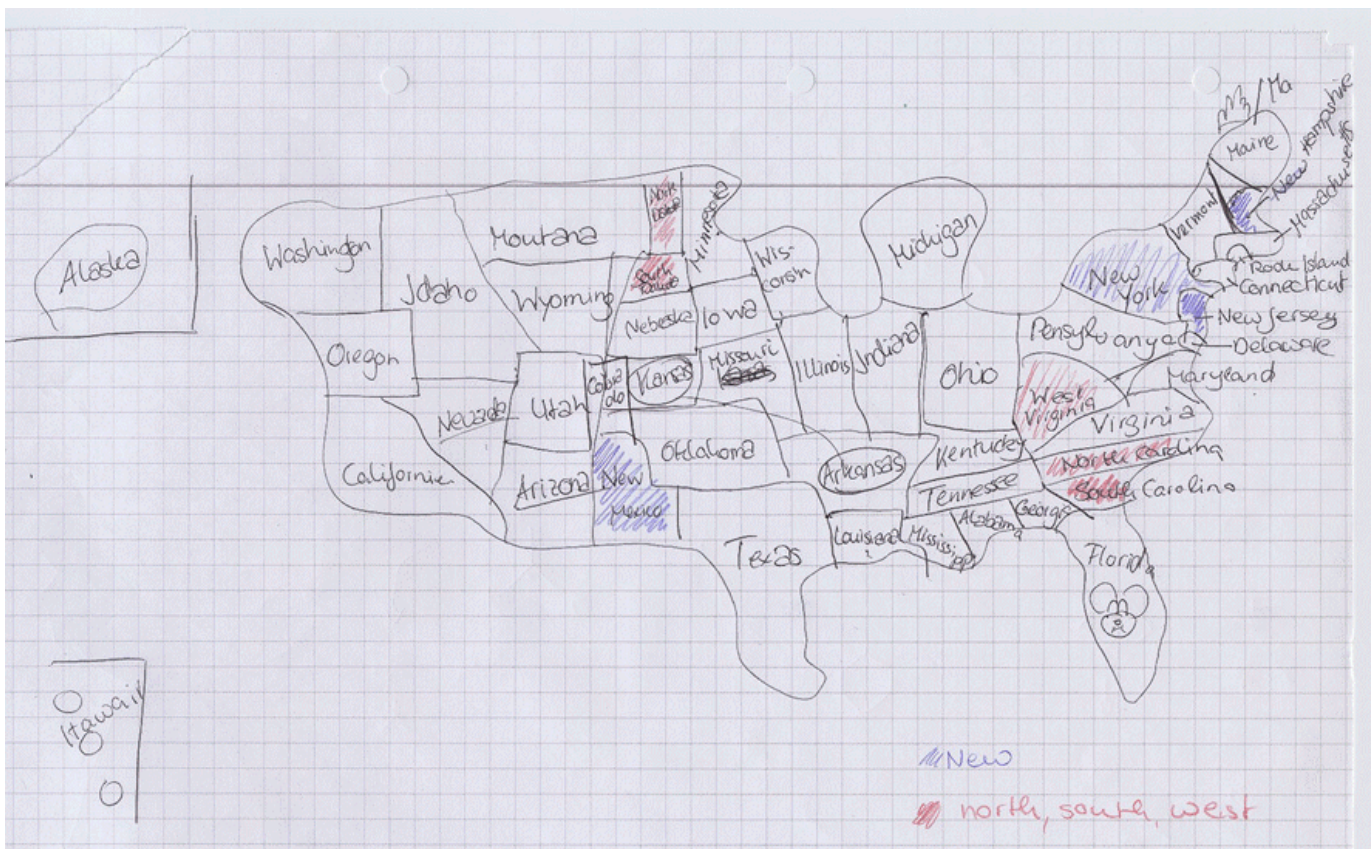
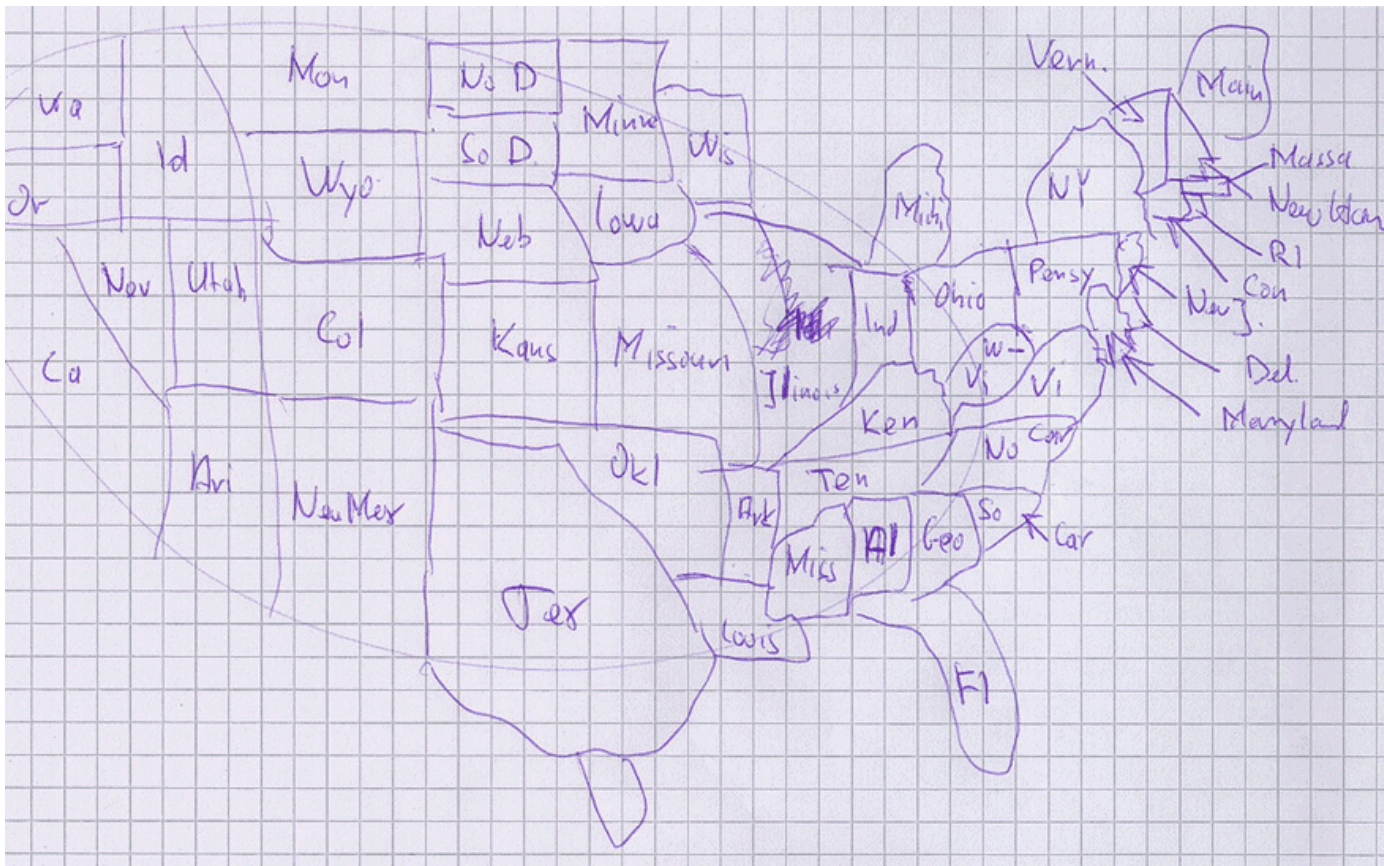
Der Lerner sieht nur diese Karte sowie die Anweisung
Wenden Sie eine sinnvolle Strategie an, sich möglichst viele Staaten einzuprägen

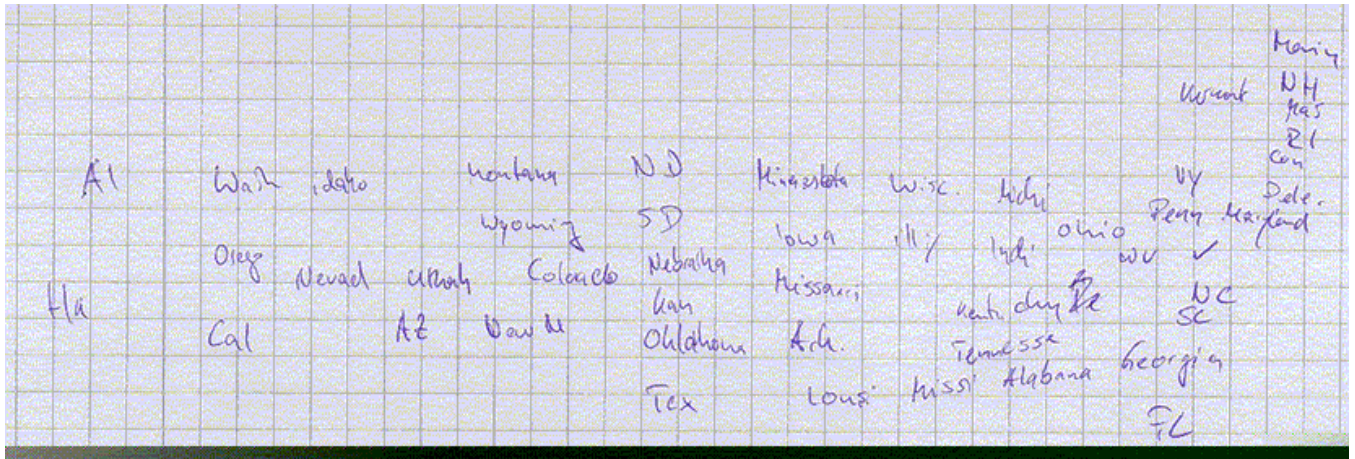
Selbsttesten mit CSA (Selbsttesten)



Der Lerner soll ein beliebiges Staatsgebiet auswählen und versuchen, sich an den Namen zu erinnern. Durch Anklicken des Gebietes von Wyoming sieht er hier den entsprechenden Namen. Beim Verlassen des Mauscur-sors aus dem Gebiet verschwindet der Name wieder und der Lerner kann das nächste Staatsgebiet auswählen. Zusätzlich hat der Lerner die Option, jederzeit in einen SO-Modus (siehe oben) zu wechseln und auf einen Blick alle Staatsnamen in den entsprechenden Gebieten zu sehen

Anhang: Notizen von Studenten während der Experimentalphase





Ausschnitte aus Notizen einiger Studenten

<p>Wyoming - grau NW</p> <p>Kansas - pink Mitte</p> <p>Nebraska - N von Kansas</p> <p>Minnesota - braun N</p> <p>Iowa - S von Minnesota</p> <p>Missouri</p> <p>Ark.</p> <p>Louis.</p> <p>Wisconsin - pink große Seen</p> <p>Illinois - S von Wisc.</p> <p>Indiana - östl. Illinois</p> <p>Kentucky</p> <p>Tennessee</p> <p>Mississippi - O von Louis.</p> <p>Alabama</p> <p>Georgia</p> <p>Florida</p> <p>South Carolina - gelb O</p> <p>Carolina - N von S. Car.</p> <p>Michigan - lila - gr. Seen</p> <p>Ohio - SO von Michigan</p>	<p>Sehr klein: Delaware Rhode Island</p> <p>klein: Connecticut New Jersey Maryland Massachusetts New Hampshire Vermont</p> <p>} East Coast</p>
	<p>Washington Oregon Nevada Arizona California</p> <p>Idaho Wyoming Montana North Dakota South Dakota Nebraska Kansas Oklahoma Colorado New Mexico Texas</p> <p>Utah</p> <p>Nevada</p> <p>California</p> <p>Arizona</p> <p>New Mexico</p> <p>Sehen</p>