

Zeitschrift für Interkulturellen Fremdsprachenunterricht

Didaktik und Methodik im Bereich Deutsch als Fremdsprache

ISSN 1205-6545 Jahrgang 19, Nummer 2 (Oktober 2014)

Der Online-Test BULATS Deutsch – ein geeignetes Instrument zur Messung von Lernfortschritten? Erste Ergebnisse im Rahmen des Projekts "Tablet-Klasse"

Kathrin Hahn

Goethe-Institut
Dachauer Str. 122
80637 München

E-mail: Kathrin.Hahn@goethe.de

Winfried Koch

BDS Koch
Bibienstraße 5
68723 Schwetzingen

E-mail: winfried.koch@bdskoch.de

Linda Fromme

Goethe-Institut
Dachauer Str. 122
80637 München

E-mail: Linda.Fromme@goethe.de

Abstract: Im Herbst/Winter 2013 wurde an drei Goethe-Instituten in der Region Nordwesteuropa der Einsatz von Tablets im Fremdsprachenunterricht pilotiert und innerhalb eines Forschungsprojekts wissenschaftlich begleitet und evaluiert. Es wurde analysiert, ob Tablets (hier iPads) für das Fremdsprachenlernen Potentiale bieten, besonders bei der Förderung des Hörverstehens und des Sprechens. Die Lernenden wurden für die Untersuchung in zwei Gruppen aufgeteilt: Eine Gruppe arbeitete mit Tablets im Unterricht (N=27), die andere ohne (N=21). Die Veränderungen im Bereich Hörverstehen wurde in beiden Gruppen durch einen Vor- und Nachtest mit dem BULATS Online-Test Deutsch gemessen. Die Testergebnisse wurden anschließend miteinander verglichen. Es blieb zunächst offen, ob die Tablets tatsächlich zu einer Verbesserung in bestimmten Bereichen des Fremdsprachenlernens führen oder ob eventuell durch den Fokus auf die Technik mit weniger Lernfortschritt oder einer Verschlechterung zu rechnen ist. Im Verlauf des Projekts zeigte sich, dass die Testgruppe (Tablet-Gruppe) erkennbar bessere Werte im BULATS-Test erzielte als die Kontrollgruppe.

In autumn/winter 2013 a rigorous research project on the use of tablets in the German foreign language classroom was conducted at three Goethe-Institutes in northwestern Europe. The purpose of the project was to examine the potential usefulness of tablets (iPads) in foreign language teaching and learning, especially when it comes to improving oral and listening skills. The participants were assigned to two different groups - one group working with tablets (N=27), the other one without (N=21). The participants' changes in their German listening skills were measured by testing and retesting with the BULATS Online-Test German. The results obtained by the two groups were then compared. The question was whether the use of tablets would lead to improvements in particular areas of learning foreign languages or whether it might be related to less progress in learning, or even a decline, possibly due to a disproportionate focus on the technology itself. It turned out that over the course of the project the test group using tablets showed demonstrably better results in the BULATS test than the control group.

Schlagwörter: Tablet, iPad, BULATS, Online-Test, computerbasiert, Hörverstehen

1. Das Projekt „Tablet-Klasse“

Der Einsatz von Tablets im (Fremdsprachen-)Unterricht wird immer beliebter. An Schulen und anderen Bildungseinrichtungen werden Tablet-Klassensätze angeschafft, um den Mehrwert, den die mobilen Geräte gegenüber herkömmlichen PCs bieten, zu nutzen. Ein Vorreiter in Deutschland ist die Kaiserin-Augusta-Schule, ein Gymnasium in Köln, an dem unter der Leitung von André J. Spang bereits seit Februar 2011 sehr erfolgreich mit Tablets im Unterricht gearbeitet wird (vgl. Spang 2013: 11). Das Land Niedersachsen führt seit September 2012 das Projekt „Mobiles Lernen mit Tablet-Computern“ durch. An insgesamt 24 Schulen in Niedersachsen wird bis zum Schuljahrsende 2015 der Einsatz von Tablets im Unterricht in einem dreijährigen Projekt pilotiert und evaluiert (vgl. Greb 2012). In Rheinland-Pfalz wurde unter der Leitung von Prof. Dr. Kerstin Mayrberger von Februar 2011 bis August 2013 das medienpädagogische Pilotprojekt „iPad im Einsatz – Education 2013“ durchgeführt (vgl. Ludwig, Mayrberger & Weidmann 2011).

Auch das Goethe-Institut hat im Laufe des Wintersemesters 2012/2013 an drei Auslandsinstituten in der Region Nordwesteuropa den Einsatz von Tablets (hier: iPads) im Fremdsprachenunterricht pilotiert und wissenschaftlich begleitet. Innerhalb des Pilotprojekts wurden verschiedene qualitative und quantitative Untersuchungen durchgeführt, die den Lernfortschritt der Lernenden im Hörverstehen und Sprechen und Veränderungen in der Lernermotivation sowie in der Lehrer-/Lernerrolle überprüfen sollten.

An dem Projekt waren die Goethe-Institute in London, Dublin und Amsterdam beteiligt. Tablets wurden bis dahin noch nicht gezielt an den Instituten eingesetzt. Sowohl das technische Setting in den Klassenräumen als auch die Erstellung der Materialien zur Arbeit mit den Geräten und die Auswahl der Apps, die im Unterricht eingesetzt werden sollten, wurden innerhalb des Projekts pilotiert. In jeweils einer B1.1-Klasse der pilotierenden Institute wurde der herkömmliche (analoge) Unterricht durch die Arbeit mit Tablets ergänzt. Die Lernenden dieser Klassen bildeten die Testgruppe im Projekt. Insgesamt bestand die Testgruppe aus 27 Lernenden, die sich wie folgt zusammensetzte: Standort London N=9, Standort Dublin N=6, Standort Amsterdam N=12. Die Kontrollgruppe bestand aus Lernenden einer weiteren B1.1-Klasse, in der „analog“ unterrichtet wurde. Die Zusammensetzung dieser Gruppe sah folgendermaßen aus: Standort London N=6, Standort Dublin N=4, Standort Amsterdam N=11. Zu beachten ist, dass die Teilnahme an dem Projekt für alle Lernenden freiwillig und unentgeltlich war. Für den Sprachkurs fielen darüber hinaus sowohl in der Test- als auch in der Kontrollgruppe die üblichen Teilnehmerkosten an. Um Unzufriedenheit zu vermeiden, musste sichergestellt werden, dass die Lernenden der Testgruppe sich bewusst für die Tablet-Klasse entschieden, eine randomisierte Verteilung war leider nicht möglich. Das Testdesign muss daher als quasi-experimentell angesehen werden.

In allen sechs Klassen (drei Tablet-Klassen für die Testgruppe und drei analoge Klassen für die Kontrollgruppe) wurde extensiv unterrichtet, das heißt, der Unterricht fand einmal pro Woche mit drei Unterrichtseinheiten (eine Unterrichtseinheit = 45 Minuten) über vier Monate hinweg statt.

Für den Fremdsprachenunterricht bieten Tablets unter anderem in den Bereichen Sprechen und Hören einen großen didaktischen Mehrwert. André J. Spang weist darauf hin, dass „gerade in den modernen Fremdsprachen [...] Apps hervorragend eingesetzt werden (können), um mit einfachsten Mitteln das freie Sprechen zu üben, aufzuzeichnen, zu beurteilen, zu verbessern und zu teilen“ (2013: 12). Die gleiche Erfahrung wurde auch im Pilotprojekt „Tablet-Klasse“ gemacht. Durch das spielerische Aufnehmen, Anhören und Korrigieren der eigenen Aussprache im Deutschen nehmen die Lernenden ihr Sprechen bewusster wahr, und es fällt ihnen leichter, darüber zu reflektieren. Zusätzlich können über die Tablets niveaugerechte Audios oder kurze Filme aus dem Internet im eigenen Lerntempo angehört werden. Über einen Splitter lassen sich an einem Tablet zwei Kopfhörer anschließen. Jeweils zwei Lernenden können so die Audios teilen und selbst entscheiden, wie oft sie sich die jeweilige Aufnahme anhören möchten. Viele Lernende nutzen darüber hinaus auch die Möglichkeit, besonders schwierige Passagen innerhalb einer Aufnahme mehrfach anzuhören.

Während viele Untersuchungen (vgl. Greb 2012; Ifenthaler & Schweinbenz 2013; Ludwig et al. 2011) im Bereich „Tablets im Unterricht“ bereits auf Vorteile in den Bereichen Lernmotivation und Lernerautonomie hinweisen,

Kathrin Hahn, Winfried Koch & Linda Fromme (2014), Der Online-Test BULATS Deutsch – ein geeignetes Instrument zur Messung von Lernfortschritten? Erste Ergebnisse im Rahmen des Projekts "Tablet-Klasse". *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 100-110. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Hahn_Koch_Fromme.pdf.

untersuchte die Begleitforschung zum Projekt „Tablet-Klasse“ im Rahmen einer quantitativen Untersuchung zum Hörverstehen darüber hinaus, ob für Lernende, die im Unterricht mit Tablets arbeiten, auch tatsächlich ein größerer Lernzuwachs zu verzeichnen ist als für Lernende, die analog unterrichtet werden. Grundsätzlich sollte hier nicht ausgeschlossen werden, dass auch eine Verschlechterung möglich ist, beispielsweise dann, wenn im Unterricht die Technik und nicht mehr das Sprachenlernen zum Fokus werden. Der Lernfortschritt im Hörverstehen wurde in beiden Gruppen durch Test und Re-Test mit dem BULATS Online-Test Deutsch gemessen. Im Folgenden soll zunächst das Forschungsdesign dieser Untersuchung beschrieben werden, um anschließend erste Ergebnisse vorzustellen.

2. Das Forschungsdesign

Für die Untersuchung zur Entwicklung des Hörverstehens in Test- und Kontrollgruppe auf Basis des BULATS Online-Test Deutsch wurden zwei Schwerpunkte gesetzt. Es ging darum, die Veränderungen, die für das Hörverstehen in Test- und Kontrollgruppe gemessen wurden, zu untersuchen und zu beschreiben und beide Gruppen miteinander zu vergleichen, um mögliche Unterschiede festzustellen.

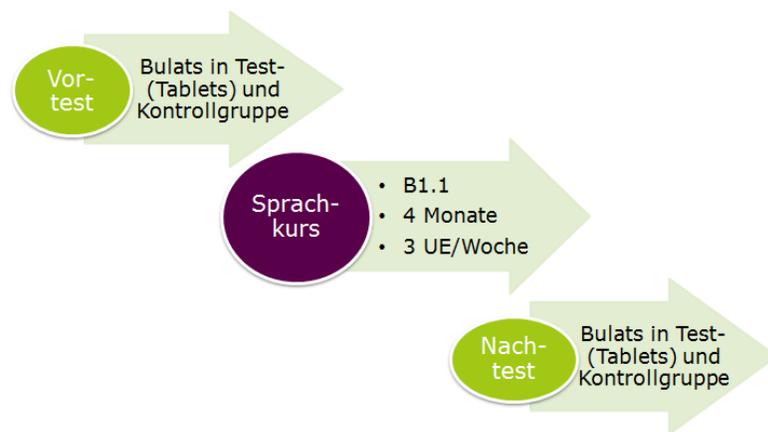


Abb. 1: Testdesign der Untersuchung

Die Zusammenstellung von Test- und Kontrollgruppe ist in Kapitel 1 bereits kurz beschrieben worden. Während in der Testgruppe Tablet-Aktivitäten zum Hörverstehen und Sprechen den Unterricht erweiterten, fand in der Kontrollgruppe konventioneller Unterricht ohne den Einsatz von Tablets oder andere digitale Medien statt. Die Zeit, die insgesamt auf die Förderung des Hörverstehens und des Sprechens verwendet wurde, war damit in der Testgruppe leicht höher als in der Kontrollgruppe. Die Orientierung an den Lehrwerken verhinderte aber ein deutliches Ungleichgewicht in diesen Bereichen.

Eine Umfrage zum Medienverhalten der Lernenden in Test- und Kontrollgruppe zu Beginn des Projekts zeigte, dass die Medienaffinität in beiden Gruppen relativ ausgeglichen war. In beiden Versuchsgruppen gab es Lernende mit viel und wenig Erfahrung im Umgang mit digitalen Medien.

Alle Lernenden in Test- und Kontrollgruppe legten zu Beginn des Semesters den BULATS Online-Test Deutsch ab. Darauf folgten vier Monate Sprachunterricht wie oben beschrieben. Am Ende des Semesters bearbeiteten beide Gruppen den Test erneut. Die Online-Version des Tests überprüft die Fertigkeiten Leseverstehen und Hörverstehen. Für die Untersuchung im Rahmen des Projekts „Tablet-Klasse“ waren nur die Ergebnisse im Hörverstehen relevant. Die Ergebnisse im Leseverstehen finden in der folgenden Untersuchung daher keine Beachtung.

Insgesamt nahmen 48 Testpersonen an der Untersuchung teil. Die Anzahl der getesteten Lernenden in der Testgruppe (Tablet-Klasse) war an allen drei Instituten höher ($N(\text{Test}) = 27$; $N(\text{Kontrolle}) = 21$) als in der Kontrollgruppe. Die Gesamtzahl von 48 Testpersonen ist relativ niedrig, was darauf zurückzuführen ist, dass alle erfassten Lernenden freiwillig an der Untersuchung teilgenommen haben. Für die Auswertung wurden nur diejenigen Lernenden berücksichtigt, die sowohl an der ersten als auch an der zweiten Testrunde teilgenommen haben.

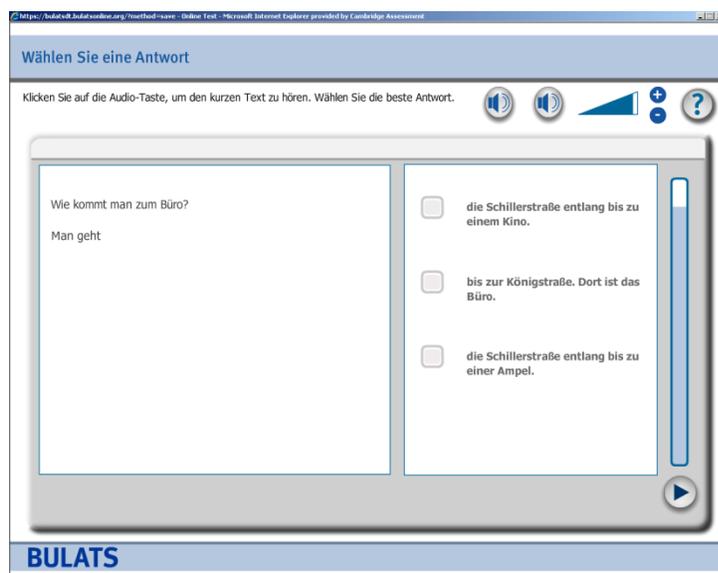
An dieser Stelle muss erwähnt werden, dass hier die Ergebnisse einer ersten Zwischenauswertung dargestellt werden, um einen Effekt nachzuweisen. Zur Gewinnung weiterer Testpersonen wurde das beschriebene Setting inzwischen an anderen Standorten wiederholt.

3. BULATS

Um den Lernfortschritt bzw. die Veränderung im Hörverstehen der Teilnehmenden objektiv darstellen zu können, entschied man sich innerhalb des Projekts für den Online-Test BULATS (Business Language Testing Service) als Vor- und Nachtest. Gründe für den Einsatz dieses Testinstruments sind einerseits, dass es alle Niveaustufen des *Gemeinsamen Europäischen Referenzrahmens für Sprachen (GER)*¹ berücksichtigt und adaptiv testet. Der Online-Test stellt den Sprachstand für die Fertigkeiten Hörverstehen und Leseverstehen (inklusive Wortschatz und Grammatik) fest und dauert durchschnittlich 60 bis 70 Minuten. Die Aufgaben werden in jedem Testlauf neu und individuell zusammengestellt, d.h. die Teilnehmenden waren bei den beiden Testläufen mit unterschiedlichen Aufgaben konfrontiert. Die Aufgabeninhalte beziehen sich im weitesten Sinne auf einen beruflichen Kontext. Für die Bearbeitung ist jedoch kein fachspezifisches Wissen erforderlich. BULATS testet alle vier Fertigkeiten (Schreiben, Sprechen, Leseverstehen und Hörverstehen). Im Projektkontext wurde die Online-Version des Tests Leseverstehen und Hörverstehen eingesetzt und bei der Analyse waren die Ergebnisse des Teils Hörverstehen relevant, die separat ausgewiesen wurden.²

3.1. Aufgaben Hörverstehen

BULATS arbeitet mit geschlossenen Aufgaben unterschiedlichen Typs. Beim Hörverstehen handelt es sich um Multiple-Choice-Aufgaben mit Antwortoptionen in Form von Grafiken, Bildern oder Text. Alle Höraufnahmen können zwei Mal gehört werden. Die Teilnehmenden bestimmen den Zeitpunkt des Abspielens selbst. Die folgenden Screenshots sollen die unterschiedlichen Aufgabentypen veranschaulichen (vgl. O'Sullivan 2006: 70-71).



Kathrin Hahn, Winfried Koch & Linda Fromme (2014), Der Online-Test BULATS Deutsch – ein geeignetes Instrument zur Messung von Lernfortschritten? Erste Ergebnisse im Rahmen des Projekts "Tablet-Klasse". *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 100-110. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Hahn_Koch_Fromme.pdf.

Abb. 2: Hören und Auswählen mit Textoptionen

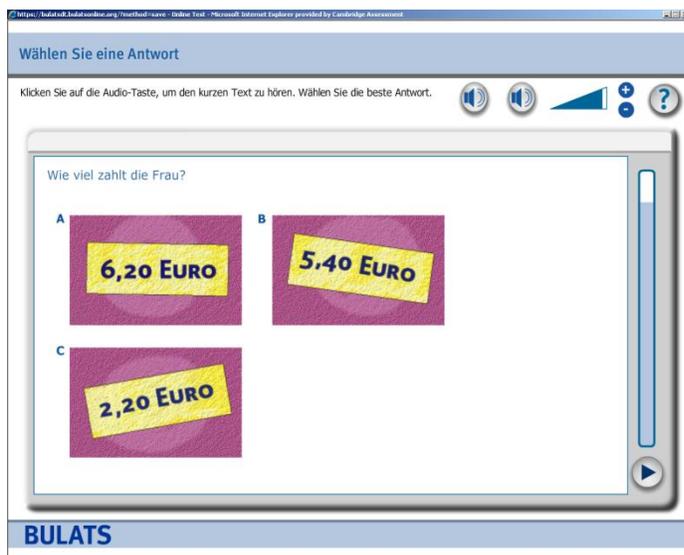


Abb. 3: Hören und Auswählen mit Grafikoptionen

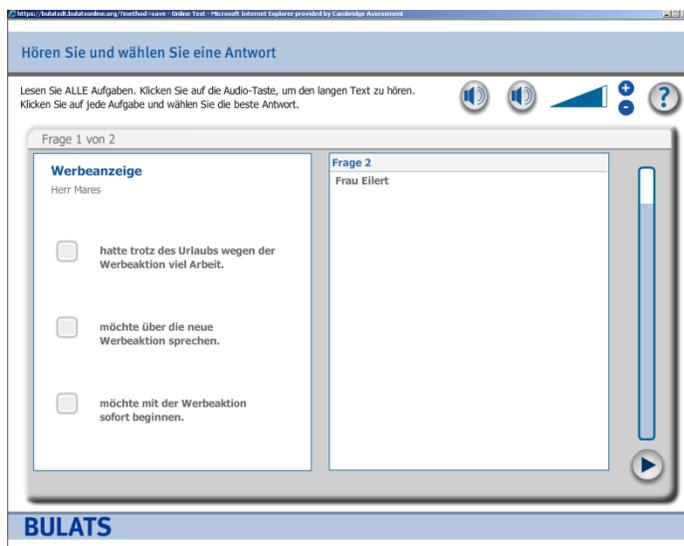


Abb. 4: Längeres Hören mit Textoptionen

Die Aufgaben sind jeweils mit unterschiedlichen Schwierigkeitsgraden entsprechend den einzelnen Niveaustufen in der Itembank hinterlegt.

3.2. Technologie und Testergebnisse

Die Testaufgaben werden je nach Niveaustufe des Kandidaten/der Kandidatin ausgewählt. Alle Teilnehmenden beginnen mit einer Aufgabe mittleren Schwierigkeitsgrades. Das Programm wählt mit Hilfe eines adaptiven Algorithmus die Aufgaben danach aus, ob der/die Testteilnehmende die vorherige Aufgabe richtig oder falsch gelöst hat.

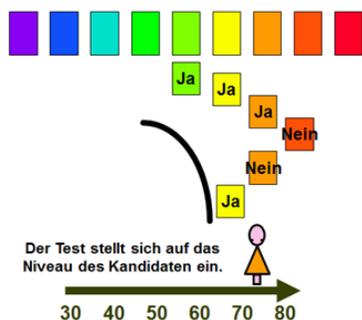


Abb. 5: Adaptives Testen

Auf diese Weise wird ein angemessener Schwierigkeitsgrad für den jeweiligen Kandidaten sichergestellt, was eine genauere Einschätzung seiner tatsächlichen Sprachkompetenz erlaubt als ein nicht-adaptiver Test mit ähnlicher Aufgabenanzahl. Die Auswertung erfolgt computerbasiert und die Ergebnisse sind direkt im Anschluss an den Test abrufbar. Für alle Sprachniveaus gibt es festgelegte Cut-off-Werte (Testtrennwerte), anhand derer die Teilnehmenden den sechs Sprachkompetenzstufen zugeordnet werden. Diese entsprechen den Stufen des GER.

Die Teilnehmenden erhalten als Ergebnis in Vor- und Nachtest eine Gesamtstufe für beide Testteile und Einzeleinstufungen für die Teile Leseverstehen und Sprachkenntnis (Wortschatz und Grammatik) sowie Hörverstehen. Die Ergebnisse werden in Form von Punkten auf einer Skala von 0 bis 100 dargestellt und dem entsprechenden Niveau des GER zugeordnet.

Stufe des Europarats (GER)	BULATS Cut-off-Werte	Beschreibung des Niveaus
C2	90–100	kompetente Sprachverwendung
C1	75–89	
B2	60–74	selbständige Sprachverwendung
B1	40–59	
A2	20–39	elementare Sprachverwendung
A1	0–19	
Vor-A1	0-9	Teilnehmende, die auf das Niveau A1 hinarbeiten

Abb. 6: Zuordnung der Stufen und Cut-off-Werte bei BULATS

Allerdings gilt hier die Einschränkung: “It is important to mention that test scores on CB BULATS do not refer to raw scores. They are actually ability estimates derived from a latent trait (Rasch) analysis, converted into BULATS scores by means of a scaling procedure.“ (Geranpayeh 2001: 15).

4. Ziele und Methoden der statistischen Auswertung

Die primäre Variable für die Analyse ist die *Veränderung der Ergebnisse im Hörverstehen* zwischen dem Testergebnis zu Beginn des Semesters und dem Testergebnis im Re-Test nach vier Monaten Sprachunterricht. Die zweiseitige Nullhypothese, die im Rahmen der Untersuchung getestet werden soll, besagt, dass es *zwischen der Test- und der Kontrollgruppe keine Unterschiede für die Veränderung im Hörverstehen gibt*. Ergänzend hierzu soll untersucht werden, inwieweit die *Veränderung der Ergebnisse im Hörverstehen vom Testergebnis im BULATS-Test zu Beginn*

des Semesters und dem Standort, an dem die getesteten Lernenden unterrichtet werden, abhängt. Die Ergebnisse dieser ersten Testrunde können außerdem Aufschlüsse und wichtige Hinweise³ für weitere Untersuchungen geben.

Quantitative Ergebnisse werden als Mittelwerte mit zugehörigem Standardfehler (SF) dargestellt. Unterschiede zwischen den beiden Testgruppen (Tablet- und Kontrollgruppe) werden als mittlere Differenzen mit 95 %-Vertrauensbereich dargestellt. Zusätzlich wird die Effektstärke als Quotient der Differenz und der gepoolten Standardabweichung innerhalb der Gruppen angegeben.

Für den statistischen Vergleich der Zuwächse im Hörverstehen zwischen den Gruppen wird wegen des Vorliegens einer annähernden Normalverteilung der t-Test für unabhängige Stichproben eingesetzt. Ein p-Wert <0.05 wird als statistisch signifikant, ein p-Wert <0.2 als Vorliegen eines statistischen Trends betrachtet. Die Abhängigkeit des BULATS-Punktezuwachses im Hörverstehen vom Ausgangswert wird mittels einfacher Regressionsanalyse pro Gruppe untersucht und grafisch dargestellt. Zusätzlich wurde in einem allgemeinen linearen Modell simultan die Abhängigkeit des Zuwachses im Hörverstehen vom Ausgangswert der Lernenden im BULATS-Test, vom Standort, an dem die getesteten Lernenden unterrichtet werden, und von der Testgruppe (Tablet-Klasse vs. analoge Klasse) untersucht.

Die Auswertungen wurden mit SAS JMP Version 10 und höher durchgeführt.

5. Ergebnisse

In der Testgruppe beträgt der Zuwachs im Hörverstehen im Mittel 7.04 (SF=2.47) und in der Kontrollgruppe im Mittel 3.33 (SF=2.80) BULATS-Punkte. Der Unterschied zwischen beiden Gruppen beträgt 3.70 BULATS-Punkte mit einem 95 % Vertrauensbereich der Differenz von [-3.8, 11.2]. Für den t-Test für unabhängige Stichproben beträgt der zweiseitige p-Wert zum Testen der zentralen Nullhypothese (es gibt keine Unterschiede zwischen Test- und Kontrollgruppe) $p = 0,326$. Er ist damit deutlich entfernt von statistischer Signifikanz ($\alpha = 0.05$). Bei den vorliegenden geringen Fallzahlen kann also zunächst einmal nicht von einer statistisch gesicherten Überlegenheit des Unterrichts mit Tablets gesprochen werden, auch wenn die Tablet-Gruppe im Mittel einen besseren Zuwachs im Hörverstehen aufweist.

Kathrin Hahn, Winfried Koch & Linda Fromme (2014), Der Online-Test BULATS Deutsch – ein geeignetes Instrument zur Messung von Lernfortschritten? Erste Ergebnisse im Rahmen des Projekts "Tablet-Klasse". *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 100-110. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Hahn_Koch_Fromme.pdf.

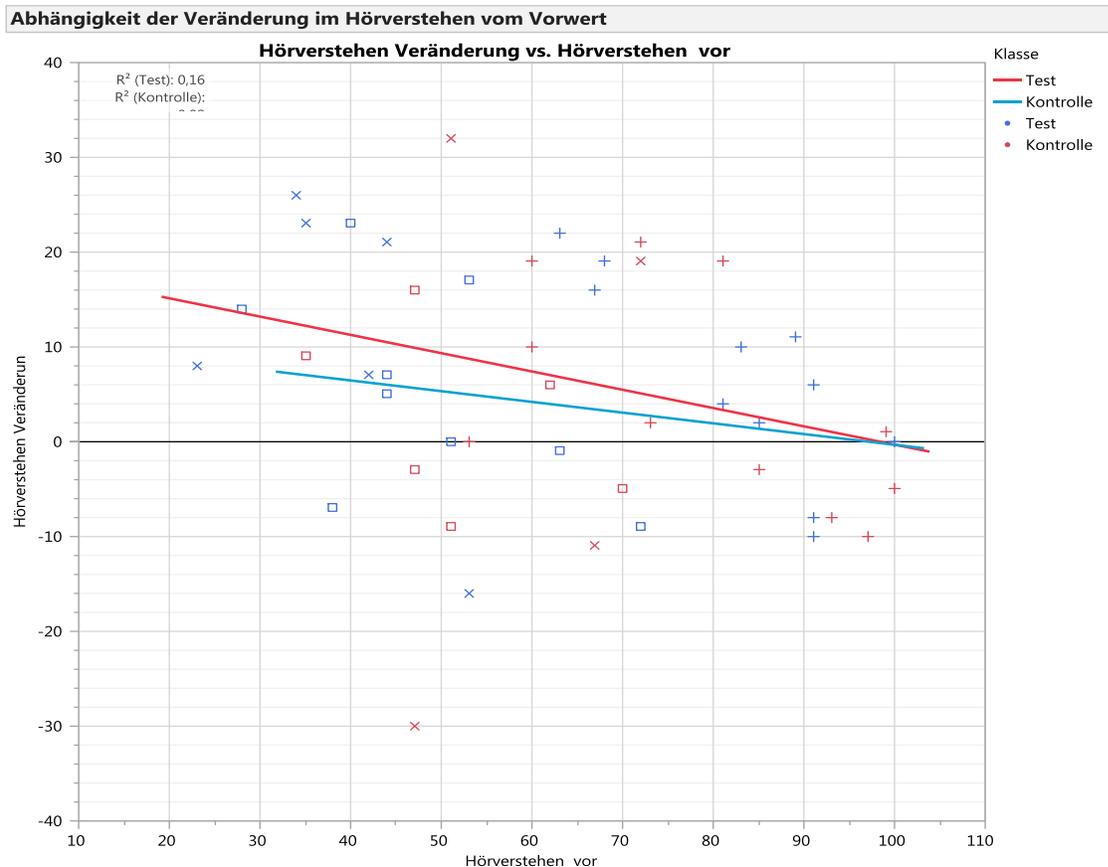


Abb. 7: Abhängigkeit der Veränderung im Hörverstehen vom Ausgangswert pro Gruppe

In Abbildung 7 ist die Veränderung im Hörverstehen für Test- und Kontrollgruppe auf der vertikalen y-Achse dargestellt. Die Werte für die Testgruppe sind rot, die Werte für die Kontrollgruppe blau markiert. Die Lernenden an den einzelnen Standorten sind durch unterschiedliche Marker gekennzeichnet. Plus-Symbole kennzeichnen die Lernenden am Standort Amsterdam, Quadrate die Lernenden am Standort Dublin und Kreuze die Lernenden am Standort London. Auf der x-Achse sind die Testergebnisse des ersten BULATS-Tests zu Beginn des Semesters abgebildet. Zu erkennen ist zunächst, dass die Werte für den ersten BULATS-Test von 20 Punkten (entspricht Niveau Anfang A2) bis hin zu Werten von 100 Punkten (entspricht Niveau C2) reichen. Diese großen Abweichungen innerhalb der Ausgangswerte sind mit der Tatsache zu erklären, dass die niederländischen Lernenden durch die Nähe des Niederländischen zum Deutschen gerade im Hör- und Leseverstehen sehr leicht sehr gute Ergebnisse erzielen, diese allerdings nicht repräsentativ für ihre tatsächliche Niveaustufe und ihre Fähigkeiten im Schreiben und Sprechen sind. Insofern müssen diese Ergebnisse relativ zu ihrem tatsächlichen Sprachstand im Deutschen und zu den anderen Fertigkeiten gesehen werden.

Die mittlere Veränderung im Hörverstehen für Test- bzw. Kontrollgruppe wurde als eine Funktion der Ergebnisse für das Hörverstehen im Test zu Anfang des Semesters mit Hilfe einer linearen Regressionsgeraden geschätzt. Folgende Beobachtungen ergeben sich aus den Regressionsgeraden der Abbildung 7:

Die mittlere Veränderung im Hörverstehen ist sowohl für die Test- als auch für die Kontrollgruppe abhängig vom *Ausgangswert* im ersten BULATS-Test zu Beginn des Semesters. Je niedriger der Ausgangswert, desto höher ist die mittlere Veränderung im Hörverstehen zwischen Vor- und Nachtest. Für Ausgangswerte ab 80 BULATS-Punkten sind zunehmend geringe Veränderungen im Hörverstehen messbar, da der Spielraum für Verbesserungen hier sehr

gering ist. Für Ausgangswerte, die nahe an 100 BULATS-Punkten liegen, geht die Entwicklung im Hörverstehen nach der Schätzung gegen Null, da es hier für die Lernenden nahezu keinen Entwicklungsspielraum mehr gibt. Damit kann bei diesen hohen Ausgangswerten auch kaum ein Unterschied zwischen den beiden Unterrichtsmethoden erkannt werden.

Auf Grundlage der vorausgegangenen Ergebnisse wurde die Subgruppe der Lernenden ausgewertet, die im ersten Test nicht mehr als 70 BULATS-Punkte erreicht hatten. Der geschätzte mittlere Unterschied für die Veränderung im Hörverstehen zwischen Test- und Kontrollgruppe liegt für diese Subgruppe bei 8,0 BULATS-Punkten mit einem 95 % Vertrauensbereich von [-2,7, 18,7]. Im Vergleich zur gesamten Testgruppe ist für den Vertrauensbereich also eine Verschiebung in Richtung einer stärkeren Überlegenheit der Testgruppe zu sehen. Der p-Wert für den Vergleich der Zuwächse mit dem t-Test ist $p=0,137$. Das bedeutet zunächst, dass auch für diese Untergruppe noch kein statistisch signifikanter Unterschied zwischen Test- und Kontrollgruppe festgestellt werden kann. Das Testergebnis liegt allerdings deutlich unter 0,2 und zeigt damit einen statistischen Trend zu einem statistisch relevanten Unterschied zwischen Test- und Kontrollgruppe zum Vorteil der Lernenden, für die der Unterricht durch den Einsatz von Tablets ergänzt wurde.

Die geschätzte relative Effektstärke liegt für den mittleren Unterschied zwischen Test- und Kontrollgruppe bei 0,578. Eine solche Effektstärke kann als Indikator für einen mittelstarken Effekt angesehen werden. Da der mittlere Unterschied von 8 BULATS-Punkten außerdem sehr nah an der Minimalpunktzahl zum Erreichen einer höheren Niveaustufe im BULATS-Test (vgl. Kapitel 3) liegt, ist jedenfalls für Lernende mit nicht am oberen Ende des Messbereichs des BULATS liegenden Ausgangswerten anzunehmen, dass das Lernen mit Tablets eine durchaus interessante und gewinnbringende Lehr-Lernmethode darstellt.

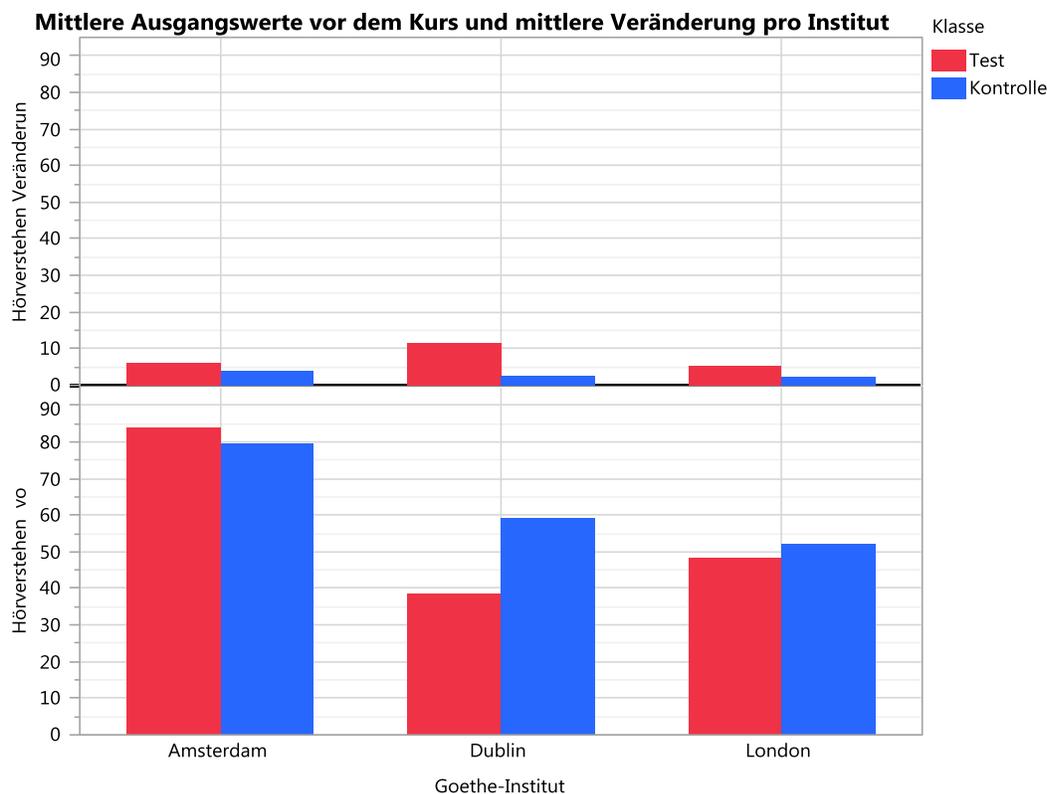


Abb. 8: Mittlere Veränderung und mittlerer Ausgangswert im Hörverstehen pro Gruppe in Abhängigkeit vom jeweiligen Standort

Abbildung 8 zeigt in Ergänzung zu den bereits dargestellten Unterschieden die Werte getrennt nach den jeweiligen Standorten und nach Test- und Kontrollgruppe. Im unteren Teil der Grafik sind die mittleren Testergebnisse der ersten Testrunde dargestellt. Auch hier ist die Testgruppe in roter, die Kontrollgruppe in blauer Farbe gekennzeichnet. Aus der Grafik ist zunächst abzulesen, dass die mittleren Ausgangswerte für Test- und Kontrollgruppe an den Standorten in London und Amsterdam gut miteinander vergleichbar sind. An beiden Instituten erreichen Test- und Kontrollgruppe ähnliche mittlere Testwerte im Ausgangstest, in Amsterdam erzielt die Testgruppe etwas bessere Ergebnisse, in London die Kontrollgruppe. Am Standort Dublin zeigt die Kontrollgruppe bessere Ausgangswerte als die Testgruppe. Zu beachten ist allerdings, dass der Standort Dublin mit insgesamt nur 10 Lernenden die kleinste Probandengruppe gestellt hat. Stärkere Streuungen in der Verteilung waren hier also zu erwarten. Deutlich zu erkennen ist im unteren Teil der Grafik außerdem die Überlegenheit der Lernenden am Standort Amsterdam. Die mittleren Ausgangswerte liegen hier bei 85 BULATS-Punkten für die Test- und 80 BULATS-Punkten für die Kontrollgruppe. Weiter oben wurde bereits erläutert, dass diese hohen Ausgangswerte der niederländischen Lernenden im Bereich Hörverstehen auf die Nähe des Niederländischen zum Deutschen zurückzuführen sind.

Der obere Teil der Grafik zeigt auf derselben Skala die mittlere Veränderung im Hörverstehen, wieder getrennt nach Standort und Test- bzw. Kontrollgruppe. An allen drei Standorten ist die Veränderung im Hörverstehen für die Testklasse höher als für die Kontrollklasse. Bedingt durch die niedrigen Werte im Ausgangstest, ist der deutlichste Unterschied für den Standort Dublin zu erkennen. Interessant ist auch, dass Amsterdam trotz der hohen Ausgangswerte eine deutliche mittlere Veränderung im Hörverstehen zeigt. Die Lernenden der Testklasse zeigen hier trotz höherer Ausgangswerte eine größere Verbesserung im Hörverstehen als die Kontrollgruppe; die Darstellung der Einzelwerte in Abbildung 7 zeigt für den Standort Amsterdam (=Marker +), dass die höchsten Zuwächse von Lernenden mit Ausgangswerten unter 75 BULATS Punkten stammen, während Lernende mit einem Ausgangswert über 90 eher eine abnehmende Tendenz im Hörverstehen zeigen. Zusammenfassend lässt sich zu Abbildung 8 sagen, dass an allen drei getesteten Standorten die Lernenden der Tablet-Klasse nach dem Re-Test im Mittel eine stärkere Veränderung im Hörverstehen zeigen als die Lernenden der Kontrollgruppe, unabhängig davon, ob die Testgruppe im Ausgangstest mehr oder weniger BULATS-Punkte erzielt hat als die Kontrollgruppe.

Bei der Auswertung der Daten in einem allgemeinen linearen Modell mit simultaner Anpassung konnten daneben fast signifikante Effekte der Einflussgrößen *Standort* ($p=0,061$) und erreichter *Ausgangswert* im BULATS-Test zu Beginn des Semesters ($p=0,011$) auf die Veränderungen im Hörverstehen von Vortest zu Nachtest identifiziert werden.

Wie oben erläutert, hatte die Nutzung der Tablets in der Testklasse einen positiven, wenn auch bei der bislang vorliegenden Fallzahl nicht signifikanten, Effekt auf die Veränderung des Hörverstehens innerhalb des Testintervalls von 4 Monaten ($p<0,2$ für die Untergruppe der Lernenden mit einem Ausgangswert ≤ 70 BULATS-Punkte). Insgesamt gehen wir davon aus, dass die Verwendung von Tablets im Klassenzimmer neben einer deutlichen Steigerung der Lernmotivation auch zu einer tatsächlichen Verbesserung des Lernerfolgs beigetragen hat. Zu Beginn wurde bereits darauf verwiesen, dass die hier vorgestellten Daten nur aus einer ersten Testgruppe von insgesamt 48 ProbandInnen gewonnen wurden. Die Erhebung weiterer Daten ist geplant; um bei einer angenommenen wahren relativen Effektstärke von 0,5 mit 80 % statistischer Power einen statistisch signifikanten Unterschied nachzuweisen, streben wir sowohl für Test- als auch für Kontrollgruppe jeweils mindestens 64 ProbandInnen an.

Als Konsequenz aus diesen Ergebnissen sollen für zukünftige Erhebungen dieser Art nur noch Lernende mit einem Ausgangswert von bis zu 70 oder maximal 80 BULATS-Punkten berücksichtigt werden. Damit soll gewährleistet werden, dass den Lernenden genügend Spielraum zur Entwicklung und damit für die statistische Auswertung genügend Spielraum für den Nachweis von Unterschieden zur Verfügung steht.⁴

6. Fazit und Ausblick

Der Einsatz von digitalen Medien im Unterricht ist heute zu einer Selbstverständlichkeit geworden. Es ist dennoch wichtig, dass neue Technologien auf ihren didaktischen Mehrwert für den Unterricht hin evaluiert werden. Da Tablets erst seit Kurzem Einzug in Klassenzimmer halten, war Ziel des Projekts, festzustellen, ob ihr Einsatz im

Kathrin Hahn, Winfried Koch & Linda Fromme (2014), Der Online-Test BULATS Deutsch – ein geeignetes Instrument zur Messung von Lernfortschritten? Erste Ergebnisse im Rahmen des Projekts "Tablet-Klasse". *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 100-110. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Hahn_Koch_Fromme.pdf.

Fremdsprachenunterricht neben motivationaler Stützung auch eine tatsächliche Leistungsverbesserung mit sich bringt.

Auch wenn aufgrund der Fallzahlen noch keine statistisch signifikanten Gruppenunterschiede ermittelt werden konnten, sind die Ergebnisse vereinbar mit der Hypothese, dass die Lernenden durch die Arbeit mit Tablets im Hörverstehen einen höheren Lernerfolg erzielen als Lernende in analogen Klassen. Um im Vor- und Nach-Test-Verfahren diesen Nachweis zu erbringen und den Lernfortschritt beider Gruppen zu vergleichen, erwies sich der BULATS Online-Test Deutsch als ein geeignetes Messinstrument.

Literaturangaben

- Geranpayeh, Ardeshir (2001), CB BULATS: Examining the reliability of a computer based test using test-retest method. *Research Notes* 5. Cambridge English Language Assessment, 14-17.
- Greb, Thorsten (2012), Projekt „Mobiles Lernen mit Tablet-Computern“ startet in den Regelbetrieb. *Magazin Digital Lernen. Das Online-Magazin zu digitalen Medien in Bildungseinrichtungen* [Online unter <http://www.digital-lernen.de/nachrichten/schulpraxis/einzelansicht/artikel/projekt-mobiles-lernen-mit-tablet-computernstartet-in-den-regelbetrieb.html>. 27.04.2014].
- Ifenthaler, Dirk & Schweinbenz, Volker (2013), The acceptance of Tablet-PCs in classroom instruction: The teachers' perspectives. *Computers in Human Behavior* 29: 3, 525-534.
- Ludwig, Luise; Mayrberger, Kerstin & Weidmann, Adrian (2011), Einsatz personalisierter iPads im Unterricht aus der Perspektive der Schülerinnen und Schüler. In: Friedrich, Steffen; Kienle, Andrea & Rohland, Holger (Hrsg.), *DeLFI 2011: Die 9. e-Learning Fachtagung Informatik – Poster, Workshops, Kurzbeiträge*. Dresden: TUDpress, 7-17 [Online unter <http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-77317>. 30.07.2014].
- O'Sullivan, Barry (2006), Issues in Testing Business English. *Studies in Language Testing* 17. Cambridge University Press, 63-77.
- Spang, André (2013), Lernen 2.0: Mobiles Lehren und Lernen mit iPad und Cloud. *Praxis Fremdsprachenunterricht* 10: 1, 11-13.

Anmerkungen

- ¹ Die BULATS-Stufen entsprechen den Niveaus von ALTE (Association of Language Testers in Europe), die durch laufende Forschungsverfahren (The ALTE Can Do-Project) an den Gemeinsamen Europäischen Referenzrahmen gebunden wurden. Für weitere Informationen zu dem Projekt siehe ALTE-Homepage <http://www.alte.org>.
- ² BULATS ist neben Deutsch auch für die Sprachen Englisch, Spanisch und Französisch verfügbar und wird im Rahmen eines Gemeinschaftsprojekts von *Cambridge English Language Assessment*, dem *Goethe-Institut*, der *Alliance Française* und der *Universidad de Salamanca* entwickelt und erstellt.
- ³ Die wichtigsten Hinweise, die aus dieser ersten Studie gewonnen werden konnten, waren, zukünftige ProbandInnen mit einem Ausgangswert nahe 100 BULATS-Punkten aus der Untersuchung auszuschließen. Daneben konnten wichtige Erkenntnisse über die Probandenzahl, die für ein signifikantes Ergebnis benötigt werden, gewonnen werden.
- ⁴ Um bei einer angenommenen wahren relativen Effektstärke von 0,5 mit 80 % statistischer Power einen statistisch signifikanten Unterschied nachzuweisen, streben wir sowohl für Test- als auch für Kontrollgruppe jeweils mindestens 64 ProbandInnen an.