Formative Leistungsmessung im naturwissenschaftlichen Unterricht

Effekte von Feedbackvariationen in formativen Leistungstests auf den Wissenszuwachs am Beispiel einer Unterrichtseinheit zum Thema „Anpassung der Vögel ans Fliegen“ in der Sekundarstufe 1

Von der Pädagogischen Hochschule Schwäbisch Gmünd
zur Erlangung des Grades einer

Doktorin der Philosophie (Dr. phil.)

angenommene DISSERTATION von

Nicole Wolf

aus

Bamberg

2014

Erstgutachter: Prof. Dr. Uwe Maier (PH Schwäbisch Gmünd)
Zweitgutachter: Prof. Dr. Christoph Randler (PH Heidelberg)

Fach: Erziehungswissenschaft

Abgabetermin der Dissertation: 27.03.2014

Inhaltsverzeichnis:

1 Einleitung 7

2 Stand der Wissenschaft 10

2.1 Einleitung 10

2.2 Formative Leistungsmessung 10

2.2.1 Grundlegende Aspekte formativer Leistungsmessung nach Black & Wiliam (2009) 17

2.2.2 Feedback als zentrale Erklärung von Effekten formativer Leistungsmessung 19

2.2.2.1 Feedback in elektronischen formativen Leistungstests 38

2.3 Biologiedidaktik 50

2.3.1 Evolutionäre Anpassung als Beispiel konzeptuellen Wissens in der Biologie 50

2.3.1.1 Falsche Schülervorstellungen innerhalb des Konzepts 52

3 Forschungsdesiderata und Hypothesen 56

4 Organisatorische und inhaltliche Vorbereitung der Studie 61

4.1 Vorüberlegungen und Zielsetzung 61

4.2 Verlauf der Studie 61

4.2.1 Entwicklung prototypischer Diagnoseaufgaben und Feedbackformate 64

4.2.2 Durchführung der Vorstudien 64

4.2.3 Vorbereitung der Lehrkräfte auf das Feldexperiment 77

4.2.4 Ablauf des Feldexperiments 79

5 Evaluationsdesign / Methodik 88

5.1 Zielgruppe 89

5.2 Die Unterrichtseinheit 89

5.2.1 Thema und Inhalt 89

5.2.2 Verankerung in den bayerischen Lehrplänen 91

5.3 Aufgaben- und Testdesign 94

5.3.1 Aufgabentypen 94

5.3.1.1 Multiple Choice / Single Choice 97

5.3.1.2 Lückentext 101

5.3.1.3 Zuordnungsaufgaben 102

5.3.2 Messinstrumente 104

5.3.2.1 Einleitung 104

5.3.2.2 Pre- und Posttest 107

5.3.2.3 Formative Leistungstests / Treatment 109

5.3.2.4 Posttest 2 – Behaltenstest 113

5.3.2.5 Motivationstests und Kurzskalen 115

5.3.3 Feedback 121

5.3.3.1 Feedback an den Lerner 126

5.3.3.2 Feedback in Bezug auf die Aufgabe 128

5.3.3.3 Feedback in Bezug auf den Lösungsprozess 130

5.3.3.4 Feedback in Bezug auf die Person 134

5.3.3.5 Feedback an den Lehrenden zur Optimierung der Lernprozesse 135

5.4 Technische Umsetzung 137

5.4.1 Vor- und Nachteile elektronischer formativer Leistungstests 137

5.4.2 Technische Umsetzung des Projekts 140

5.4.2.1 Software 140

5.4.2.1.1 Das Autorensystem zur Erstellung der Testfragen 140

5.4.2.1.2 Lernplattform zur Auswertung der Schülerleistungen 146

5.4.2.2 Hardwarekomponenten 149

5.4.2.2.1 Server und Accesspoint 149

5.4.2.2.2 Endgeräte 151

6 Evaluationsergebnisse 158

6.1 Einleitung 158

6.2 Ergebnisse des Feldexperiments 158

6.2.1 Allgemeine Informationen über die Stichprobe ... 158

6.2.2 Vorbereitungen für die Ergebnisauswertung ... 159

6.2.3 Reliabilität der Tests ... 164

6.2.4 Hypothesenprüfung ... 165

6.2.5 Motivationale Auswirkungen ... 171

6.2.6 Änderung der Schülervorstellungen zum Konzept der evolutionären Anpassung ... 174

6.2.7 Der Umgang mit den Tablets , den elektronischen Tests und dem Feedback ... 181

6.2.8 Die Bewertung der Unterrichtseinheit ... 188

6.2.9 Teilnehmende Beobachtung ... 189

6.3 Diskussion der Ergebnisse ... 191

6.3.1 Methodisches Vorgehen ... 191

6.3.1.1 Die Unterrichtssequenz ... 191

6.3.1.2 Das Treatment ... 193

6.3.1.3 Das Feedback ... 195

6.3.1.4 Die Technik ... 199

6.3.1.5 Organisatorisches / Ablauf der Evaluation ... 204

6.3.2 Die Messinstrumente ... 205

6.3.2.1 Reliabilität der Tests ... 205

6.3.2.2 Pre- und Posttest ... 205

6.3.2.3 Formative Leistungstests ... 206

6.3.2.4 Behaltenstest ... 207

6.3.3 Hypothesenprüfung ... 207

6.3.3.1 Faktenwissen ... 209

6.3.3.2 Konzeptwissen ... 210

6.3.3.3 Fehlvorstellungen zur „Evolutionären Anpassung“ ... 216

6.3.4 Motivationale Auswirkungen ... 219

7 Ausblick ... 222

7.1.1 Nächste Schritte in der Forschung 222

7.1.2 Empfehlungen für die Schulpraxis 225

8 Literaturverzeichnis I

9 Eigene Veröffentlichungen XVII

10 Abbildungsverzeichnis XVIII

11 Tabellenverzeichnis XX

12 Inhaltsverzeichnis des Anhangs XXII

13 Danksagung XXIII

Anmerkung zur geschlechtsneutralen Formulierung:

Aus Gründen der besseren Lesbarkeit wird im Text auf eine geschlechtsneutrale Formulierung, wie z.B. LehrerInnen oder SchülerInnen verzichtet. Selbstverständlich richten sich alle Formulierungen gleichermaßen an beide Geschlechter.

1 Einleitung

Durch die immer schneller voranschreitenden veränderten Bedürfnisse unserer Gesellschaft, was die Ausbildung und die Fähigkeiten der Jugendlichen beim Start in das Berufsleben angeht, werden Forderungen nach Veränderungen im Bildungssystem laut. Zum immer stärker werdenden innerdeutschen Vergleich der Schüler (Jahrgangsstufentests, etc.) kommt die Internationalisierung und der damit einhergehende internationale Druck (PISA, etc.). Auf der Suche nach geeigneten Wegen, diesen Entwicklungen und Bedürfnissen gerecht zu werden, gibt es viele Diskussionen und auch Umstrukturierungen im Bildungssystem der Länder. Neben größeren Veränderungen, wie zum Beispiel der Umstellung der gymnasialen Oberstufe auf eine verkürzte Schulzeit (G8 statt G9), um im internationalen Vergleich bestehen zu können, wenn es um das Alter der Schul- und Hochschulabsolventen geht, sollen auch kleinere innere Strukturen, die Curricula und den Unterricht selbst betreffend, verändert werden. Dabei stehen vor allem die Forderungen nach individuelleren und differenzierenden Methoden und Unterrichtsangeboten im Vordergrund. So werden reformpädagogische Ansätze, die versuchen, die Kinder mit ihren eigenen Fähigkeiten in den Mittelpunkt zu stellen und diese dort abzuholen, wo sie sich in ihrer Entwicklung befinden (zum Beispiel Montessoripädagogik) und auch alternative Lehr- und Lernmethoden interessant.

In diesem Zusammenhang lohnt sich auch ein genauerer Blick auf das Prüfungssystem der Schulen. Der Prüfungsalltag in den Schulen ist geprägt von herkömmlichen summativen Tests, die am Ende einer Unterrichtseinheit aufdecken sollen, wie gut die Schüler die Lerninhalte verstanden haben, um diese anschließend anhand ihrer Leistungen zu bewerten. Leider werden die Leistungen selbst nicht näher in Augenschein genommen, um den nachfolgenden Förderbedarf der einzelnen Schüler festzustellen und darauf einzugehen. Vielmehr wird die Lerneinheit abgeschlossen und mit der nächsten begonnen. Dadurch entstehen vor allem in Fächern, wo verschiedene Inhalte aufeinander aufbauen (Mathematik, Chemie, Fremdsprachen, etc.), nach und nach im größere Wissenslücken, die später kaum mehr oder nur durch große Anstrengungen zu schließen sind. Ein Weg, diese Wissenslücken zu verhindern, können formative Leistungsmessungen sein. Diese gelten in der Lehr- und Lernforschung inzwischen als effektiver Weg zur Steigerung von Schülerleistungen und Lernmotivation. Durch häufige und in den Lehr-Lernprozess eingebaute Rückmeldungen sollen Lehrende und Lernende Informationen zur Op-

timierung ihres Lehr- und Lernverhaltens bekommen (z.B. Black & Wiliam 1998; Maier 2010). Formative Leistungsdiagnostik soll die Lehrenden über den Wissensstand der Schüler informieren und Fördermöglichkeiten der Schüler aufzeigen. Sie ermöglicht der Lehrkraft dem Lerner bei der Entwicklung von Lernstrategien zur Seite zu stehen, die Schüler selbst in ihrem Lernprozess zu begleiten und auch den nachfolgenden Unterricht an die Bedürfnisse der Klasse ständig anzupassen. Die Lerner erhalten immer wieder Rückmeldungen zu ihrem Wissensstand und Lernfortschritt und lernen Wissenslücken eigenständig zu isolieren und diese zu schließen. Dies setzt jedoch eine zeitsparende, wiederholbare aber dennoch valide Erfassung von grundlegendem fachspezifischem Wissen innerhalb einer Unterrichtssequenz voraus (Black & Wiliam 2009). Ebenso sollte formatives Feedback so gestaltet sein, dass sich Anschlüsse für den laufenden Unterricht oder selbstgesteuerte Lernprozesse ergeben. Je nach Altersstufe und Fach sind vielfältige Formate für formative Leistungsdiagnosen und den damit einhergehenden Rückmeldungen denkbar.

Da es bisher nur eine ökologisch valide Experimentalstudie gibt, in der Feedbackmerkmale im Rahmen eines systematischen Konzepts formativer Leistungsmessung variiert wurden, in der aber aufgrund forschungsmethodologischer Probleme keine Leistungseffekte nachgewiesen werden konnten (Shavelson et al. 2008), sollen in dieser Arbeit die kognitiven und motivationalen Effekte einer feedbacktheoretisch begründeten Rückmeldestrategie empirisch untersucht werden. Dazu werden die formativen Leistungstests innerhalb einer Unterrichtseinheit zum Thema „Vogelflug“ in Biologie durchgeführt. Während dieser beiden Testzeitpunkte werden die teilnehmenden Klassen in drei Gruppen unterteilt (zufällige Zuordnung) und bearbeiten diagnostische Aufgaben mit unterschiedlichen Rückmeldungen. Durch dieses Forschungsdesign werden Effekte der Lehrervarianz minimiert. Die Effekte des Treatments auf Fachwissen, Interesse und Motivation sollen mit Hilfe von Motivations-, Vor- und Nachtests geprüft werden.

Die Dissertationsarbeit entstand im Rahmen eines von der DFG geförderten Projekts zum Thema „Formative Leistungsmessung im naturwissenschaftlichen Unterricht“ und ist folgendermaßen aufgebaut: In einem Theorieteil wird zu Beginn der aktuelle Forschungsstand in den Bereichen „Formative Leistungsmessung“, „Feedback“ und „Einsatz elektronischer formativer Leistungstests“, vor allem im naturwissenschaftlichen Unterricht näher beleuchtet. Nachdem die

Forschungsdesiderate in diesen Bereichen aufgezeigt und die zu überprüfenden Hypothesen aufgestellt wurden, werden das Design und die Durchführung der Studie genauer beschrieben. Im Anschluss daran, sollen die Ergebnisse der Studie dargelegt, diskutiert und mit deren Hilfe die Richtigkeit die zuvor aufgestellten Hypothesen überprüft werden. Zuletzt werden Empfehlungen für anschließende Forschungsprojekte und den Einsatz elektronischer formativer Leistungstests in der Schulpraxis ausgesprochen.

2 Stand der Wissenschaft

2.1 Einleitung

In den folgenden Abschnitten soll näher beleuchtet werden, welche Forschungsansätze und Ergebnisse es bereits im Bereich der formativen Leistungsmessung gibt und wie wichtig bei dieser Form der Leistungsdiagnostik ein überlegter Umgang mit Rückmeldungen ist. Die systematische Variation von Feedback in formativen Leistungstests und dessen empirische Untersuchung ist zentraler Punkt der Studie. Ein weiterer Hauptaspekt besteht im Einsatz formativer Diagnostik im naturwissenschaftlichen Unterricht, denn besonders wenn es darum geht, den Schülern naturwissenschaftliche Konstrukte und Konzepte näher zu bringen und im Anschluss zu überprüfen, inwieweit die vermittelten Zusammenhänge von den Schülern verstanden und verinnerlicht wurden, können spezielle Aufgaben während und geeignete Rückmeldeformate nach den Tests von Bedeutung sein. Da naturwissenschaftliche Fächer besondere Anforderungen an das Design der Diagnoseaufgaben und die eingesetzten Rückmeldeformate stellen, gilt es herauszufinden, welche sich in bisherigen Forschungsprojekten als besonders effektiv bzw. ineffektiv erwiesen haben. Neben den biologiedidaktischen Fragestellungen, treten mediendidaktische Fragen in den Fokus des Interesses, denn um bestimmte Arten von Feedback geben zu können (individuelles, adaptives oder auch unmittelbares), benötigt man computerunterstützte formative Test- und Diagnoseverfahren. Darum sollen die bisher eingesetzten Möglichkeiten elektronischer formativer Leistungsmessung genauer beleuchtet werden. Zum Abschluss des Theorieteils sollen die auftretenden Forschungsdesiderata der zuvor genannten Bereiche und die Anknüpfungspunkte der Studie an diese Defizite aufgezeigt werden. Die dort aufgestellte Hypothese / Forschungsfrage dient als Leitfaden der Untersuchung und wird bei der Diskussion der Ergebnisse auf ihre Richtigkeit hin überprüft.

2.2 Formative Leistungsmessung

In Zeiten, wo es in bildungspolitischen Diskussionen immer häufiger um die Individualisierung von Lernprozessen, eine individuelle und differenzierte Förderung der Schüler und kompetenzorientierten Unterricht (siehe Bildungsstandards) geht, rückt auch die Frage nach effektiven Diagnoseinstrumenten, die die Erreichung dieser Ziele unterstützen können, weiter in den Fo-

kus des Interesses. Um Schüler in ihrem Lernprozess individuell begleiten und fördern zu können, bedarf es einer genauen und immer wiederkehrenden, in den Unterrichtsprozess integrierten Diagnose ihres Wissensstands und ihrer Lernstrategien. Und nicht nur die Diagnose dieser Aspekte ist von großer Bedeutung, sondern vor allem die Schlüsse, die die Lehrkraft daraus zieht und wie sie anschließend mit Hilfe der erzielten Resultate den weiteren Unterrichtsverlauf individuell erarbeitet und diesen immer wieder an die veränderten Bedürfnisse der Lerner anpasst. Im Zuge dieser Diskussionen und aufgrund der veränderten Bedürfnisse einer Wissensgesellschaft, wird die Forderung nach neuen Lehr- und Lernkonzepten immer stärker (Segers, Dochy & de Corte 1999) und die traditionellen Formen der Leistungsmessung zunehmend in die Kritik genommen (Achtenhagen & Baumert 1999; Dochy 2001; Pellegrino et al. 2001; Broadfoot & Black 2004; Birenbaum et al. 2006). Unter traditionellen Formen verstehen sich vor allem summative Leistungstests, die zu einer abschließenden Bewertung (Notengebung) führen und vor allem Selektionsfunktion haben (Scriven 1967). Zusätzlich zu dem Umstand, dass es gerade bei der Notengebung Mängel in Hinsicht auf die zentralen Gütekriterien (Objektivität, Reliabilität und Validität) gibt (Hochweber 2010), kommt belastend hinzu, dass summative Tests nur zur Beurteilung, aber nicht zur Förderung der Schülerleistung herangezogen werden. Pellegrino, Chudowsky & Glaser (2001) unterstreichen, dass beim summativen assessment allein das Lernergebnis im Vordergrund steht und die Prozesse, die zu diesem Resultat geführt haben, nicht berücksichtigt werden (vgl. Bürgermeister 2014). Harlen & Winter (2004) weisen darauf hin, dass Zensuren nur rückblickend fungieren und keinen Aufgabenbezug herstellen, also keine Informationen darüber preisgeben, in welchen Bereichen sich die Schüler verbessern sollten. Noten und somit summative Leistungsbeurteilungen dienen also eher der Kontrolle, als der Information über Fort- und Rückschritte, und den weiteren Lernprozess der Schüler. Die Notenvergabe sorgt dabei häufig für Leistungsdruck und Demotivation bei den Schülern. Auf der anderen Seite werden Noten benötigt, um laufbahnbedingte Entscheidungen für die Lerner zu treffen und sind wahrscheinlich deshalb auch in Zukunft nicht aus dem Unterrichtsalltag wegzudenken.

Ein Ausweg aus diesem Dilemma kann sein, alternative, förderorientierte Methoden der Leistungsbeurteilung, die den Lernprozess der Schüler begleiten, in den Unterricht einzubinden.

Eine Möglichkeit förderorientierter Leistungsmessung stellen formative Leistungstests dar, die immer wieder in den Lernprozess (die Unterrichtseinheit) integriert werden können, um die

Lernfortschritte der Schüler zu überprüfen, Defizite bei den Lernern zu isolieren und daraufhin deren individuelle Förderung, sowie den nachfolgenden Unterricht zu planen. Anders als eine summative Evaluation, die zu einer abschließenden Bewertung führt, zielt eine formative Evaluation auf die Veränderung laufender Programme (Scriven 1967).

Bei Smit (2009) wird der Begriff der formativen Beurteilung dann verwendet, „wenn die aus der Beurteilung resultierenden Informationen in den Unterrichts- und Lernprozess mit einer möglichst förderlichen Wirkung zurück gespeist werden“ (S. 32). Die Kompetenzen der Schüler sollen also durch die kontinuierliche und optimale Ausrichtung der Unterrichtsprozesse an deren Entwicklungsstand gefördert werden (vgl. Bürgermeister 2014). Dabei sollen die Lernenden aktiv mit einbezogen werden und zusammen mit der Lehrkraft gemeinsame Lernziele formulieren (Heritage 2007).
Black und Wiliam (2009, S.9) verdeutlichen dies in ihrer Definition des *“formative assessments”*: “Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited."

Auch Bloom (1974) hat bereits eine optimale Adaption des Unterrichts durch die Lehrkraft und häufige Überprüfungen des Wissenszuwachses der Lernenden durch Leistungstests mit schülerspezifischem Feedback während der Unterrichtseinheiten gefordert. Die Ergebnisse dieser Tests soll die Lehrkraft nutzen, um den folgenden Unterricht effektiver zu planen und individuelle Maßnahmen zu entwickeln, die die Lerner während ihres Lernprozesses unterstützen.

Formative Leistungsmessungen sollen demnach dabei helfen:

- Den momentanen Wissenstand der Schüler zu diagnostizieren.
- Anhand der Diagnose und der damit aufgedeckten Defizite der Schüler, individuelle Fördermaßen für die einzelnen Lerner zu entwickeln (Lernstrategien, Übungsaufgaben, etc.) und auch
- den nachfolgenden Unterricht für die gesamte Klasse anzupassen.

Ein weiteres, langfristigeres Ziel ist es, die Schüler soweit anzuleiten und zu begleiten, bis diese selbst dazu in der Lage sind, eigene Wissenslücken aufzudecken und gezielt Lernstrategien zu entwickeln, um diese Lücken zu schließen. Formative Leistungsmessung unterstützt und fördert also selbstgesteuertes Lernen (Dochy 2001). Klieme et al. (2003) äußern sich dazu wie folgt: „Das ‚Lernen des Lernens' ist die grundlegende, für das gesamte Leben unverzichtbare Kompetenz, die in modernen, offenen Gesellschaften in schulischer Arbeit generalisiert werden muss."

Durch den Einsatz formativer Leistungstests verspricht man sich neben einer besseren Begleitung der Schüler in ihrem Lernprozess auch eine Verbesserung der Diagnosekompetenz von Lehrkräften und einen eher auf die individuellen Voraussetzungen der Schüler abgestimmten und damit effektiveren Unterricht (Maier & Randler 2011).

Smit (2008, S.385) schreibt zu diesem Thema:
„Spätestens seit Black und Williams (1998) Metaanalyse zur Bedeutung von formativer Beurteilung beginnt sich die Erkenntnis durchzusetzen, dass sich eine bessere Abstimmung von Beurteilung, Curriculum und Lernen förderlich auf das Lernen der Schüler und Schülerinnen auswirkt."

Das bedeutet, dass Unterricht nur erfolgreich sein kann, wenn es eine ständige Wechselbeziehung zwischen Lehrendem, Lernenden und Lernumwelt gibt. Nach Hattie & Timperley (2007) und Black & Wiliam (2009) sind schulische Leistungsdiagnosen dann als formativ zu bezeichnen, wenn sie folgende Aspekte berücksichtigen:

1) Die Lehrperson führt in regelmäßigen Abständen diagnostische Tests durch, die aufzeigen, wo die Lerner im Lernprozess stehen und nehmen diese Resultate zum Anlass, den weiterführenden Unterricht und die ausstehenden Lernziele immer wieder an die Bedürfnisse der Schüler zu adaptieren.
2) Die Lehrkraft gibt den Lernern eine individuelle Rückmeldung zu ihrem Wissensstand und hilft bei der Erarbeitung weiterführender Lernschritte.
3) Die Schüler lassen sich in regelmäßigen Abständen über ihren Leistungszuwachs informieren und ziehen im positivsten Fall Rückschlüsse darauf, wie ihre zukünftige Lernstrategie aussehen muss, um Wissenslücken zu schließen.

Dabei kann *formative assessment* aber nicht nur im Einsatz geplanter formativer Leistungstests bestehen, sondern auch spontan während des Unterrichts stattfinden, daher wird in diesem Zusammenhang auch zwischen formellen und informellen Beurteilungsmethoden unterschieden (Bell & Cowie 2001; Ruiz-Primo & Furtak 2007). Shavelson et al. (2008, S. 300) schreiben über informelle Beurteilungsmethoden folgendes: „On-the-fly formative assessment arises when a "teachable moment" unexpectedly occurs, for example, when a teacher circulating and listening to the conversation among students in small groups overhears a student say that, as a consequence of her or his experiment." Formelle Beurteilungsmethoden werden von Shavelson et al. (2008) hingegen, als geplant und auf ein bestimmtes Ziel ausgerichtet beschrieben. In Tabelle 1 werden die beiden Beurteilungsmethoden noch einmal genauer gegenübergestellt:

Tabelle 1: Formative assessment – Beurteilungsmethoden (vgl. Bürgermeister 2014)

Formelle Beurteilungsmethode	Informelle Beurteilungsmethode
o Im Voraus geplante Beurteilungen o Zur Dokumentation des Lernstandes der Schülerinnen o Beinhaltet das Sammeln und Interpretieren von Informationen über den Wissenstand der Schüler und o die darauffolgende Planung des weiteren unterrichtlichen Vorgehens. o Bezeichnet als „planned-for interaction" (vorab geplante Frage-Antwort Sequenzen) und „curriculum–embedded assessment" (fortlaufende Beurteilungsaktivitäten) (Heritage 2007, Shavelson et al. 2008)	o Spontane Erhebung des Lernstandes im Unterrichtsalltag (zum Beispiel in Unterrichtsgesprächen) o Eher improvisierte Situationen zwischen Schülern und Mitschülern, Kleingruppen, der gesamten Klasse oder der Lehrkraft. o Kann auch nonverbal ablaufen (durch Beobachtungen) (Ruiz-Primo & Furtak 2007) o Bestehen vor allem aus lernförderlichen Rückmeldungen an die Schüler, so dass eine Vertiefung, Klärung und Ergänzung des Wissens stattfinden kann (vgl. Bürgermeister 2014). o Bezeichnet als „On the fly assessment" (Heritage 2007, Shavelson et al. 2008)

Der Prozess der formativen Beurteilung selbst wird oft als zyklisch bezeichnet. Bell & Cowie (2001) schreiben zum Beispiel, dass formative Leistungsbeurteilung verschiedene Phasen beinhaltet, die voneinander abhängig sind und im Unterrichtsprozess fortlaufend wiederholt werden können. Diese Phasen beinhalten die Gewinnung der Informationen, deren Interpretation und das daran anschließende pädagogische Handeln.

Bürgermeister (2014) verortet in diesem Kontext auch den Förderkreislauf, der im deutschsprachigen Raum als praxisorientiertes Konzept dient, eine förderorientierte Beurteilung in den Unterrichtsalltag zu integrieren (Smit 2009).

Dieser Kreislauf beinhaltet vier Schritte - das Festlegen von Zielen, das Beobachten, das Beurteilen und das Fördern:

1) Ziele festlegen: Dabei ist es wichtig, die Lernziele gemeinsam mit den Lernern zu erarbeiten, diese gelegentlich zu überdenken und auch, wenn nötig, anzupassen und diese im Unterricht konsequent zu verfolgen.
2) Die Phasen der Beobachtung und Beurteilung (3) umfassen die Diagnose der Lernvoraussetzungen und Lernentwicklungen der Schüler.
4) Bei der Förderung geht es darum, den Schüler individuell in seinem Lernprozess zu begleiten und gemeinsam mit ihm Lernstrategien zu entwickeln, um aufgestellte Ziele zu erreichen und Wissenslücken zu schließen. Kernelement dieser Phase sind die differenzierten Rückmeldungen (Smit 2009).

Harlen (2011, S.90) betont bei seinem zyklischen Modell (Siehe Abbildung 1), das immer der Lernende im Zentrum des Interesses steht und sich dieser stets im Austausch mit der Lehrkraft befindet (Bürgermeister 2014).

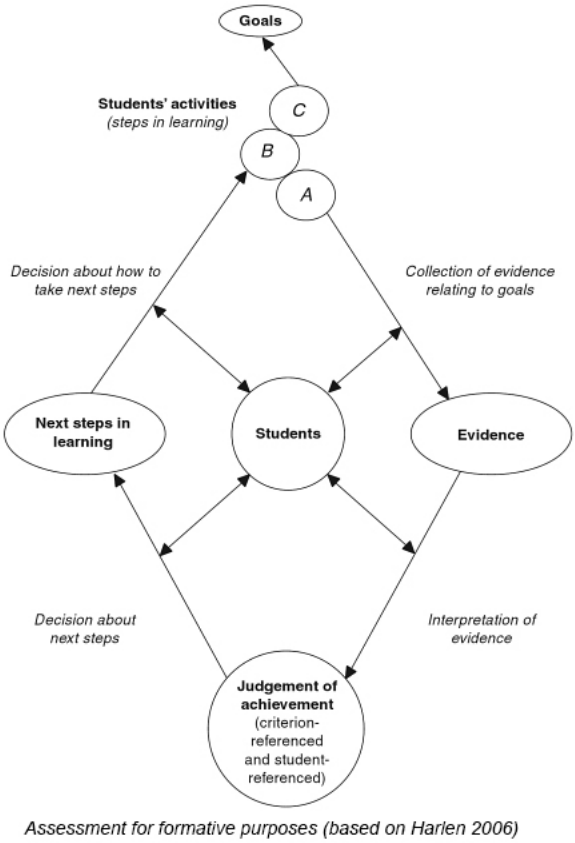


Abbildung 1: Formative Assessment - Zyklisches Modell nach Harlen (2011)

Besonders geprägt wurde der Begriff der formativen Leistungsdiagnostik durch einige Metaanalysen und Literaturübersichten, die den Fokus auf Leistungsmessungen als "Unterrichtsintervention" auf Schülerleistungen und motivationale Variablen legen (Fuchs & Fuchs 1986; Crooks 1988; Natriello 1987; Black & Wiliam 1998).

Es zeigte sich, dass Leistungsmessungen, die die unten aufgeführten Merkmale aufwiesen, zu hohen Effektstärken führten (z.B. Black & Wiliam 1998, 2009; Harlen & Winter 2004):

- Zur Optimierung der Lehr- und Lernprozesse wurden Ergebnisse / Informationen über Lernprozesse und Lernprodukte herangezogen.
- Es fand bereits vor der Leistungsmessung ein Austausch zwischen Lehrkraft und Lernern über die Ziele des Unterrichts, die Lernziele und die Kriterien zur Bewertung der Schülerleistungen statt.
- Die Schüler wurden an der Messung und bei der Beurteilung ihrer eigenen Leistungen aktiv mit einbezogen (*peer and self assessment*).

Die Ergebnisse der Arbeit von Black und Wiliam (1998) führten in Großbritannien dazu, dass Projekte initiiert wurden, die den Einsatz formativer Leistungstests, in einem von *high stakes testing* geprägten Schulsystem vorantrieben und auch in der angloamerikanisch geprägten Literatur zeigt sich verstärkt, dass Effekte von Leistungsmessungen auf Motivation und Lernerfolg betont und auch immer häufiger erforscht werden.

2.2.1 Grundlegende Aspekte formativer Leistungsmessung nach Black & Wiliam (2009)

Das folgende Zitat von Maier & Randler (2011) zeigt die Vielschichtigkeit von *formative assessment* Strategien: „Durch die vielfältigen empirischen Belege ist der Begriff FLM sehr attraktiv und gleichzeitig aber auch problematisch, weil damit sehr unterschiedliche Methoden der Leistungsmessung (zeitliche Reichweite, Format, Technikunterstützung, etc.) in unterschiedlichen institutionellen Kontexten (Primarstufe, Sekundarstufe, Universität) und Fachdomänen betrachtet werden."

So unterschiedlich die Ansätze und Ausprägungen formativer Leistungsmessung auch sein mögen, so gibt es doch einen „kleinsten gemeinsamen Nenner", also ein paar grundlegende Aspekte der formativen Leistungsdiagnostik, die bei Ramaprasad (1983) und Black & Wiliam (2009) folgendermaßen zusammengefasst werden:

(1) Lernziele sollen gemeinsam mit den Lernern geklärt und die Kriterien für die Zielerreichung konkretisiert werden;
(2) Es sollen verschiedene Anlässe geschaffen werden (zum Beispiel formative Leistungstests / Diagnoseinstrumente), in denen Hinweise auf das aktuelle Verständnis der Lerninhalte generiert werden;
(3) Dabei gilt es, effektive und zielorientierte Rückmeldungen zu formulieren und zur Verfügung zu stellen;
(4) Durch das Feedback sollen weitere Lernschritte aufgezeigt und dadurch die Selbstregulationskompetenz der Schüler und die Entwicklung individueller Lernstrategien durch die Lerner gefördert werden;
(5) Der Lehr-Lernprozess soll aufgrund der Ergebnisse formativer Leistungsmessung im weiteren Verlauf angepasst werden (individuelle Lernstrategien und adaptierter Unterricht).

„Wie diese Prinzipien technisch und unterrichtsmethodisch umgesetzt werden ist allerdings sehr verschieden. Auch werden nicht immer alle Aspekte formativer Leistungsmessung gleichermaßen stark berücksichtigt.“ (Maier & Randler 2011)

Diese Gegebenheit zeigt, dass viele Aktionen, die im Klassenzimmer zur Beurteilung herangezogen werden, auf formative Art und Weise verwendet werden können. Lehrkräfte können ihre Schüler dabei beobachten, wie sie lernen, wie sie sich in Lerngruppen einbringen und wie sie mit Aufgaben umgehen, die ihnen gestellt werden. Weiterhin könne sie diagnostische Tests stellen und diese auswerten, um Empfehlungen für individuelle Lernstrategien der Schüler auszusprechen, individuelle Lernpläne zu erstellen oder den nachfolgenden Unterricht adaptiv zu gestalten (vgl. Smit 2008).
In diesem Zusammenhang kann man auch summative Tests dazu heranziehen, um Auskunft über den Wissenstand und die Wissenslücken der Lernenden zu bekommen und das zukünftige Lernverhalten zu modifizieren. Letztendlich ist das Ziel formativer Leistungsmessung, das selbstgesteuerte Lernen der Schüler zu fördern. Dies bedeutet, dass Lehrer ihren Schülern immer wieder bei der Auswertung und Interpretation diagnostischer Informationen helfend zur Seite stehen sollten, um sie in diesen Prozessen zu schulen. Eine bleibende Änderung im Lern- und Arbeitsverhalten der Schüler zu verwirklichen, dauert allerdings sehr lange. Es muss immer wieder daran gearbeitet werden, um bestimmte Abläufe und Arbeitsweisen zu automatisieren.

Eine der genauesten Metaanalysen zum Thema *formative assessment* wurde von Kingston und Nash (2011) veröffentlicht. Ihre Arbeit basiert auf der Metaanlyse von Black und Wiliam (1998), allerdings berücksichtigen sie nur die Studien, denen ein sorgfältiges empirisches Design zugrunde lag. Nach dem „Aussortieren“ blieben lediglich 13 Studien mit einer mittleren Effektgröße von nur .20 übrig. Eine genauere Analyse zeigte, dass das Fach von großer Bedeutung ist. So ist formative assessment zum Beispiel im Englischunterricht effektiver, als in Mathematik oder in naturwissenschaftlichen Fächern. Dieses Ergebnis nimmt auch Bezug auf Bennett's (2011) Aussage, dass die Fächer in den meisten Studien nicht adäquat berücksichtigt wurden.

Zusammenfassend lässt sich sagen, dass sich formative Leistungsmessungen dadurch auszeichnen, dass sie als Diagnosesystem immer wieder während des Lernprozesses /der Unterrichts-

einheit eingesetzt werden, um den momentanen Wissensstand der Schüler festzuhalten. Die daraus resultierend Erkenntnisse ermöglichen es der Lehrkraft, den Schülern individuelle und differenzierte Rückmeldungen zu geben und diese bei darauffolgenden Lernprozessen durch geeignete Fördermaßnahmen zu unterstützen. Die Ergebnisse der formativen Leistungstests werden außerdem dazu herangezogen, den weiteren Unterricht und nachfolgende Messzeitpunkte zu planen und zu entwickeln. So ist formative assessment ein zyklischer Prozess, bei dem der Lerner mit seinen Bedürfnissen im Mittelpunkt und im ständigen Austausch mit der Lehrkraft steht. Letztendlich soll formative assessment den Lerner dazu bringen, selbst seine Wissenslücken zu erkennen und Lernstrategien zu entwickeln, um diese zu schließen. Die Hinführung zu selbstgesteuerten Lernprozessen benötigt jedoch Zeit, denn dies geht nicht nur mit Veränderungen von Schülereinstellungen, sondern auch mit Modifizierungen des Lernverhaltens einher. Wie effektiv formative Leistungstests sind, hängt von mehreren Faktoren ab (Art der Rückmeldung, Fach, Schülerpersönlichkeit, etc.).

Das Feedback an die Schüler wird dabei als Kernelement der formativen Leistungsmessung angesehen (Heritage 2007) und soll in den nächsten Abschnitten näher untersucht werden. Es wird aufgezeigt, wie unterschiedlich Rückmeldungen gestaltet sein können (hinsichtlich ihrer Ausführlichkeit, Bezugsnorm, Formulierung, etc.), und inwieweit es einen Effekt hat, wer das Feedback gibt (Lehrer, Computer, Mitschüler), oder wie offen der Lerner für die Annahme des Feedbacks ist (Aspekte der Schülerpersönlichkeit).

2.2.2 Feedback als zentrale Erklärung von Effekten formativer Leistungsmessung

Rückmeldungen sind ein wichtiger Aspekt, wenn es um das Lehren und Lernen geht. Unter dem Begriff Feedback können im Kontext von Lehr- und Lernsituationen Informationen verstanden werden, die einem Lernenden während oder nach der Bearbeitung einer Aufgaben von einem Erwachsenen, fähigeren Peer oder einem computergestützten tutoriellen System angeboten werden. Dessen Ziel ist es, lösungsförderlich für die aktuelle oder eine folgende Aufgabe zu sein (Narciss 2006). Hattie (2003) definiert Feedback als „actions or information provided by an agent (e.g. teacher, peer, book, parent, experience) that provides information regarding aspects of one's performance or understanding" (S. 2).

Viele Lehrmaterialien und auch Lehrmethoden beinhalten alle möglichen Arten von Feedback. Auch im normalen Unterrichtsgespräch kommt es immer wieder zu Rückmeldungen durch den Lehrer und die Mitschüler.

Dabei können sich diese Rückmeldungen in sehr vielen Aspekten unterscheiden – in der folgenden Aufzählung befinden sich einige dieser Unterscheidungsmerkmale:

1) In ihrer Darreichungsform: schriftlich am Ende einer Probearbeit, mündlich während des Unterrichtsgespräches durch den Lehrer oder die Mitschüler, oder nach einer Testung vom Computer.
2) In ihrer Ausführlichkeit: Dichotom (richtig – falsch), ausführlich, oder gar mit weiterführenden Informationen zu angrenzenden Themen und Inhalten.
3) Im Zeitpunkt der Darreichung: Feedback kann direkt nach einer Aktion gegeben werden (spontan), oder später, wenn Probearbeiten erst korrigiert werden müssen, oder auch gleich während der Arbeit an einem Thema, wenn die Rückmeldetexte vom Computer kommen.
4) In ihrem Inhalt und ihrer Informationsqualität: Feedback kann in Bezug auf die Aufgabe Aufgabenebene), den dazugehörigen Lösungsweg (Aufgabenprozesse), oder auf einer Metaebene formuliert werden.
5) In ihrer Komplexität: Rückmeldetexte können sehr allgemein gehalten, oder aber auch individuell auf einen Lerner zugeschnitten sein, oder sich im besten Fall sogar adaptiv an den Lernprozess des Schülers anpassen.
6) In ihrer sozialen Dimension: Self Assessment, Peer Assessment, Lehrperson, anonymisiertes Feedback.
7) In dem Aspekt, welche Wissensart und Komplexität des Lernziels sie ansprechen (prozedurales, oder deklaratives Wissen (Fakten- und Konzeptwissen))
8) Feedback kann negativ, positiv oder neutral formuliert werden.
9) …

Die von mir erstellte Aufzählung ist nicht vollständig, soll aber zeigen, dass Rückmeldung nicht gleich Rückmeldung ist und es viele Arten von Feedback gibt, die auf ihre Effektivität beim Wissenszuwachs und die Schülermotivation hin untersucht werden können.

Für alle gilt: Durch Rückmeldungen sollen Schüler zu einem selbstgesteuerten Lernen geführt werden und Feedback ist der zentrale Aspekt formativer Bewertungen. Daher gibt es eine Vielzahl an Forschungsprojekten in sehr unterschiedlichen Bereichen des Lernens und Erziehens, vor allem aber in der Lern- und Motivationspsychologie, die den Einsatz unterschiedlicher Rückmeldestrategien näher untersuchen. Im Fokus dieser Untersuchungen steht nicht, ob die Leistungsmessungen an sich objektiv oder fair verlaufen sind, sondern ob sie dazu beitragen, dass der Lerner sich gerne weiter mit dem behandelten Thema auseinandersetzt und wie erfolgreich er dabei ist.

Die zahlreichen Ergebnisse der Feedbackforschung wurden in Metaanalysen zusammengefasst (z.B. Bangert-Drowns et al. 1991; Kluger & DeNisi 1996). Aus diesen Befunden wurden Aspekte abgeleitet, wie schulische Leistungsdiagnosen lernförderlich gestaltet werden können (z.B. Hattie & Timperley 2007). Dabei ging man davon aus, dass nicht die Leistungsmessung an sich, sondern die Qualität der Rückmeldung und die Art und Weise, wie Lerner mit dieser umgehen, entscheidend dafür sind, wie erfolgreich das nachfolgende Lernen verläuft. (Maier & Randler 2011)

In den folgenden Abschnitten sollen die Ergebnisse der oben angesprochenen Metaanalysen genauer betrachtet werden. Insgesamt lässt sich jedoch feststellen, dass die Effektgrößen in den Feedbackstudien sehr widersprüchlich ausfallen. Die unterschiedlichen Feedbackeffekte können dadurch erklärt werden, dass verschiedene Rückmeldeformate eingesetzt wurden und auch unterschiedliche äußere Bedingungen während der Untersuchungen vorlagen (Lerninhalte, Unterrichtsfach, Alter der Lerner, Aufgabenkomplexität, Schülerpersönlichkeit, etc.).

Die umfassendste Metaanalyse über Feedbackinterventionen stammt von Kluger & DeNisi (1996). Sie schätzten Feedback, als ein gutes Werkzeug ein, um Lernprozesse zu optimieren, trotz der negativen, oder auch nicht vorhandenen Effekte früherer Studien. Ihre Bestrebungen beziehen sich darauf, eine ausführliche Feedbackinterventionstheorie zu entwickeln.
Dabei haben sie unterschiedliche Feedbackmerkmale isoliert und untersucht, um zu erklären, wie es zu den verschiedenen Effektstärken bei den Rückmeldungen kam. „Kluger & DeNisi (1996, 1998) gehen davon aus, dass Feedbackinterventionen die Aufmerksamkeit steuern kön-

nen und dass der Ort der Aufmerksamkeitsfokussierung wesentlich zur Erklärung der Effektivität von Feedbackinterventionen beitragen kann.“ (Maier & Randler 2011)

Die drei Ebenen der Aufmerksamkeitsfokussierung von Kluger & DeNisi (1996, 1998) wurden von Hattie und Timperley (2007) aufgegriffen und um eine weitere Ebene ergänzt:

1) Ebene der Aufgabenlösung (*FT: feedback about task performance*),
2) Ebene der Aufgabenbearbeitungsprozesse (*FP: feedback about the process of understanding how to do a task*);
3) Ebene der Selbstregulation (*FR: feedback about metacognitive regulation*);
4) Ebene der selbstbezogenen Kognitionen (*FS: feedback about the self*).

Dabei betonen Hattie und Timplerley (2007), dass ein wirksames Feedback auf drei Ebenen verortet sein sollte, um den Lernprozess der Schüler zu unterstützen (Ebenen 1-3). Die vierte, von den Autoren hinzugefügte Ebene, die oft durch Rückmeldungen im Unterricht angesprochen wird (zum Beispiel durch ein „Gut gemacht“) gilt jedoch als wenig informativ und daher als wenig förderlich, wenn es um die positive Entwicklung von Lernprozessen geht (vgl. Bürgermeister 2014).

Diesen vier Ebenen können nun die Ergebnisse der Feedbackforschung zugeordnet werden:

Zu 1) FT: Unmittelbares Feedback auf der Aufgabenebene (z.B. *knowledge of results*)

- führt bei wenig komplexen Aufgaben (z.B. Silben memorieren) zu hohen Effektstärken (vgl. Bangert-Drowns et al. 1991). Diese Effekte sind allerdings nicht auf andere Aufgaben übertragbar. Dieser Feedbacktyp ist umso effektiver, je schneller die Rückmeldung erfolgt.
- Positive Rückmeldungen zur Aufgabenlösung sind hoch effektiv; negative Rückmeldungen sind nur dann erfolgreich, wenn Korrekturhinweise gegeben werden, den Lernern also das richtige Ergebnis zur Verfügung gestellt wird (*knowledge of correct result*).

Zu 2) FP: Feedback zum Aufgabenlösungsprozess

- bringt die Schüler dazu, ihre eigenen Lösungswege kritisch zu hinterfragen und zu überprüfen, ihre Fehler zu reflektieren und über neue Lösungsansätze nachzudenken. Diese Art von Rückmeldung kann zu einem tieferen Verständnis der Lerninhalte und Konzepte führen und ist effektiver, wenn sie zeitverzögert erfolgt.
- Außerdem wird angenommen, dass ein Transfer auf ähnliche Aufgabenstellungen auf dieser Ebene eher möglich ist und somit FP zu einem nachhaltigeren Lernen führen kann.

Zu 3) FR: Feedback zur Selbstregulation

- Setzt sich mit der Frage auseinander, inwieweit Schüler ihre Lernstrategien an das zu erreichende Lernziel anpassen und im weiteren Lernverlauf weiter überwachen.
- Zur Selbstregulation gehören konkrete Handlungsstrategien, Einstellungen und Wahrnehmungsmuster.
- Rückmeldungen, die die Selbstregulation betreffen, haben nichts mit den rein fachlichen Aufgabenbearbeitungsprozessen zu tun. In der Feedbackinterventionstheorie geht man deshalb davon aus, dass Effekte, die durch Feedback zu Selbstregulationsstrategien entstehen, nur über einen langen Zeitraum hinweg feststellbar sind.

Zu 4) FS: Die Feedbackinterventionstheorie besagt,

- dass bei einer Aufmerksamkeitsverlagerung auf das Selbst, ausgelöst durch Lob, Kritik oder normative Hinweise (z.B. Vergleich mit anderen Schülern), eine eher geringere Leistungssteigerung zu erwarten ist, weil der Lernende durch die Beschäftigung mit dem eigenen Ich von der Aufgabenbearbeitung abgelenkt wird.
- Dass diese Annahme durchaus richtig sein kann, zeigen Teilbefunde der Metaanalysen, die darauf hinweisen, dass computerbasierte Rückmeldungen effektiver sind als verbale Rückmeldungen.
- Rein kriteriumsbezogene Rückmeldungen sind insgesamt effektiver. Allerdings bedarf es einer Differenzierung dieses Ergebnisses: Bei komplexeren Aufgaben kommt es zu reduzierten Effekten bei der Leistung, weil kognitive Kapazitäten

durch selbstbezogene Gedanken abgezogen werden. Bei einfachen Aufgaben hingegen, führt selbstbezogenes Feedback (Lob) zu größerer Anstrengung im Lernprozess und damit auch zu einer Leistungssteigerung.

Weiterhin beschreiben Kluger & DeNisi (1996) in ihrer Feedbackinterventionstheorie, dass auch die Persönlichkeitsmerkmale von Lernenden einen großen Einfluss auf Feedbackeffekte haben. (vgl. Maier & Randler 2011)

Darüber hinaus wurde in verschiedenen Studien herausgefunden, wie man Feedback und dessen Effekte auf die Lernleistung verbessern kann:

Zu einer Steigerung der Leistungsmotivation kommt es, wenn die Lehrkraft eine Rückmeldung gibt, inwieweit die Leistung eines Schülers den Anforderungen der gestellten Aufgabe entsprochen hat (kriteriale Bezugsnorm), also an welcher Stelle der Lernentwicklung sich der Schüler befindet (Heritage 2007). Auch ein Vergleich mit vorhergehenden Leistungen des Schülers (individuelle Bezugsnorm), führt zu einer höheren Motivation. Diese Form des Feedbacks kommt in der Schule allerdings relativ selten vor.

Lehrer können den Lernenden bereits in der Phase der Zielsetzung Hinweise darauf geben, welche Zielorientierung von Bedeutung ist (Butler 2005). Dies wird erreicht, durch das gemeinsame Besprechen der Lerninhalte und Beurteilungskriterien, die bei der abschließenden Bewertung herangezogen werden. Auch bei der Aufrechterhaltung der Lernmotivation und der abschließenden Bewertung der Lernergebnisse sollte das Lehrerfeedback zu den festgelegten Lernzielen der Schüler passen (zielrelevante Rückmeldeinformationen).

Darüber hinaus sollte das Feedback Informationen enthalten, wie weit der Lernende vom festgesetzten Lernziel entfernt ist und wie er die nachfolgenden Lernanstrengungen gestalten kann, um das angestrebte Ziel zu erreichen (Black & Wiliam 2009).

Sadler (1998) hebt hervor, dass die Qualität der Rückmeldung (Akkuratheit, Angemessenheit und Verständlichkeit) und wie diese vermittelt wird essentiell für eine positive Unterstützung des Lernprozesses ist (in: Bürgermeister 2014).

In der Schule bekommen Lernende häufig Feedback für ihre Leistungen, sei es verbal während des Unterrichtsgesprächs durch ihre Mitschüler oder die Lehrkraft oder schriftlich bei der Beurteilung der Leistungstests. Häufig reflektiert man als Lehrkraft während des Unterrichts nicht

vor jeder Rückmeldung, die man dem Schüler gibt, wie man diese am besten formuliert (positiv, negativ oder neutral) und ob in einem bestimmten Fall eine kurze Rückmeldung effektiver wäre, als eine ausführliche Schilderung des kompletten Sachverhaltes.
Man reagiert eher intuitiv und versucht auf die Schülerpersönlichkeit und die die äußeren Bedingungen (verbleibende Zeit, etc.) Rücksicht zu nehmen. Umso interessanter ist es, sich anzusehen, welche Art von Feedback besonders effektiv sein kann, um die Leistung und die Motivation der Schüler zu steigern. Bisherige Studien zeigen, dass Feedbackeffekte sehr inkonsistent sind und von bestimmten Faktoren abhängen (Fach, Lerninhalte, Komplexität der Aufgaben, etc.).

Hierzu gibt es bereits grundlegende Aussagen der (psychologischen) Feedbackforschung, die für den Einsatz und das Design formativer Leistungstests in einem schulischen Kontext interessant sind und dafür herangezogen werden können. Bangert – Drowns et al. (1991) zum Beispiel, beschreiben in ihrer Metaanalyse zu Feedbackeffekten Variablen, die höhere Effektstärken erklären können:
Zum Beispiel ist der Informationsgehalt der Rückmeldung von Bedeutung, wenn es um Effektstärken geht – so gab es meist keine Effekte bei einfachem Feedback, bei dem nur das Ergebnis angezeigt wurde (*knowledge of results*). Zu höheren Effekten kam es, wenn die richtige Antwort rückgemeldet wurde (*knowledge of correct result*) und wenn den Lernern weiterführende Informationen zur Lösung oder zum Lösungsweg zur Verfügung gestellt wurden (*elaborated feedback*). Bangert – Drowns et al. (1991) unterscheiden dabei verschiedene Arten elaborierten Feedbacks:

- *Task specific elaboration:* Die komplette Aufgabenstellung mit der richtigen Antwort wird zur Verfügung gestellt,
- *instruction based elaboration:* Der im Unterricht behandelte Lösungsweg für die Aufgabe wird noch einmal erläutert und
- *extra-instructional elaboration:* Dem Lerner werden zusätzliche, weiterführende oder vertiefende, noch nicht im Unterricht besprochene Inhalte, Beispiele, Veranschaulichungen oder Erklärungen präsentiert.

Die Durchführung eines Pretests, wirkt sich negativ auf die Feedbackeffekte aus und reduziert diese.
(vgl. Maier & Randler 2011)

Stellt man den Lehrenden in den Fokus lassen sich folgende Empfehlungen für die Feedbacknutzung geben:
Lehrkräfte können ihre Diagnoseergebnisse nutzen, um Fördermaßnahmen zu entwickeln und ihren nachfolgenden Unterricht zu planen, also adaptiv an die Schülerbedürfnisse anzupassen (Harlen & Winter 2004, Maier 2010). Dabei sollten die Unterrichtsprozesse kontinuierlich an die Entwicklungen der Schüler angepasst werden (Smit 2009).
Darüber hinaus können Lehrer die Informationen, die sie aus den formativen Leistungstests über die Schüler und ihren Stand im aktuellen Lernprozess erhalten dazu verwenden, um die bestehenden Lernziele zu überdenken, oder diese neu zu formulieren und zukünftige Lernschritte (Einsatz passender Unterrichtsmethoden und –materialien) festzulegen (Hattie 2009).

Möchte man auf den Zusammenhang zwischen Feedback und Schülermotivation näher eingehen lohnt es sich einen Blick auf die beiden Studien von Rakoczy et al. (2008) zu werfen (die einzigen Studien in der deutschsprachigen Lehr- Lernforschung zum Thema Feedback und Leistungsmessung): Diese untersuchten, wie sich Rückmeldungen von Lehrkräften im Rahmen von *classroom assessment* auf Schülerleistung und Schülermotivation auswirken. Innerhalb einer deutsch-schweizer Videostudie, wurden die Aufnahmen in Hinsicht auf die Lehreräußerungen analysiert. Dabei wurde unterschieden zwischen „informatives vs. evaluatives/bewertendes Feedback“ und „positives vs. negatives Feedback“. Die Feedbackmethoden der Lehrkräfte wurden während der Evaluation nicht beeinflusst oder experimentell variiert. Die Analysen zeigen, dass informatives und positives, bewertendes Feedback positiv mit Leistungsmotivation korreliert. Negatives, bewertendes Feedback hängt hingegen nicht mit motivationalen Variablen zusammen. Allerdings steht keiner der Rückmeldetypen in einem substanziellen Zusammenhang mit den Schülerleistungen.
Crooks (1988) vertritt in diesem Zusammenhang die Auffassung, dass eine aussagekräftige Rückmeldung an die Schüler, deren Motivation und Lernleistung in hohem Maße beeinflussen kann.

Betrachtet man die Forschungsbereiche „formative Leistungsmessung“ und „Feedback“ genauer, muss man feststellen, dass es im Bereich des naturwissenschaftlichen Unterrichts noch nicht so viele Studien gibt. Bisher wurden formative Leistungstests vor allem im Bereich der Sonder-

pädagogik (Fuchs et al. 1994; Fuchs 2004; Codding et al. 2005; Walter 2008) und beim Erlernen von Grundkompetenzen, wie Lesen, Schreiben und Rechnen eingesetzt (Walter 2008). Dort wurden bereits formativ-diagnostische Testverfahren entwickelt, um die Leistungsentwicklung von Schülern zuverlässig, valide und änderungssensibel erfassen zu können. (vgl. Maier & Randler 2011)

Allerdings verändert sich dieser Schwerpunkt langsam, so dass der Einsatz von FLM Techniken und deren Untersuchung in anderen Fächern immer mehr in den Fokus rückt. So wird in Deutschland im Zuge der jüngsten Diskussionen über eine stärkere Kompetenzorientierung, Individualisierung und Differenzierung des Unterrichts auch in der Mathematikdidaktik verstärkt über Aufgabenstellungen mit diagnostischem Potenzial nachgedacht (Leuders 2004, 2006; Büchter & Leuders 2005; Blum et al. 2006; Sjuts 2006, 2007). Eine Möglichkeit, das tiefere konzeptuelle Verständnis der Schüler zu überprüfen, ist zum Beispiel, eine Multiple Choice Frage um eine offene Fragestellung zu erweitern. Bei der Beantwortung dieser zusätzlichen Frage, sollen die Lerner ihre getroffene Auswahl begründen und nochmals reflektieren.

Maier, Hofmann und Zeitler (2012) ziehen als Aufgabenbeispiel eine Mathematikaufgabe aus VERA 8, Mathematik Testheft I; Aufgabe 7 „Brettspiel“ (IQB, 2011), heran:

Bei einem Brettspiel wird ein sechsseitiger Spielwürfel mit den Zahlen 1 bis 6 verwendet. Der Spielstein darf immer um genau so viele Felder weitergeschoben werden, wie die Augenzahl beim einmaligen Werfen des Würfels anzeigt.
Jan wird das Spiel gewinnen, wenn er beim nächsten Wurf eine Fünf wirft.
Wie groß ist die Wahrscheinlichkeit für Jan, mit dem nächsten Wurf das Spiel zu gewinnen?

Kreuze an. ☐ $\frac{1}{5}$ ☐ $\frac{1}{6}$ ☐ $\frac{5}{6}$ ☐ 5

Ergänzt wird: „Begründe deine Antwort…“

Die Vorteile dieser Vorgehensweise liegen darin, dass die Schüler zunächst intuitiv an die Aufgabe herangehen, sich alle 4 Antwortmöglichkeiten ansehen und womöglich erst einmal eine für sie plausible Lösung auswählen. Dabei ist allerdings noch nicht gesagt, warum sie sich genau für diese Lösung entschieden haben. Nur durch die schriftliche Begründung kann die Lehrkraft nachvollziehen, ob der Schüler das grundlegende Konzept (hier: Wahrscheinlichkeit bei einem einfachen Zufallsexperiment) auch wirklich verstanden hat (vgl. Maier, Hofmann, Zeitler 2012).

Dass die Diagnose von Schülerkompetenzen mit herkömmlichen Testaufgaben nur sehr unzureichend gelingt, wurde in der naturwissenschaftlichen Literatur bereits häufig erläutert (Harlen & Winter 2004; Shavelson & Ruiz-Primo 1999).
Keogh & Naylor (2000), McConnell et al. (2006), Harlen (2007), Brown (2008) und Tan & Towndrow (2009) machen darauf aufmerksam, dass es neuer Diagnosesysteme bedarf, um hinreichend auf die Schüler und anspruchsvolle Lernziele eingehen zu können. Als Beispiele für Diagnosemöglichkeiten werden die Beobachtung von Schülern bei der Durchführung von Experimenten, die gemeinsame Analyse von Arbeitsprodukten, der Einsatz von Concept-Maps, kurze Vorwissenstests oder videogestützte Feedbackverfahren genannt.

Einige verschiedene Methoden, die dazu entwickelt wurden, um Fehlvorstellungen von Schülern innerhalb naturwissenschaftlicher Konzepte aufzudecken sollen im Folgenden kurz beispielhaft aufgelistet und erläutert werden:
Shavelson & Ruiz-Primo (1999) beschreiben in ihrem Artikel über „Leistungsbewertung im naturwissenschaftlichen Unterricht“, welche unterschiedlichen Arten naturwissenschaftlichen Wissens es gibt und welche Diagnosemöglichkeiten geeignet sind, um diese verschiedenen Wissensarten abzuprüfen. Dabei bauen diese Wissenslevel in ihrer Komplexität aufeinander auf (siehe Tabelle 2):

Tabelle 2: Wissensarten in den Naturwissenschaften nach Shavelson & Ruiz-Primo (1999)

Wissenslevel	Welche Art von Wissen beinhaltet dieser Wissensaspekt?	Geeignete Diagnosesysteme
Propositionales oder deklaratives Wissen	Wissen über Fakten, Konzepte und Prinzipien. Beispiel: Schüler wissen, was die Begriffe „Kraft“ und „Masse“ in der Physik bedeuten und auch, „dass ein in Bewegung befindlicher Gegenstand seine Bewegung unendlich fortsetzt, wenn nicht andere Kräfte auf ihn einwirken.“	Mit den üblichen Multiple Choice und Short – Answer – Tests gut messbar.
Konzeptuelles Wissen/ Deklaratives Expertenwissen (hoch strukturiert) / Anfängerwissen (weniger stark organisiert)	Anwendung und Verknüpfung der erlernten Fakten, Konzepte und Prinzipien. Die mentale Organisation des Wissens im Schülergedächtnis.	Mit herkömmlichen Paper-Pencil Tests schwer abprüfbar. Mögliche Lösung: Concept Maps oder „Concept Similarity Judgments“, um die kongnitive Struktur des Schülerkonzeptes zu erschließen.
Prozedurales Wissen	Vor allem in den Naturwissenschaften wichtig, weil es zu den fachspezifischen Arbeitstechniken gehört: Wissenschaftler stellen Vermutungen auf und überprüfen diese Theorien mit Hilfe von Experimenten. Es werden Beobachtungen gemacht, während einzelne Variablen verändert und andere kontant gehalten werden.	Tests, bei denen sie Schüler eigenhändig Untersuchungen durchführen. Shavelson & Ruiz-Primo (1999) nennen diese Art von Tests, die in einem Labor oder mit Hilfe einer Computersimulation (nachgestellte Laborsituation am Computer) stattfinden „performance assessment“.
Strategisches Wissen	Konzepte und Verfahren werden in Form von Regeln miteinander kombiniert. Nach diesen Regeln richtet sich dann das Handeln der Experten unter bestimmten Aufgabenstellungen und Arbeitsbedingungen. Man ist sich darüber im Klaren, welches spezifische Wissen wann wie anzuwenden ist. Experten strukturieren dieses Wissen in Form von mentalen Modellen.	Zur Messung dieses Wissensaspektes gibt es bisher keine systematische Forschung und Entwicklung.

Shavelson & Ruiz-Primo (1999) untersuchten, wie Lerner ihr Wissen innerhalb eines naturwissenschaftlichen Konzepts oder Zusammenhangs anordnen und organisieren. Um diese Wissenskonstrukte im Schülergedächtnis aufzudecken, verwendeten sie sowohl Multiple Choice Aufgaben, als auch unterschiedliche Arten von Concept Maps und überprüften deren Qualität. Sie fanden heraus, dass „sich Concepttests und Multiple Choice Tests überschneiden, aber den-

noch unterschiedliche Aspekte deklarativen Wissens messen“ und sich „vor allem strukturelle Aspekte des deklarativen Wissens mit Concept Maps gut gemessen werden können“. Um ihr prozedurales und strategisches Wissen zu zeigen, mussten die Schüler eigenständig ein Experiment durchführen, um eine Hypothese zu überprüfen oder ein Problem zu lösen. Zur Messung der Schülerleistung gab es verschiedene Verfahren: Beobachtungen des Schülers, Auswertung der Schülerprotokolle, Durchführung von computersimulierten Untersuchungen und Paper-Pencil Tests, in denen die Lerner erläuterten, welche Probleme es bei der Planung, Durchführung und Interpretation des Experimentes gab. Ergebnis war, dass die gerade beschriebenen ausführungsbezogenen Leistungstests wichtige Aspekte des prozeduralen Wissens verhältnismäßig gut messen können, zur „Messung struktureller Aspekte prozeduralen und strategischen Wissens aber noch kaum Forschungsergebnisse vorliegen“.

Bei Anderson et al. (2007) sollte das Schülerwissen innerhalb des Unterrichtsgespräches durch Fragen der Lehrkraft und die sorgfältige Untersuchung der Schülerantwort herausgearbeitet werden. Dabei stützen sich Anderson et al. (2007) auf die zunehmende Forderung, dass im naturwissenschaftlichen Unterricht nicht mehr nur über Naturwissenschaft gelesen und geschrieben, sondern auch vermehrt darüber gesprochen werden und dadurch gelernt werden soll. Gespräche sind dynamischer, als vorgegebene Texte und fordern die Lerner dazu auf sinnvoll zu argumentieren und zu reflektieren. Außerdem findet eine bessere Vernetzung des Wissens und eine Art situatives Lernen statt. In der Studie wurden kleinere Schülergruppen von 3-4 Schülern gefilmt (in jedem der beiden Implementationszyklen), während sie sich über Inhalte aus dem Astronomieunterricht austauschten. Gesprächsanlass waren auch die Fragen aus den durchgeführten Tests. Dabei gingen Anderson et al. (2007) davon aus, dass sich die Gespräche der Schüler durch die Aktivitäten des formative assessment verändern und durch das Coaching und die Übung verfeinern ließen. Außerdem würden durch die Interaktion mit anderen Lernern, neue Denkmodelle und veränderte Sichtweisen in die Problematik mit einfließen. Bei der Auswertung wurden sowohl verbale (Gesprächsinhalte), als auch nonverbale (Gestik und Mimik) Aspekte berücksichtig und auch festgehalten, zwischen welchen Schülern die Interaktionen stattfanden. Sie fanden heraus, dass durch die Gruppendynamik und die Förderung durch die Lehrkraft, die wissenschaftlichen Gespräche differenzierter und reflektierter wurden.

Mit Hilfe von "Concepttests" (McConnell et al. 2006; Donovan 2008) soll durch die sorgfältige Ausarbeitung von Multiple Choice Aufgaben herausgefunden werden, was die Schüler verstanden und wo sie noch Fehlvorstellungen haben. Dabei konzipierten McConnell et al. (2006) Concepttests, die aus Multiple Choice Fragen höherer Ordnung bestehen und sich auf die Schlüsselkonzepte der Lernziele beziehen. Für die Studie wurden über 300 Concepttests im Bereich der Geowissenschaften entwickelt. Es zeigte sich, dass die Arbeit mit den Concepttests zu einer höheren Schülerzufriedenheit und besseren Leistungen führte. Durch die große Datenbank an ausgearbeiteten Fragen, empfanden die Lehrkräfte die Einbindung von Concepttests in ihren Unterricht als unkompliziert. Sie benötigten nur 30 Minuten Vorbereitungszeit für ihre Klasse.

In einer qualitativen Studie von Furtak und Ruiz-Primo (2008) fand man heraus, dass vor allem offene Diagnoseaufgaben zu ausführlicheren Schülerantworten führten, die einen besseren Einblick in den Denkprozess der Lerner erlaubte und deren Verständnis naturwissenschaftlicher Konzepte (hier: Schwimmen und Sinken) detaillierter zeigte.

Ein großer Anteil der Forschung untersucht allerdings vor allem die Reliabilität und Validität von *formative assessment* Methoden im naturwissenschaftlichen Unterricht. Nur wenige Studien beschäftigen sich mit den Effekten formativer Leistungsmessung auf den Lernprozess.

Da in der vorliegenden Studie elektronische formative Leistungstests zum Einsatz kommen, soll im folgenden Abschnitt kurz auf ein paar Studien eingegangen werden, die sich mit dem Einsatz computergestützter formativer Diagnosesysteme im naturwissenschaftlichen Unterricht beschäftigen:

Trumpower & Sarwar (2010) setzten in ihrer Studie die Conceptmapping - Software „Pathfinder" ein, um das konzeptuelle Wissen zu physikalischen Zusammenhängen bei 24 high school Schülern zu überprüfen. Diese setzten sich mit den physikalischen Themen „Arbeit, Energie und Kraft" auseinander und mussten mit Hilfe der Software Relationen zwischen den verschiedenen Konzepten herstellen. „Aus diesen Einzelbewertungen wurde eine graphische Darstellung der konzeptuellen Repräsentationen errechnet und mit einer Standardlösung verglichen" (Maier 2014). Als Feedback erhielten die Schüler also zusätzlich zu ihrem Pathfinder Ergebnis auch ein Referenzergebnis, welches sie mit ihrem eigenen vergleichen konnten. Dadurch konnten sie ihr

Ergebnis noch einmal überdenken und eventuell bereits zu diesem Zeitpunkt einige Denkfehler berichtigen. Mit Hilfe von Pathfinder wurden die Unterschiede / Fehlkonzepte isoliert, um daraus individuelle Übungsaufgaben und Fördermaßnahmen zu entwickeln, die diese Fehlkonzepte ausmerzen sollten. Es stellte sich heraus, dass Schüler, die eine Kombination aus Diagnose und Förderung erhalten am besten abscheiden, es also viel größere Effekte gab, wenn die Schüler nicht nur Feedback bekamen, sondern auch anschließende Fördermaßnahmen erhielten.
Ein eher komplexes, webbasiertes System, das an den Physiklehrplan angelehnt ist und sich mit dem Abprüfen konzeptuellen Wissens aus der Physik auseinandersetzt, ist DIAGNOSER (Thissen-Roe et al. 2004). Darin sind vor allem Fragen (Testitems) zu physikalischen Schlüsselkonzepten (Kraft, Energie, etc.) enthalten. Multiple Choice Distraktoren können Fehlvorstellungen innerhalb dieser Konzepte bei den Schülern aufdecken. Dafür müssen die Lerner zunächst eine Frage beantworten und in einem zweiten Schritt eine gesonderte Erklärung (abstraktere Ebene) für ihre gewählte Antwort geben. Diese beiden Schritte sorgen dafür, dass die Schüler nicht nur raten und geben Aufschluss über das, was sich die Lerner bei dieser Antwort gedacht haben. Es stellte sich heraus, dass Schüler, die mit DIAGNOSER arbeiteten im landesweiten naturwissenschaftlichen Test besser abschnitten, als Schüler, die DIAGNOSER nicht zur Verfügung hatten.

Wang et al. (2006) und Wang (2011) entwickelten unterschiedliche *formative assessment* Strategien, um das selbstregulierte Lernen von Schülern einer *junior high school* in Taiwan innerhalb einer eLearning Umgebung zu fördern und zu unterstützen. Diese Strategien sind Teil der eLearning Umgebung und bieten interaktive Feedbackformen mit anderen Schülergruppen und Lehrkräften (zum Beispiel können die Schüler Fragen stellen, etc.) und auch aufgabenbezogenes Feedback (z.B. kann man Testitems so lange wiederholen, bis man sie richtig beantwortet hat) an.
Wang (2011) beschreibt in seinem Artikel, welche 5 Hauptstrategien das webbasierte System PDA-WATA beinhaltet:

1) Adding answer notes: Dabei handelt es sich um eine kognitive Strategie, die es den Lernern erlaubt, ihre Antworten selbst zu beurteilen und zu begründen. Während dieses Vorgangs können die Schüler ihre Antwort noch einmal reflektieren und strukturieren. Auch eine wiederholte Beschäftigung mit dem bisher gesichteten

Lernmaterial wird dadurch gefördert. Alle nachfolgenden Strategien erfolgen auf einer metakognitiven und selbstregulierenden Ebene:

2) Stating confidence: Hier sollen die Schüler ihr eigenes Vertrauen in die gegebenen Antworten und die dazugehörigen Begründungen und Notizen beurteilen, also darüber reflektieren, wie sicher sie sich bei der Beantwortung waren.
3) Reading peer answer notes
4) Recommending peer answer notes und
5) Querying peers' recommendation on personal answer notes (S. 1801). Bei den Strategien 3) bis 5) können die Lerner ihre eigenen Antworten mit denen der anderen Schüler / Schülergruppen vergleichen. Dabei sehen sie auch, wie gut die eigenen Antworten und Ausführungen sind und wie weit man im Lernprozess vorangekommen ist. Außerdem kann die Beschäftigung und Auseinandersetzung mit den anderen Schülerantworten dazu führen, die eigenen Notizen nochmals zu überarbeiten und neu zu strukturieren. Durch die gegenseitige Bewertung und Einschätzung erhält man ein individuelles Feedback. Dabei wirken sich eine positive Beurteilung und die Wertschätzung anderer auch positiv auf die Motivation der Lerner aus.

Die Schüler der Kontrollgruppe hatten die gleichen eLearning Tools zur Verfügung und arbeiten die gleiche Zeit an den Lerninhalten. Außerdem bearbeiteten sie die gleichen formativen Aufgaben, bekamen aber nur ein einfaches dichotomes Feedback, ob ihre Antwort richtig oder falsch war. Die Schüler wurden zufällig einer der beiden Gruppen zugeordnet und mussten nach einer zweiwöchigen eLearning Phase einen Posttest schreiben. Dabei fand man folgendes heraus: Die Treatmentgruppe hatte signifikant bessere Ergebnisse im summativen Posttest und im Motivationstest. Die Schüler dieser Gruppe waren offener für die Angebote der und die Arbeit mit der eLearning Plattform und versuchten stärker an der Verbesserung ihrer Lernleistung und an selbstregulierten Lernstrategien zu arbeiten. Außerdem zeigte sich, dass es innerhalb der Treatmentgruppe keine signifikanten Unterscheide in der Lernleistung gab, egal, ob die Schüler über ein hohes oder niedriges Level selbstregulierter Lernprozesse verfügten. In der Kontrollgruppe war dies nicht der Fall. Diese Ergebnisse wurden mit Theorien zum selbstregulierendem Lernen begründet: Unterstützt man selbstregulierte Lernprozesse in der eLearning Umgebung führt dies zu einer höheren Qualität der selbstregulierten Lernaktivitäten und zu höherer Effek-

tivität. Die Bereitstellung verschiedener Feedbackstrategien führt also zu besseren Ergebnissen. Außerdem führt die Bereitstellung verschiedener Strategien (wie es in der Treatmentgruppe der Fall ist) und somit ein erweiterter Zugang zum Computer zu einem immer versierteren Umgang mit der Mensch – Maschine Schnittstelle, was sich neben der Möglichkeit sich gegenseitig zu helfen und zu unterstützen, ebenfalls positiv auf die Schülermotivation auswirkt (Wang 2008). An der Studie nahmen 123 Schüler der siebten Jahrgangsstufe aus 4 verschiedenen Klassen teil.

Hickey und Zuiker (2012) untersuchten die Effekte des interaktiven eLearning Programmes GenScope auf die Schülermotivation und Schülerleistung. In einem virtuellen Modell (fiktive Drachenpopulation) konnten Studenten Schlüsselkonzepte der Genetik genauer unter die Lupe nehmen, indem sie unterschiedliche Variablen veränderten (Spezies, Vererbungsregeln, genetische Zusammensetzung, etc.). Durch die Simulation dieser Veränderungen und deren Auswirkungen, bekamen die Schüler ein direktes Feedback. Das Programm hält verschiedene Rückmeldeformate für die Lerner bereit: informelles bzw. formelles Feedback und sofortiges Feedback bzw. verzögerte Rückmeldungen. Es stellte sich heraus, dass die Schüler, die mit der Software arbeiteten zunächst keine besseren Ergebnisse im summativen Test erzielten, als Lerner aus anderen Klassen. Erste Vorteile der eLearning Umgebung und er ausgearbeiteten Strategien zeigten sich erst 3 Jahre nach der Implementation: Es gab zum Teil sehr hohe Effekte bei den Schülerleistungen und der Problemlösefähigkeit (allerdings nicht bei allen Lehrkräften).

Zusammenfassend kann man festhalten, dass Feedback besonders dann erfolgreich und effektiv ist, wenn es die Aufmerksamkeit, auf eine der höheren Eben der Feedbackinterventionstheorie lenken kann, also weg von der Ebene der reinen Aufgabenlösung. Dabei kann Lob oder Kritik, zu geringen oder sogar negativen Effekten führen, wenn sie bei Rückmeldungen zu komplexen Aufgaben zum Einsatz kommen, denn dann wird der Lerner durch die Beschäftigung mit sich selbst zu sehr von der Aufgabenstellung und deren Anspruch abgelenkt. Positive Effekte erhält man in diesem Zusammenhang nur dann, wenn sich das Feedback lobend auf einfache Aufgaben bezieht und den Lerner dadurch dazu bringt, sich mehr anzustrengen. Betrachtet man die Aufgabenkomplexität genauer, kann man festhalten, dass Rückmeldungen zu komplexeren Aufgaben dann effektiver sind, wenn sie ausführlicher ausfallen und die Lerninhalte und

Zusammenhänge noch einmal genau beschreiben (elaboriertes Feedback). Die Leistungsdiagnose selbst sollte, wenn sie zu einer effektiven Rückmeldung für die Lernenden führen soll, möglichst informativ sein, also etwas über das Lernergebnis selbst aussagen und Einblicke in die erlernten Vernetzungen der fachlichen Zusammenhänge gewähren. Einfachere Sachverhalte sollten durch häufigere kurze Rückmeldungen in ihrer Richtigkeit bestätigt oder negiert werden. Dabei ist es effektiver, das Feedback möglichst zeitnah zur Verfügung zu stellen, weil die Schüler schon während sie das Feedback erhalten, aus ihren Fehlern lernen können. Das richtige Ergebnis (knowledge of results) zur Verfügung zu stellen wirkt sich ebenfalls positiv aus. Ein detailliertes und ausschweifendes Feedback hingegen wäre in diesem Fall eher hinderlich und würde den Schüler vom eigentlichen Lerninhalt ablenken. Bespricht die Lehrkraft mit den Schülern schon zu Beginn der Unterrichtseinheit, wie die Lernziele aussehen und welche Beurteilungskriterien bei der Auswertung der Tests herangezogen werden, wirkt sich das positiv aus die Schülermotivation und die Lernleistung aus.
Weiterhin hat man festgestellt, dass „personalisierte Rückmeldungen Nachteile im Vergleich zu schriftlichen Rückmeldungen oder Rückmeldungen durch den Computer mit sich bringen. Der Lerner fühlt sich weniger normativ evaluiert und kann sich besser auf die Details der Rückmeldung beziehen.“ (Maier, Hofmann, Zeitler 2012)

Um einen besseren Überblick über die Ergebnisse der Feedbackforschung zu gewährleisten, wurden die wichtigsten Resultate in einer Tabelle zusammengefasst:

Tabelle 3: Ergebnisse der Feedbackforschung

Untersuchtes Feedbackmerkmal		**Ergebnis**	**Quelle / Literatur**
Feedback auf der Aufgabenebene	Komplexität der Aufgabe	Hohe Effektstärken bei einfachen Aufgaben.	Bangert-Drowns et al. (1991)
	Zeitlicher Aspekt	Dieser Feedbacktyp ist umso effektiver, je schneller die Rückmeldung erfolgt.	Feedbackinterventionstheorie (FIT) nach Kluger & DeNisi (1996)
	Formulierung der Rückmeldung (positiv / negativ / neutral)	Positives Feedback zur Aufgabenlösung ist hoch effektiv, negativ formuliertes Feedback ist nur dann erfolgreich, wenn Korrekturhinweise (richtiges Ergebnis) gegeben werden	

Untersuchtes Feedbackmerkmal		Ergebnis	Quelle / Literatur
Feedback zum Aufgabenlösungsprozess FIT		Bringt die Schüler dazu, ihre eigenen Lösungswege kritisch zu hinterfragen und führt zu einem tieferen Verständnis der Lerninhalte.	
	Zeitlicher Aspekt	Sollte zeitverzögert erfolgen	
Feedback zur Selbstregulation FIT	Zur Änderung von Lernstrategien und zur Hinführung zu selbstgesteuertem Lernen	Änderungen von Einstellungen und Verhaltensweisen der Schüler und die daraus resultierenden Entwicklungen hin zum selbstreguliertem Lernen sind erst nach einem langen Zeitraum feststellbar.	
Feedback an die Schülerpersönlichkeit (Ebene der selbstbezogenen Kognitionen) FIT	Lob, Kritik oder normative Hinweise	Geringere Leistungssteigerung, da die Aufmerksamkeit des Schülers von der Aufgabe auf sich selbst gelenkt wird (vor allem bei komplexen Aufgaben reduzierte Effekte); bei einfachen Aufgaben führt Lob zu größerer Anstrengung und einer Leistungssteigerung	
	Elektronisches Feedback	Computerbasierte Rückmeldungen sind effektiver als verbale Rückmeldungen, denn der Lerner fühlt sich weniger normativ beurteilt und kann sich besser auf die Details der Rückmeldung konzentrieren.	
Lehrerfeedback an die Schüler hinsichtlich	der Bezugsnorm (BN)	Höhere Leistungsmotivation bei: o Kriterialer BN: Feedback, inwieweit die Leistung des Schülers den Anforderungen der Aufgabe entsprochen hat o Individueller BN: Vergleich der Schülerleistung mit seinen vorherigen Leistungen o Lernzielorientierter BN: Besprechung der Lernziele und der Bewertungskriterien bei der Beurteilung der Tests	Dweck (1986), Butler (2005)
	der erreichten Lernziele	Das Feedback sollte Informationen enthalten, wie weit der Lerner genau vom festgesetzten Lernziel entfernt ist, also an welcher Stelle der Schüler innerhalb seines Lernprozesses steht. Dies führt zu größeren Anstrengungen.	Heritage (2007) Black & Wiliam (2009)

Untersuchtes Feedbackmerkmal		Ergebnis	Quelle / Literatur
	ihrer Qualität	Die Qualität des Feedbacks in Hinsicht auf seine Akkuratheit, Angemessenheit und Verständlichkeit, ist neben einer verständlichen Vermittlung essentiell für eine positive Unterstützung des Lernprozesses (in: Bürgermeister 2014).	Sadler (1998)
	ihres Informationsgehalts	Meist keine Effekte bei einfachem Feedback, wo nur angezeigt wurde, ob richtig geantwortet wurde. Höhere Effekte gibt es, wenn die richtige Antwort rückgemeldet wurde und wenn elaboriertes Feedback gegeben wurde (Weiterführende Informationen zur Lösung oder zum Lösungsweg)	Bangert-Drowns et al. (1991)
Die Feedbacknutzung		Ist besonders effektiv, wenn es dazu herangezogen wird, um den nachfolgenden Unterricht adaptiv zu gestalten, die Unterrichtsprozesse kontinuierlich an die Entwicklungen der Lerner anzupassen und die Lernziele zu überdenken und gegebenenfalls neu zu formulieren.	Harlen & Winter (2004) Maier (2010) Smit (2009) Hattie (2009)
Feedback und Schülermotivation		Informatives und positives, bewertendes Feedback korreliert positiv mit der Schülermotivation. Negatives bewertendes Feedback hängt hingegen nicht mit motivationalen Effekten zusammen.	Radoczy et al. (2008)
Feedback in Kombination mit	Fördermaßnahmen	hat einen höheren Effekt auf die Schülerleistung, als reines Feedback.	Trumpower & Sarwar (2010)
	gezielten Unterrichtsgesprächen und Coaching durch die Lehrkraft	führt durch die Interaktion mit anderen und deren Sichtweisen zu reflektierten Ansichten und vertieften konzeptuellen wissenschaftlichen Verständnis.	Anderson et al. (2007)
Peer Feedback		Erhält man ein positives Feedback von seinen Mitschülern auf seine Antwort und wird diese dadurch wertgeschätzt, hat dies einen hohen Effekt auf die Motivation.	Wang (2011)

Zusätzlich können auch Persönlichkeitsmerkmale einen Einfluss auf die Effektivität der Rückmeldung haben. Kluger & DeNisi (1996) beschreiben in ihrer Feedbackinterventionstheorie, dass auch die Merkmale von Lernenden einen großen Einfluss auf Feedbackeffekte haben, die-

se Einschätzung spiegelt sich auch in folgendem Zitat von Maier, Hofmann und Zeitler (2012) wieder: „Außerdem hängt die Wirkung der Rückmeldung auch von Persönlichkeitsmerkmalen ab (z.B. Selbstwertgefühl, Kontrollüberzeugungen oder Leistungsorientierung). Beispielsweise tendieren Personen mit geringem Selbstwertgefühl und hoher Ängstlichkeit dazu, negatives Feedback zu vermeiden."

Und neben den Persönlichkeitsmerkmalen kann es weitere „äußere Faktoren" geben, die die Effektstärken beeinflussen. So beschreiben Kingston & Nash (2011) in ihrer Metaanalyse, dass die Moderatorenanalyse gezeigt hat, dass das Unterrichtsfach den größten Einfluss auf die Höhe der Effektstärke hat: Studien im Fach Englisch kommen durchschnittlich zu höheren Effekten, als Studien in Mathematik oder gar naturwissenschaftlichen Fächern – dort sind die Effekte besonders gering.

Einen umfassenden internationalen narrativen Forschungsüberblick zu 37 empirischen Studien aus verschiedensten Fächergruppen, die sich „mit der Entwicklung, der Implementation und den Effekten elektronischer formativer Leistungstests in der Schule beschäftigen", gibt Maier (2014) in seiner Publikation „Computergestützte, formative Leistungsdiagnostik in Primar- und Sekundarstufe".

Um die elektronische Unterstützung beim Einsatz formativer Diagnosesysteme geht es auch im folgenden Kapitel.

2.2.2.1 Feedback in elektronischen formativen Leistungstests

Durch die zunehmende Technisierung in allen Sparten des Lebens (Arbeitswelt, Freizeit, etc.), werden auch die immer wiederkehrenden Forderungen, digitale Medien im Unterricht einzusetzen, legitim. Leider sieht die Realität, wenn man den Einsatz elektronischer Medien in der Bildung näher betrachtet eher gegenteilig aus. Ludwig (2013, S. 24) zitiert dazu: „Beinahe jede Neuerung auf dem Technikmarkt der letzten Jahre wird auf Unterrichtstauglichkeit getestet. So kamen die „Schulen ans Netz", wurden „1001 Notebooks für den Schulranzen" angeschafft, die „digitale Schultasche" versprochen und das „Ende der Kreidezeit" heraufbeschworen. Zum Standard im Unterricht gehören heute jedoch weder USB-Sticks statt Büchern und Heften noch ein mobiler Computer oder digitale Schultafeln". So gibt es sowohl bei Schülern, die Computer und Internet in der häuslichen Vorbereitung und beim Lernen sehr häufig einsetzen und Leh-

rern, die den Computer regelmäßig für die Unterrichtsvorbereitung nutzen, eine gravierende Diskrepanz zwischen häuslicher Ausstattung und institutioneller Nutzung (Ludwig 2013).
Dabei wären die Einsatzmöglichkeiten elektronischer Materialien und Geräte im Unterricht sehr vielfältig.

Die Realität zeigt jedoch (auch nach meinen Erfahrungen im Schuldienst), dass elektronische Medien bisher nur in Form von Desktop PCs an den Schulen angekommen sind und diese zusätzlich kaum ins Unterrichtsgeschehen integriert werden. So gibt es überwiegend ausschließlich Computerräume, die mit einem Klassensatz PCs ausgestattet wurden.

Diese Rechnerräume werden aber aus verschiedenen Gründen von den Lehrkräften selten bis gar nicht genutzt:

1) Der Wechsel vom normalen Klassenzimmer in den Computerraum ist mit Zeitverlust und häufig unruhigem Verhalten der Schüler verbunden (Aufenanger & Schlieszeit 2013). Oft lohnt es sich nicht, wegen einer kurzen Recherche oder einem kleinen Arbeitsauftrag das Zimmer zu wechseln. Außerdem sind die Schüler in der gewohnten Umgebung des Klassenzimmers meist aufmerksamer und ruhiger.
2) Viele Lehrkräfte scheuen davor zurück, sich mit der im Computerraum bereitstehenden Technik auseinanderzusetzen und haben zusätzlich Angst zu versagen: Wie funktioniert das Smartboard, wie kann ich die Arbeitsplätze der Schüler freischalten, oder einzelne Schülerergebnisse mit dem Beamer anzeigen lassen? Funktioniert das mit dem einloggen überhaupt und blamiere ich mich nicht vor den Schülern, wenn ich mich mit den technischen Gegebenheiten nicht auskenne?
3) Oft sind die Computerräume durch den Informatikunterricht besetzt und nicht frei, wenn man ihn gerade bräuchte (Aufenanger & Schlieszeit 2013).
4) Schüler versuchen die Gelegenheit zu nutzen und sich am Computer mit fachfremden Dingen zu beschäftigen.
5) Das Starten der Computer und das Einloggen benötigt einiges an Zeit und funktioniert nicht immer reibungslos.
6) …

Meine Erfahrungen als Mathematiklehrerin am Gymnasium haben gezeigt, dass häufig nur die Informatiklehrer und mit etwas Glück der eine oder andere Mathematiklehrer, den vorhanden Computerraum auch wirklich nutzen.

Ludwig (2013, S. 25), hat in einer Lehrerbefragung (Erhebung: *N* = 13, Notenskala 1-9) herausgefunden, wo die Hinderungsgründe beim Einsatz digitaler Medien im Allgemeinen liegen und diese in einem Ranking zusammengestellt (siehe Abbildung 2). Auch hier zeigt sich, dass die ungenügende Geräteausstattung der Schulen, der hohe Zeitbedarf bei der Integration in den Unterricht und Organisationsprobleme bei der Raumbelegung Gründe sind, den Computer nicht zu nutzen. Diese eher infrastrukturellen Faktoren könnten mit dem Einsatz von Tablet PCs und Faktoren, wie die fehlende Medienkompetenz auf Lehrerseite könnten durch geeignete Fortbildungsangebote eliminiert werden.

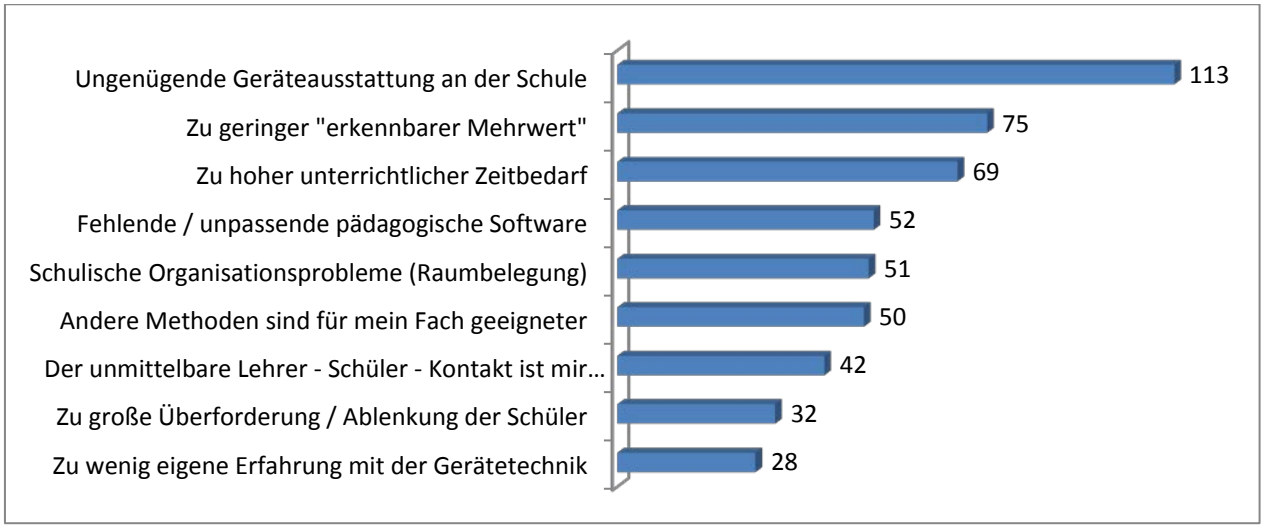


Abbildung 2: Hinderungsgründe beim Einsatz digitaler Medien im Unterricht (Ludwig (2013)

Das Raumproblem und den Zeitverlust durch den Raumwechsel hat man versucht, durch sogenannte Medienecken im Klassenzimmer zu lösen: An manchen Schulen gibt es deshalb heute in jedem Klassenzimmer ein oder zwei PCs, die aber natürlich bei weitem nicht ausreichen, damit alle Lerner der Klasse gleichzeitig an einer Aufgabe arbeiten können. „Die Einrichtung von Medienecken im Klassenraum hat die Abhängigkeit von Zeiten und Raumnutzung des Computerraums gelöst, zugleich aber die Verwendung des Computers dahingehend eingeschränkt, dass nur wenige Computer nutzbar waren“ (Aufenanger & Schlieszeit 2013, S.8). In Medienecken

sind Computerarbeitsplätze also jederzeit verfügbar, werden aber dann aus anderen Gründen letztendlich doch äußerst selten genutzt.

Noch seltener kommt es vor, das Netbooks (in sogenannten Laptopklassen) oder gar Tablets in den Unterricht integriert werden. Nur sehr vereinzelt findet man Projektschulen, in denen es Klassen gibt, die flächendeckend mit Tablets arbeiten. Dabei könnten gerade diese kleinen, mobilen Endgeräte relativ schnell und unkompliziert in den Unterrichtsalltag integriert werden und würden viele Vorteile (auch der Mediennutzung im Allgemeinen) mit sich bringen:

1) Elektronische Unterrichtsmaterialien oder Schulbücher können sehr viel anschaulicher gestaltet werden, als ein einfaches Arbeitsblatt oder traditionelle Schulbücher. So besteht zum Beispiel die Möglichkeit, in Geschichte einen kurzen Film zu einer historischen Begebenheit einzuspielen, in Physik eine Animation zu einem Experiment zur Verfügung zu stellen (Aufenanger & Schlieszeit 2013), in dem die Schüler einzelne Variablen ändern und abschließend beobachten können, inwieweit sich die restlichen Parameter im Versuchsaufbau verändern, in Musik Tonbeispiele einzubinden, oder auch in Englisch auf Knopfdruck, Texte oder Wörter in der richtigen Aussprache lesen zu lassen. Dies sind nur einige wenige Beispiele dafür, wie man den Lernprozess anschaulicher und motivierender/interessanter gestalten kann. Durch den schnellen Einsatz (kein Zeitverlust beim Startvorgang), können die Tablets jederzeit als interaktives Buch dienen.
2) Über eine Cloud (Datenwolke) können alle Schüler und Lehrer einer Klasse gemeinsam an Dokumenten arbeiten. Ein gemeinsamer Server, der sich durch Datenverschlüsselung hinreichend schützen lässt, dient als eine Art „Dropbox“, in der die Dokumente gespeichert und bearbeitet werden können (Schlieszeit 2013).
3) Leistungstests, Rätsel und Übungsaufgaben können direkt in das elektronische Schulbuch oder Arbeitsblatt eingebunden und abwechslungsreich gestaltet werden.
4) Eine automatische Auswertung / Korrektur erleichtert den Lehrkräften die Arbeit.
5) Es gibt bereits eine große Anzahl an Apps, die in der Bildung eingesetzt werden können. Wie geeignet die unterschiedlichen Angebote für den eigenen Unterricht sind sollte man vorher genau überprüfen (Schlieszeit 2013).

6) Die Schüler können selbst aktiv werden: Im Internet recherchieren, Filme drehen, Dialoge im Fremdsprachenunterricht aufnehmen, Fotos für eine Kunstaustellung aufnehmen, etc.
7) Schülervorträge können ebenfalls in elektronischer Form gestaltet und gehalten werden. Über WLAN und Beamer kann man per Knopfdruck (z.B. via AirPlay beim iPad) die Inhalte vom Tablet „als Großprojektion für alle sichtbar machen und in der Anwendung arbeiten“ (Schlieszeit 2013, S. 10). Es besteht also die Möglichkeit, die Tabletinhalte aller Schüler und Lehrer über den Beamer zu zeigen und das von jedem Platz des Klassenzimmers aus. Die Fähigkeiten, Inhalte anschaulich aufzubereiten und diese in einem Vortrag ansprechend zu präsentieren, benötigen die Kinder heute in sehr vielen Berufen, so dass der Umgang mit den technischen Medien sehr wichtig geworden ist.
8) Adaptive Tests sind eigentlich nur elektronisch durchführbar, denn adaptive Diagnosesysteme benötigen eine automatisierte Auswertung, einen systematischen Abgleich mit Teilschritten des Problemlöseprozesses der Aufgabe und ein unmittelbares Feedback an den Schüler. Durch adaptive Systeme können gezielt Denkfehler auf Lernerseite diagnostiziert und isoliert werden. In einem zweiten Schritt besteht die Möglichkeit, dem Schüler genau die Übungsaufgaben oder Lernstrategien an die Hand zu geben, die gezielt das isolierte Problem angehen (z.B. Lehker 2009).
9) Durch die Mobilität (geringe Größe und geringes Gewicht, Unabhängigkeit vom Stromnetz), können Tablets überall mit hingenommen werden. „Unterricht wird dadurch realitätsnah und öffnet sich gegenüber neuen Lernwelten“ (Baumgartner et al. 2011). Dadurch besteht auch die Möglichkeit, elektronische Tests zu jeder Zeit und an jedem Ort (auch zu Hause) durchzuführen.
10) Netbooks und auch Tablets „unterstützen Gruppenprozesse, indem sie den Schülern wichtige Kollaborations- und Kommunikationswerkzeuge (Kamera, Mikrofon, Web – Tools) in die Hand geben“ (Baumgartner et al. 2011). Durch die vielen Möglichkeiten, sich mit anderen Schülern und den Lehrkräften auszutauschen und mit diesen zu interagieren, werden kollaborative Lernprozesse unterstützt. Roschelle et al. (2010) fanden in einer Studie heraus, dass kollaboratives Lernen mit Hilfe von Tablets einen positiven Effekt auf den Wissenszuwachs der Schüler hat: Mit Hilfe des serverbasierten Systems TechPALS wurden kooperative Lernprozesse für eine Unterrichtseinheit in Mathematik

bereitgestellt. Die quasiexperimentelle Studie fand an drei amerikanischen Grundschulen im Schuljahr 2007 – 2008 statt. Die Treatmentgruppe arbeitete in kleinen Gruppen mit dem Technology-mediated, Peer-Assisted Learning System (TechPALS) auf Tablet PCs, um Bruchrechenprobleme zu lösen. Dabei hatten sie die Möglichkeit Fragen zu stellen, Begründungen zu formulieren und unterschiedliche Sichtweisen auf das bestehende Problem zu diskutieren. Einzelne Schüler erhielten also auch ein Feedback von der Gruppe. Die Kontrollgruppe löste in Einzelarbeit Probleme der Bruchrechnung und erhielt ein individuelles Feedback zu ihrer Arbeit. Es zeigte sich, dass die Schüler, die mit dem TechPALS System arbeiteten, mehr lernten, als die Schüler der Kontrollgruppe. Außerdem stellte man durch Beobachtungen fest, dass die Schüler der Treatmentgruppe verstärkt an den durch die Technologie bereitgestellten sozialen Lerngefügen (Fragen, Erklären und Diskutieren) teilnahmen.

11) Durch Mindmapping-Tools können Lernprozesse und Gedanken besser organisiert und strukturiert werden.
12) …

All diese Möglichkeiten könnten mit Tablet PCs realisiert werden, denn sie sind Fotoapparat, Computer, Filmkamera, Fernseher und Suchmaschine in einem. Es gibt sie in unterschiedlichen Größen und Preisklassen und sie können überall mit hingenommen werden.
Auch die Akzeptanz dieser Geräte bei Schülern und Studenten ist sehr hoch. An den Hochschulen werden bereits zunehmend iPads eingesetzt. „Ergebnisse einer ersten Befragung an der Universität Kassel zeigen, dass Studierende, die ihre Texte und Materialien über das iPad bekommen, mehr und intensiver lesen bzw. nutzen als Studierende ohne dieses Gerät“ (Aufenanger & Schlieszeit, 2013, S. 9).

Multimediale Lernumgebungen selbst regen durch ihre Beschaffenheit ebenfalls zu einem motivierten Umgang mit den Lerninhalten an. Laut Narciss (2006) weisen elektronische Lehr- und Lernumgebungen folgende Charakteristika auf:

1) Eine nicht lineare Struktur
2) Multimodalität (die Informationsdarbietung erfolgt auf mehreren sensorischen Kanälen, sowie mittels unterschiedlicher semantischer Codes)

3) Interaktivität (Lernende entscheiden selbst, welche Inhalte, sie wann bearbeiten möchten und können sich mit anderen Schülern elektronisch austauschen)

Durch diese Charakteristika wird eine Individualisierung der Lernprozesse erst ermöglicht. Vor allem durch die interaktive Struktur sollen die Lerner aktiviert werden, sich intensiv mit den Lerninhalten auseinanderzusetzen und diese nicht nur flüchtig zu konsumieren.

Elektronische Unterrichtsmaterialien, Leistungstests, oder eLearning Kurse sind im schulischen Kontext allerdings noch nicht sehr weit verbreitet - eLearning Plattformen werden vor allem im Hochschulbereich verwendet – dort werden bereits häufiger Onlinekurse und auch elektronische Tests angeboten und durchgeführt.

Das Defizit von elektronischen Medien in der Schule, spiegelt sich auch im Angebot der Schulbuchverlage wieder: Durch den geringen Absatzmarkt, wiederum hervorgerufen, durch die geringe Verbreitung von Lesegeräten (Tablets), führt dazu, dass es nur vereinzelt elektronische Schulbücher gibt und diese zusätzlich nur die digitale Version der traditionellen Schulbuchausgabe sind und dadurch kaum zusätzliche Funktionen haben (Schlieszeit 2013, Ludwig 2013, Ott 2013). Es besteht gerade mal die Möglichkeit ein Lesezeichen oder eine Notiz einzufügen. Der Mehrwert, den das elektronische Medium mit sich bringen könnte, wird bei weitem nicht ausgeschöpft. Schlieszeit (2013, S. 12) schreibt zu dieser Problematik. „ Zwar haben sich einige Bildungsverlage zusammengeschlossen und als Pedant zum eBook das „digitale Schulbuch“ ins Leben gerufen. Allerdings handelt es sich hier bisher lediglich um bessere digitalisierte Bücher, ähnlich einem PDF-Dokument mit Blätter- und Markierfunktion ohne echte interaktive Funktionen“.
Die Vorteile dieser klassischen eBooks gegenüber der gedruckten Version liegen laut Ott (2013) bei folgenden Faktoren:

1) Die Bestellung und Auslieferung von E-Lehrmitteln ist für Schulen bedeutend einfacher und kostengünstiger als Print- Lehrmittel.
2) E-Books kosten weniger, da sie kein Papier benötigen, nicht gelagert und nicht geliefert werden müssen.

3) Auf einem einzigen Lesegerät lassen sich hunderte von Büchern speichern. Die Schüler müssen dadurch deutlich weniger Gewicht mit sich herumschleppen.
4) Die Schriftgröße kann jederzeit individuell angepasst werden.
5) Man kann innerhalb der Bücher die komfortable Suchfunktion nutzen, um gezielt bestimmte Begriffen ausfindig zu machen.
6) Digitale Bücher können schneller überarbeitet und auf den neuesten Stand gebracht werden. Danach braucht es keinen neuerlichen Druck – ein kurzer Upload der neuesten Version per Knopfdruck genügt.
7) Multimediale und interaktive Inhalte könnten eingebunden werden.

Ott (2013, S. 46) schreibt: „ Das Angebot an multimedialen Schulbüchern ist heute noch sehr klein, denn die Entwicklung von multimedialen und interaktiven digitalen Büchern ist für die Lehrmittelverlage momentan deutlich aufwendiger und kostspieliger als die Herstellung herkömmlicher Printlehrmittel.“ Hinzu kommt der bereits erwähnte geringe Absatzmarkt. Dennoch denkt Ott (2013), dass „der Trend zu kleinen, multifunktionalen und leistungsfähigen Geräten, den digitalen Schulbüchern zum Durchbruch verhelfen dürfte“ (S. 46).

Im Moment unterscheidet man zwischen drei Arten von digitalen Schulbüchern:
Dem bereits erwähnten klassischen eBook, dem Enhanced Book und dem Flexbook.
In seinem Artikel „Trends bei digitalen Schulbüchern“ stellt Ott (2013) diese drei Arten einander gegenüber:

Tabelle 4: Spektrum digitaler Schulbücher (aus: Ott 2013)

	Klassisches eBook	**Enhanced Book**	**Flexbook**
Eingebundene Medientypen	Text und Bild	Text, Bild, Ton und Film	Text, Bild, Ton und Film
Interaktiver Inhalt	Nein	Ja	Ja
Kollaborative Funktionen	Nein	Ja	Ja
Anpassbarer Inhalt	Nein	Nein	Ja
Funktionalität	zunehmend		
Anzahl der Bücher (2012)	abnehmend		

Enhanced Books können demnach auch multimediale, kollaborative und interaktive Funktionen enthalten und Flexbooks sogar von der Lehrkraft individuell verändert und für die jeweilige Klasse gestaltet und angereichert werden. Auf dem deutschen Markt gibt es bisher vornehmlich klassische eBooks, ganz vereinzelt Enhanced Books. Flexbooks sind ausschließlich in Englisch erhältlich. Das liegt daran, dass vor allem im amerikanischen und asiatischen Markt die Digitalisierung der Schulbücher bereits viel weiter vorangeschritten ist, als im Europäischen oder gar deutschen Raum. Laut Ott (2013) soll in den USA bereits im Jahr 2016 35% des Schulbuchumsatzes durch digitale Schulbücher gedeckt werden und Südkorea möchte schon 2015 alle Schulbücher digital anbieten.

Mit der Weiterentwicklung der digitalen Schulbücher könnte auch die Einbindung elektronischer formativer Leistungstests einhergehen. Die Vorteile computergestützter formativer Diagnoseverfahren ist ja nicht von der Hand zu weisen: schnelle und automatische Korrekturen, individuelles und direktes Feedback, Bereitstellung individueller Fördermaßnahmen und Übungsmaßnahmen, Einbindung adaptiver Systeme, etc. Dennoch gibt es sie bisher kaum und die vorhandenen Systeme wurden in der Forschung auch selten näher untersucht. Über dieses Desiderat schreibt Maier (2014, S. 69): „Digitale Informations- und Kommunikationstechnologien eröffnen neue Möglichkeiten, diagnostisches Handeln im schulischen Kontext zu gestalten... Die Nutzung von Computern, Internet und mobilen Endgeräten für eine formative Leistungsdiagnostik in Schulen ist jedoch gering.“ Und „In der pädagogischen Diagnostik sind computergestützte Verfahren nicht mehr wegzudenken (vgl. Jurecka & Hartig 2007). Verschiedene Vorteile sind damit verbunden. … Es liegt auf der Hand, dass diese Argumente auch für eine EDV-Nutzung im Rahmen der formativen Leistungsdiagnostik sprechen (vgl. Koedinger, McLaughlin & Heffernan 2010; Russell 2010).“ (S. 71)

Neben der beschriebenen geringen Anzahl interaktiver Schulbücher, gibt es in deutschen Verlagen auch kaum kommerzielle Systeme, um elektronische Leistungsmessungen durchzuführen (z.B. Antolin.de) und diese sind bisher nicht empirisch überprüft worden.
In Amerika bieten einige kommerzielle Anbieter computergestützte Systeme an, die es erlauben, Leistungsmessungen häufiger (2-5-mal pro Woche) im Mathematik- oder Deutschunterricht (Lesen) einzusetzen.

Allerdings bestehen diese Leistungsdiagnosen vor allem aus Multiple-Choice-Aufgaben (z.B. *Rapid Formative Assessment* oder *Interactive Computerised Assessment System: InCAS*). Lehrkräfte können dabei die Testaufgaben selbst zusammenstellen und damit individuelle und für die Klasse bzw. einzelne Schüler passende Leistungsdiagnosen erstellen. Studien haben gezeigt, dass Lehrer dadurch die Schülerlernvoraussetzungen stärker berücksichtigen und häufiger mit den Lernern über Lernziele und Bewertungskriterien sprechen. Dadurch, dass die Erstellung und Korrektur der Tests, sowie die Erteilung des Feedbacks automatisiert ablaufen, werden die Lehrkräfte in ihrer Arbeit unterstützt und entlastet (Yeh 2006, 2007, 2009; Merrell & Tymms 2007).

An amerikanischen Hochschulen werden auch elektronische *classroom response systems* (CRS) verwendet, um den momentanen Wissensstand des kompletten Kurses zu erfahren: Dafür stellen die Dozenten eine Frage und die von den Studierenden gewählten Antwortalternativen werden sofort auf einem Display dargestellt (Feldman & Capobianco 2008). So kann die Lehrkraft sofort sehen, ob die vorher durchgenommenen Lerninhalte verstanden wurden, oder ob es im nachfolgenden Unterricht noch Klärungsbedarf gibt. Die Technik ist einfach und flexibel einsetzbar und deshalb weit verbreitet. Allerdings ist durch das feste Aufgabenformat (Multiple Choice) eine kumulative Erfassung von Lernprozessen, vor allem, wenn es um die Überprüfung von komplexen Sachverhalten und vertieftem Verständnis geht, nicht möglich (Beatty & Gerace 2009; Maier & Randler 2011).

Doch nicht nur in Hochschulen, sondern auch in allgemeinbildenden Schulen, könnten eLearning Systeme neue Wege eröffnen, um formative Leistungsdiagnosen einzusetzen.
Russell (2010) unterscheidet vier Hauptbereiche, in denen sich Technologien, die formative Leistungsmessungen elektronisch unterstützen können, bereits sehr schnell entwickeln:

1) Online - Diagnosesysteme, die den Lern- und Entwicklungsprozess der Schüler überwachen. In diesem Bereich gibt es ein großes Entwicklungspotential, aber auch einen großen Entwicklungsbedarf. Je komplexer die Programmierung hinter dem Diagnosesystem, desto umfassender und individueller können die Lerner in ihrem Lernprozess begleitet und unterstützt werden. Das beste Diagnosesystem wäre adaptiv, so dass die Fehler der Schüler innerhalb des Problemlöseprozesses genau identifiziert und isoliert

werden könnten, um in den nachfolgenden Lerneinheiten genau auf diese Defizite mit Hilfe von speziellen Übungsaufgaben und individuell generierten Testfragen einzugehen.

2) Technologien zur Identifikation von Fehlvorstellungen der Lerner innerhalb der Lerninhalte. Vor allem im naturwissenschaftlichen Unterricht, wo es oft um das Verständnis komplexer Zusammenhänge geht, bedarf es neuer Diagnosesysteme, die falsche Vorstellungen innerhalb der Konzepte aufdecken können (vgl. Kapitel 2.2.2). Dabei könnten neue Aufgabentypen helfen, wie zum Beispiel Drag & Drop Aufgaben, bei denen eine Concept Map zu einem bestimmten naturwissenschaftlichen Zusammenhang mit den richtigen Begriffen, Bildern oder Abhängigkeiten gefüllt werden müssen.
3) Automatische Auswertesysteme der elektronischen Tests. Die Automatisierung bei der Korrekturarbeit kann die Lehrkräfte in ihrer Arbeit unterstützen und entlasten. Inzwischen gibt es sogar Bemühungen, durch automatisierte Textanalysen (Texterkennung und -analyse) auch komplexere Problemlösefähigkeiten auf Schülerseite zu überprüfen. Zusätzlich erlaubt die Speicherung jeder gegebenen Schülerantwort, auch später einzelne Entwicklungsschritte der Lerner nochmals nachzuvollziehen und deren Wissenszuwachs genau festzuhalten.
4) Oben genannte “Classroom response systems”, die sofort nach der Fragestellung der Lehrkraft, die gewählten Antworten der Studenten rückmelden, erlauben einen guten Einblick in den momentanen Wissenstand der Schüler. Gleich im Anschluss kann der Lehrer im Unterricht auf die aufgedeckten Wissenslücken auf Schülerseite nochmals genauer eingehen, also die folgende Lerneinheit individuell an die Bedürfnisse der Klasse/ des Kurses anpassen.

Dies alles sind Technologien, die nicht nur in der Hochschule, sondern auch in der Schule ohne Probleme eingesetzt werden könnten.

Die formativen Leistungstests selbst, sowie das dazugehörende Feedback können in eine Lernplattform eingebettet werden, um sie den Schülern zur Verfügung zu stellen. Dieses Vorgehen bietet sich an, weil die unterrichtende Lehrkraft über keine Programmierkenntnisse verfügen muss, sondern auf ein bestehendes System, dessen Organisationsstrukturen, Datenbanken und Vorlagen zurückgreifen kann.

Doch was versteht man unter einer Lernplattform? Technisch gesehen sind Lernplattformen webbasierte Systeme. Dabei wird die Software auf einem Server installiert, den man mit Hilfe eines Webbrowsers über das Internet erreichen kann. Zusätzliche Programme auf den Endgeräten sind daher nicht nötig – sie müssen lediglich internetfähig sein und eine WLAN Anbindung haben (König 2011). Laut Schulmeister (2005, S. 10) „helfen Lernplattformen bei der Organisation, der Initiierung und Aufrechterhaltung von Lehr-Lernprozessen über das Internet“. Inzwischen gibt es zahlreiche nicht kommerzielle (Ilias, Moodle, etc.) und kommerzielle (Fronter, Sharepoint, etc.) Programme mit ähnlichen Basisfunktionalitäten auf dem Markt (König 2011).

Abbildung 3 von König (2011) zeigt, welche Funktionen administrativer und didaktisch - methodischer Natur in Lernplattformen angeboten werden:

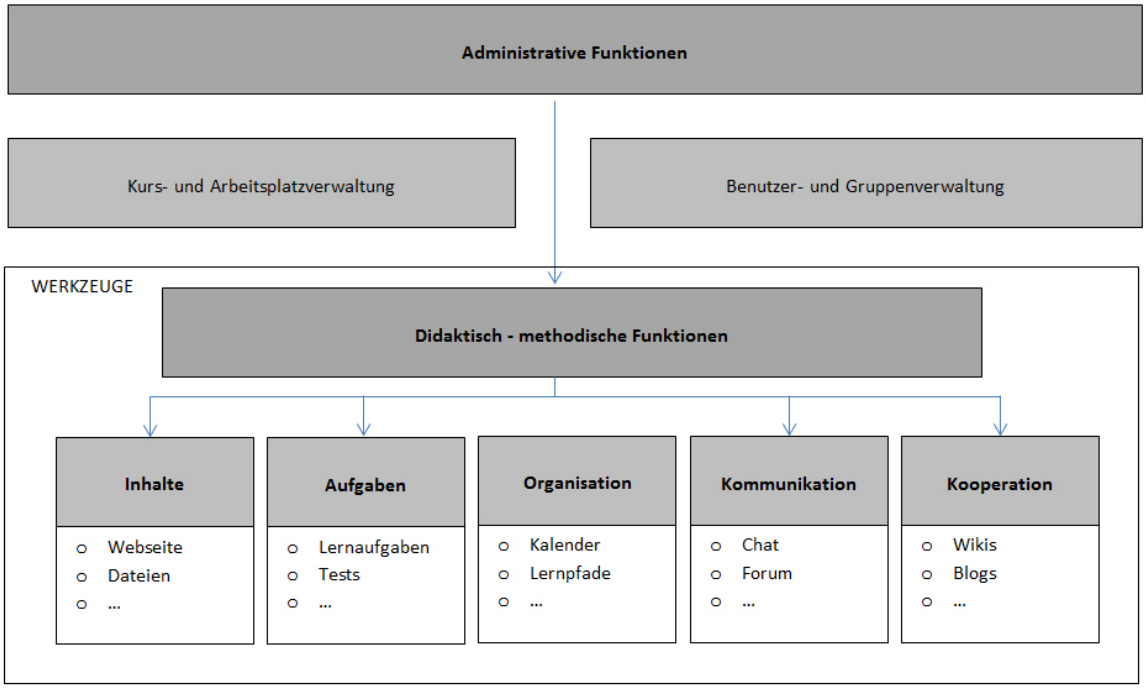


Abbildung 3: Funktionalitäten von Lernplattformen (König 2011)

Durch den Einsatz von Lernplattformen verspricht man sich eine stärkere Hinwendung der Schüler zum selbstregulierten Lernen. In der vorliegenden Studie wurde Moodle als Lernplattform gewählt, weil Moodle sowohl als Autorensystem, als auch als Datenbank verwendet werden kann und somit keine weitere Software bei der Evaluation eingebunden werden musste. Außerdem wird Moodle bereits häufig in Bildungseinrichtungen verwendet und steht kostenfrei

zur Verfügung. Genaueres über Moodle und dessen Struktur kann man in Kapitel 5.4.2.1 nachlesen.

Zusammenfassend lässt sich sagen, dass der Einsatz von Tablet PCs so viele Möglichkeiten und Vorteile im schulischen Kontext mit sich bringen würde, wie kein anderes elektronisches Endgerät. Sie zeichnen sich aus durch ihre Mobilität, intuitive Bedienung und ihre vielfältigen interaktiven, sowie multimedialen Komponenten. Mit einer größeren Verbreitung dieser Minicomputer in der Bildung würde der Absatzmarkt für interaktive Schulbücher, pädagogische Apps und elektronische Diagnosesysteme rasant anwachsen und auch die damit einhergehenden Vorteile mit sich bringen: Elektronische interaktive Schulbücher erlauben es, Rohstoffe, wie Papier in großem Maße einzusparen. Außerdem fallen größere Kosten für Lagerung, Druck und Auslieferung für die Verlage weg. Diese eingesparten Gelder und Ressourcen könnten in die Weiterentwicklung interaktiver und sogar adaptiver pädagogischer Softwareanwendungen gesteckt werden. Computergestützte Diagnosesysteme entlasten die Lehrkräfte bei ihren Korrekturaufgaben und ermöglichen es, Wissensdefizite bei den Schülern aufzudecken und automatisch individuelle Fördermaßnahmen vorzuschlagen oder direkt anzubieten. Die eingesparte Zeit könnte in effektive Fortbildungsmaßnahmen im mediendidaktischen Bereich investiert werden. Auch sollten bereits in der ersten Phase der Lehrerbildung genügend Fort- und Weiterbildungsmöglichkeiten in dieser Hinsicht angeboten und etabliert werden. Die künftigen Lehrergenerationen werden in Sachen Medienkompetenz, dadurch dass sie bereits in einer stärker multimedial geprägten Umwelt aufgewachsen sind, weniger Schwierigkeiten haben. Und durch die Einbindung der elektronischen Unterrichtsmaterialien und Tests in eine Lernplattform / ein Content Management System benötigen Lehrkräfte keine Programmierkenntnisse. Mit der stärkeren Verbreitung der Tablets im Unterricht, wird sich die Forschung ebenfalls verstärkt mit den damit einhergehenden Veränderungen auseinandersetzen müssen und auch die Untersuchung elektronischer Leistungstests weiter in den Fokus gerückt werden.

2.3 Biologiedidaktik

2.3.1 Evolutionäre Anpassung als Beispiel konzeptuellen Wissens in der Biologie

In Kapitel 2.2.2 wurden bereits einige Studien, die formative Leistungstests und das damit einhergehende Feedback in naturwissenschaftlichen Fächern untersuchen und dessen Effekte be-

schreiben, näher beschrieben. Im folgenden Kapitel soll genauer auf das biologische Konzept der evolutionären Anpassung, dessen Komplexität und den damit verbundenen Schwierigkeit im unterrichtlichen Kontext eingegangen werden. Auch die weit verbreiteten Fehlvorstellungen auf Schülerseite, hinsichtlich dieses umfassenden Konstruktes werden isoliert und vorgestellt.
Doch zunächst soll kurz erläutert werden, warum ich gerade den Begriff der evolutionären Anpassung als Beispiel für konzeptuelles Wissen in Biologie für die Unterrichtseinheit meiner Studie gewählt habe:
Die evolutionäre Anpassung streift viele Teilgebiete der Genetik und der Evolution, die in den Curricula fast aller Jahrgangsstufen auftauchen. Eine Unterrichtssequenz über evolutionäre Anpassung kann biologische Fachbegriffe beinhalten, wie Evolution - Anpassung - Ökologische Nische - Selektion - Genmutationen – Vererbung - genetische Variabilität – Population, etc., die geklärt und noch wichtiger, in einen Zusammenhang gestellt werden können. Die Wechselbeziehung der Teilkonzepte zueinander sollte nach dem Unterricht von den Schülern verstanden sein.

Zusätzlich zur allgemeinen Klärung, was man unter evolutionärer Anpassung versteht, kann man anhand des Vogelfluges beispielhaft den Anpassungsvorgang sehr gut veranschaulichen und den Lernern näher bringen. In diesem Zusammenhang gibt es zahlreiche Schülerexperimente und Möglichkeiten um zu Mikroskopieren, so dass fachspezifische Arbeitsweisen zusätzlich in das Unterrichtsgeschehen mit einbezogen und geübt werden können.

Außerdem wurde durch viele Studien (Baalmann et al. 2004; Kattmann et al. 2009; Nehm, Schonfeld 2008) belegt, dass Schüler aller Altersgruppen und sogar Biologiestudenten oft in einigen wichtigen Punkten eine falsche Vorstellung vom evolutionären Anpassungsbegriff haben. Häufig werden Ausprägungen der individuellen und evolutionären Anpassung vermischt oder miteinander gleichgesetzt.

Aufgabe des Unterrichts und der Tests ist es, diese falschen Vorstellungen offenzulegen und zu berichtigen. Welche dies sind, soll im nächsten Kapitel genauer beleuchtet werden.

2.3.1.1 Falsche Schülervorstellungen innerhalb des Konzepts

Engel Clough & Wood - Robinson (1985) stellten fest, dass es für viele Schüler, aber auch Erwachsene schwierig ist, das Thema der „Evolutionären Anpassung“ vollständig zu durchdringen, oder auch nur zu umreißen.

Dies hat mehrere Gründe:
Zum einen handelt es sich bei der Evolution um ein sehr komplexes Konstrukt, das in viele Teilbereiche der Biologie hineinreicht. Zum anderen schreibt Schmidt (2006, S. 69) über die Problematik, dass sich biologische Fachbegriffe, wie Evolution und Evolutionstheorie „gemäß der Dynamik der Wissenschaft und des Zeitgeistes gewandelt haben und dabei auch verwässert worden sind und so schon innerhalb des Faches unterschiedlich oder unstimmig gebraucht werden.“ Außerdem kann das Verstehen der Evolution durch religiöse Weltvorstellungen beeinflusst werden (Bishop & Anderson 1990).

Hinzu kommt, dass die Schüler vorunterrichtliche Vorstellungen, geprägt durch „körperlich-psychische, soziale und umweltliche Erfahrungen, die bereits seit frühester Kindheit an gemacht wurden“ (Baalmann et al. 2004, S. 9) mitbringen, die im Unterricht mit der wissenschaftlichen Perspektive konfrontiert werden.
In diesem Zusammenhang merkt Gropengießer (2003) an, dass sich diese mitgebrachten Vorstellungen häufig auch nicht durch Unterricht in eine wissenschaftliche Richtung lenken lassen. Dies unterstreicht auch eine Studie von Bishop & Anderson (1990), in der der Unterricht trotz großer Bemühungen, die Lerner zu einer wissenschaftlichen Vorstellung von Evolution zu führen, nur mäßigen Erfolg hatte (vgl. Johannsen & Krüger 2005).

Alle oben angeführten Punkte tragen dazu bei, dass das Konzept der „Evolutionären Anpassung“ aufgrund seiner Vielschichtigkeit und Komplexität nicht leicht zu erfassen und noch dazu schwierig zu vermitteln bzw. zu berichtigen ist.

Johannsen & Krüger (2005) untersuchten in diesem Zusammenhang die Vorstellung der Schüler zu verschiedenen Aspekten der Evolution in einer quantitativen Studie, um diejenigen Vorstellungen zu isolieren, mit denen die Lehrkräfte im Unterricht rechnen müssen und dadurch die Planung des Unterrichts zu erleichtern. In der Studie wurden 100 Schüler (10. Jahrgangsstufe), die noch keinen Unterricht zur Evolution erhalten hatten und 206 Kursteilnehmer der Kollegstufe, die bereits Evolutionsunterricht bekommen hatten, gebeten einen Fragebogen auszufüllen. Dabei zeigte sich, dass „finale, anthropomorphe und lamarckistische Vorstellungen prominent in allen Klassen auftraten“ (Johannsen & Krüger 2005, S. 23).

Unter Finalität versteht man eine Zweckgerichtetheit: Merkmale entstehen nicht zufällig, sondern zum Beispiel durch die Steuerung des Lebewesens, damit sie einen bestimmten Zweck erfüllen.
Bei anthropomorphen Vorstellungen werden menschliche Maßstäbe auf Tiere oder Pflanzen übertragen. Tiere handeln hier eigenständig und zielgerichtet und beeinflussen aktiv Veränderungen. Lamarckistische Vorstellungen beziehen sich auf „die Annahmen von Lamarck (1744-1829), dass der Gebrauch und die Nutzung von Organen zu deren Vergrößerung bzw. deren Nichtgebrauch zu deren Verkümmerung führen und dass diese erworbenen Eigenschaften an die folgenden Generationen weitergegeben werden“ (Johannsen & Krüger 2005, S. 26).

Baalman et. al (2004), haben in ihrem Artikel „Schülervorstellungen zu Prozessen der Anpassung – Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion“, die falschen Schülervorstellungen zum Begriffskonstrukt der evolutionären Anpassung isoliert und näher erläutert. Dazu wurden in einer qualitativen Untersuchung entlang eines Leitfadens offene und problemzentrierte Einzelinterviews mit 10 Schülerinnen und Schülern der Klassenstufe 11 bis 13 über den Prozess der Anpassung geführt. Die Interviews beinhalteten verschiedene Teilbereiche der Anpassung: „Natürliche Selektion“, „Variabilität“, „Kampf ums Dasein“ und „Überleben des Tauglichsten“.

In der nachfolgenden Tabelle werden die Fehlvorstellungen zusammengefasst und dargestellt, die sich während der Interviews herauskristallisiert haben:

Tabelle 5: Fehlvorstellungen innerhalb des Konzeptes "Evolutionäre Anpassung"

Falsche Vorstellung	Dieses Verständnis vom Anpassungsbegriff wäre richtig
Gezieltes adaptives Handeln von Individuen: Anpassung wird durch gezieltes und aktives Handeln erreicht. Lebewesen führen bewusst Handlungen aus, die zur Anpassung führen.	Anpassung passiert nicht durch aktives Handeln, sondern passiv durch die **zufällige Mutation von Genen** und die darauf folgende **natürliche Selektion. Unbewusst**!
Vererbung antrainierter Merkmale: Schüler denken, dass adaptive Merkmale durch den wiederholten Gebrauch ausgeprägt oder verstärkt und dann an die nächste Generation weitergegeben werden. Beispiel: Der Giraffenhals wird länger, wenn man sich stärker strecken muss.	**Antrainierte Merkmale sind nicht im Erbmaterial gespeichert*** und werden auch nicht an die Nachkommen weitergegeben. Vielmehr setzen sich zufällig entstandene Merkmale (Genmutation) durch, wenn sie sich als **vorteilhaft erweisen**, da die Lebewesen mit diesem Merkmal **überleben** und sich fortpflanzen.
Absichtsvolle genetische Mutation: Das genetische Material wird vom Organismus oder Körper zum Zweck der Anpassung abgeändert.	**Gene können nicht durch** Individuen und ihr **Handeln verändert werden.**
Anpassung passiert aus einer **Notwendigkeit** heraus. Zum Beispiel, um zu überleben!	Anpassung passiert **zufällig**, weil sich Genmerkmale verändern und sich die vorteilhaften Merkmale durchsetzen.
Anpassung **betrifft die gesamte Population gleichermaßen** (alle sehen gleich aus und haben die gleichen Merkmale / Eigenschaften).	Anpassung betrifft zwar die gesamte Population, ist dabei aber nicht einheitlich, sondern tritt in Variationen auf. (**Genetische Variabilität**)
Manche Schüler denken, Merkmale können sich **innerhalb eines Lebens / einer Generation** anpassen.	Der zeitliche Rahmen einer Anpassung betrifft **mehrere Generationen.**
Natürliche Selektion = Der **Stärkste überlebt**	Natürliche Selektion = Wechselspiel zwischen Lebewesen und Umwelt. **Derjenige, der am besten angepasst ist, überlebt.**
Selektion **hört auf**, sobald eine optimale Anpassung an die Umwelt stattgefunden hat.	Anpassung **endet nie** – es gibt immer wieder veränderte Umweltbedingungen, die Anpassungsvorgänge auslösen.

* Neueste Forschungsergebnisse bei Mäusen zeigen, dass die Speicherung von außen zugefügter Veränderungen in den Genen eventuell doch möglich ist (Epigenetische Effekte (Verhalten im Genom und Weitervererbung)). Allerdings handelt es sich hier nur um erste Ansätze und es wird noch einige Zeit dauern, dass es in diesem Forschungsbereich zu gesicherten Ergebnissen kommt und dann noch einmal eine gewisse Zeit, bis diese Ergebnisse in die klassischen Bildungspläne Einzug halten und im Schulunterricht vermittelt werden.

Die oben angeführten Fehlvorstellungen werden in der vorliegenden Studie wieder aufgegriffen. Um diese bei den teilnehmenden Schülern zu eliminieren oder wenigstens zu verringern, wurden verschiedenste Testaufgaben entwickelt und in den formativen Leistungstests eingesetzt.

Ein Beispiel für eine Testfrage, kann man sich in Tabelle 6 genauer ansehen – alle anderen Testfragen und deren Auswertung befinden sich in Kapitel 6.2.6.

Tabelle 6: Beispiel einer Testaufgabe zur "Evolutionären Anpassung"

Fragentext	Antwortalternativen	Richtig
In einer Population von Schmetterlingen sind fast alle Schmetterlinge hell, nur ein Schmetterling hat eine dunkle Farbe. Was hat ihn wohl dunkel werden lassen?	Innerhalb einer Population kann es gar keine Schmetterlinge mit unterschiedlichen Farben geben.	
	Eine zufällige Veränderung im Erbmaterial.	x
	Im neuen Lebensraum mit dunklen Bäumen hat sich der schlauste Schmetterling getarnt, um zu überleben.	
Wenn das dunkle Merkmal die Chance zu überleben erhöhen würde, was würde dann mit der Population nach einer längeren Zeitspanne wohl passieren?	Alle Schmetterlinge würden dunkel werden,	
	Die Anzahl der dunklen Schmetterlinge würde sich erhöhen, es würde aber dennoch hellere Schmetterlinge geben,	x
	Es würde sich nichts ändern,	
weil	sich alle Lebewesen anpassen, um leichter zu überleben	
	vorwiegend die Lebewesen mit dem dunklen Merkmal überleben und dann vermehren	x
	sich die geringe Anzahl dunkler Schmetterlinge nicht durchsetzen könnte	
und	Merkmale, wie die Farbe nicht im Erbmaterial gespeichert sind und somit nicht vererbt werden können.	
	durch Variationen im Erbmaterial nie die ganze Population gleich aussehen wird.	x

Einzelne Testitems sind in Anlehnung an Johannsen & Krüger (2005), bzw. Kattmann et al. (2009) und Murray (1996) entstanden. Um welche Items es sich genau handelt, kann man dem Dokument „Fragensammlung“ entnehmen (bei Interesse anfordern).

3 Forschungsdesiderata und Hypothesen

Trotz der sehr hohen Resonanz formativer Leistungsmessung in der internationalen *Assessment*-Literatur bleiben, wie bereits in den vorhergehenden Kapiteln angedeutet, einige Forschungsdefizite bestehen, die die vorliegende Studie aufgreifen möchte:

Empirische Grundlage der FLM Forschung bilden Metaanalysen, die sich überwiegend auf wenig ökologisch valide Untersuchungen beziehen. Hinzu kommt, dass es bisher nur eine Studie gibt, bei der versucht wurde, die Merkmale der Rückmeldung experimentell zu variieren und daraus resultierende Effekte zu testen (Yin et al. 2008). Dabei gibt es schon seit Langem Forderungen nach einer umfassenderen empirischen Forschung in dieser Hinsicht (Bennett 2011; Kingston & Nash 2011). In der vorliegenden Studie sollen in einer randomisierten Feldstudie unterschiedliche Rückmeldeformate experimentell variiert und auf ihre Effektivität hin untersucht werden.

Im naturwissenschaftlichen Unterricht gibt es bisher nur wenige Studien, die sich mit Feedbackinterventionen und dem Einsatz und der Methodik formativer Leistungsmessung beschäftigen (z.B. Anderson et al. 2007).

Auch der Einsatz von elektronischen Diagnosesystemen und digitalen Endgeräten und deren Effektivität bei der formativen Leistungsmessung im naturwissenschaftlichen Unterricht wurde nur vereinzelt untersucht. Erschwerend kommt hinzu, dass der Fokus dieser Projekte nicht auf das Thema „Feedback“ gerichtet war (z.B. Wang et al. 2006 und 2011, Thissen-Roe et al. 2004, Hickey & Zuiker 2012). Vielmehr wurde der Einsatz unterschiedlicher Arten von Software, die zum Teil kollaborative Aspekte beinhalteten oder verstärkt Aspekte des selbstregulativen Lernens förderten, genauer betrachtet. In meiner Studie soll überprüft werden, welche Effekte verschiedene technologisch unterstützte Feedbackvariationen auf den Wissenszuwachs und die Motivation der Schüler im naturwissenschaftlichen Unterricht haben können.

Einige Studien (Shavelson & Ruiz-Primo 1999, Anderson et al. 2007, Mc Conell et al. 2006, Furtak & Ruiz-Primo 2008, Trumpower & Sarwar 2010) befassen sich mit den unterschiedlichen

Wissensleveln, die im naturwissenschaftlichen Unterricht auftauchen und für die diese Fächergruppe auch typisch ist (neben deklarativem Wissen, auch konzeptuelles, prozedurales und strategisches Wissen) und damit, welche elektronischen Diagnosesysteme sich am besten dafür eigenen, diese verschiedenen Arten von Wissen abzuprüfen. Dabei werden verstärkt Concept Maps, Concepttests und Diagnosesysteme mit offenen Fragestellungen eingesetzt. Allerdings befassen sich diese Studien vorwiegend mit der Frage, welche Aufgabenformate dafür am besten geeignet sind, stellen aber nicht das Feedback in den Mittelpunkt.

Neben dem gerade beschriebenen Forschungszweig, gibt es einen Forschungsschwerpunkt in der Biologiedidaktik, der sich mit dem Konzept der evolutionären Anpassung und den damit einhergehenden Fehlvorstellungen von Schülern befasst. Hierzu zählen Studien von Baalmann et al. (2004), Gropengießer (2003), Bishop & Anderson (1990) oder auch Johannsen & Krüger (2005). Allerdings geht es in diesen Projekten darum, die Fehlvorstellungen zu identifizieren, zu benennen und herauszufinden, wodurch sie verursacht werden. Auch werden unterrichtliche Empfehlungen in methodischer und didaktischer Hinsicht ausgesprochen, die helfen sollen, diese Fehlvorstellungen zu berichtigen. In der vorliegenden Studie möchte ich versuchen, die auftretenden Fehlvorstellungen zum Konzept der Evolutionären Anpassung mit Hilfe geeigneter Aufgabenformate und dem dazu passenden Feedback in formativen Leistungstests auszumerzen, also die beiden zuvor genannten Forschungsstränge zu verbinden. Hinzu kommt, dass ich die Tests in elektronischer Form abhalten möchte, also mit Hilfe von Moodle (Lernplattform) und Tablet PCs (digitale Endgeräte). Dieses Vorgehen stellt eine weitere Verknüpfung zu den Forschungsgebieten der nachfolgenden Unterpunkte dar:

Mit den technologischen Möglichkeiten hinsichtlich formativer Leistungsmessung und elektronischer Diagnosesysteme beschäftigt sich zum Beispiel Russell (2010). Er beschreibt verschiedene Hauptbereiche, in denen sich Technologien, die formative Leistungsmessungen elektronisch unterstützen können, bereits sehr schnell entwickeln. Feldman & Capobianco (2008) untersuchen *classroom response systems* und Yeh (2006, 2007, 2009), sowie Merrell & Tymms (2007) betrachten zum Beispiel *Interactive Computerised Assessment Systems (InCAS)* genauer. Beide Systeme sollen Lehrkräfte bei der Leistungsmessung und der Diagnose von Leistungsdefiziten bei Schülern unterstützen.

Mit den Vorteilen, die der Einsatz mobiler Endgeräte im Unterricht mit sich bringt, beschäftigen sich unter anderem Aufenanger & Schliezeit (2013) und Baumgartner et al. (2011). Ludwig (2013) versucht herauszufinden, warum die neuen Technologien, trotz der unbestreitbaren Vorteile, die sie mit sich bringen nur sehr selten in den Unterricht integriert werden und Roschelle et al. (2010) legen ihren Forschungsschwerpunkt auf kollaborative Aspekte des eLearnings. Allerdings bleiben diese Forschungen eher bei allgemeinen Faktoren des Einsatzes elektronischer Medien im Unterricht. In der vorliegenden Studie soll vor allem der Einsatz multimedialer Endgeräte in Bezug auf formative Leistungsmessung und Rückmeldeformate näher untersucht werden.

Aus der Auflistung der Forschungsdesiderata wird ersichtlich, das es noch kaum ökologisch valide Experimentalstudien in einem schulischen Kontext im Allgemeinen und im naturwissenschaftlichen Unterricht im Besonderen gibt, in denen Feedbackmerkmale im Rahmen eines systematischen Konzepts formativer Leistungsmessung variiert werden und bisher nicht untersucht wurde, in welchem Maße neue informationstechnische Möglichkeiten (Lernplattformen, Tablet-PCs) für formative Leistungsmessungen herangezogen werden können.

Daher rücke ich folgende Aspekte in den Mittelpunkt der Evaluation:

- Den **naturwissenschaftlichen Kontext**. Die formativen Leistungstests werden innerhalb einer, im Lehrplan verankerten, **Unterrichtseinheit** in Biologie durchgeführt. Dabei wird beim Design der Sequenz und bei der Ergebnisauswertung besonderer Wert auf die Unterscheidung der Effekte bei der Aneignung, dem vertieften Verständnis und der Behaltensleistung von ***konzeptuellem Wissen*** und ***Faktenwissen***, gelegt. Diese Unterscheidung wird vorgenommen, da es besonders in naturwissenschaftlichen Fächern sehr viele komplexe Zusammenhänge und Konstrukte gibt, die weit über reines Faktenwissen hinausgehen und daher auch einen besonderen Aspekt des naturwissenschaftlichen Unterrichts darstellen. In die Studie werden die Konzepte „Evolutionäre Anpassung“ und „Physikalischer Auftrieb“ als Beispiele konzeptuellen Wissens eingebunden.

- Die **empirische Untersuchung** verschiedener Feedbackmethoden: Die **Rückmeldungen** werden innerhalb des Treatments **systematisch variiert** und hinsichtlich ihrer Effekte auf die Schülerleistung und –motivation untersucht.
- Den Einsatz **elektronischer** formativer Leistungstests (Tablet PCs), denn gerade diese ermöglichen es, den Schülern eine direkte, individuelle, ausführliche und sogar adaptive Rückmeldung zu geben (Russell 2010). Dabei wird Moodle als Lernplattform und Tablet PCs als Endgeräte verwendet. Moodle stellt eine Reihe unterschiedlicher Aufgabenformate und Rückmeldeoptionen zur Verfügung und dient darüber hinaus als Datenbank, in der alle Schülerdaten und Testergebnisse automatisch gespeichert werden.

Maier (2014, S. 69) beschreibt die Forschungssituation folgendermaßen: „Schulpraktisch einsetzbare Systeme und gut abgesicherte Leistungseffekte findet man vor allem bei der computergestützten Diagnose und Förderung überwiegend prozeduraler Basisfertigkeiten. Ebenso zeichnen sich interessante Entwicklungen im Bereich der Diagnose konzeptuellen Wissens ab. Die Nutzbarmachung computergestützter, formativer Diagnostik für die tägliche Unterrichtspraxis bleibt jedoch ein großes Desiderat.“
Und auch Wang (2011) fordert, weitere *formative e-assessment* Strategien zu entwickeln, um deren Effektivität im Einzelnen zu untersuchen.

In Bezug auf frühere Forschungsergebnisse, denke ich, dass eine ausführliche Rückmeldung (elaboriertes Feedback) zu den Fehlvorstellungen der einzelnen Schüler innerhalb konzeptueller Wissenszusammenhänge effektiver sein wird (Kluger & DeNisi 1996, Bangert-Drowns et al. 1991), als ein einfaches dichotomes Feedback, wo der Lerner nur darüber informiert wird, ob er richtig oder falsch geantwortet hat. Weiterhin sollte die Formulierung der Rückmeldungen möglichst auf der Ebene des Aufgabenlösungsprozesses (Kluger & DeNisi 1996) stattfinden und rein informativ (ohne Lob oder Kritik), also auf die Aufgabe und nicht auf die Person gerichtet sein, damit sie nicht vom Lerninhalt ablenkt (Dweck 1986, Butler 2005, Radoczy et al. 2008). Außerdem vermute ich, dass bestimmte Persönlichkeitsmerkmale eine wichtige Rolle dabei spielen werden, die Feedbackeffekte zu erklären (Maier, Hofmann & Zeitler 2012).

Dazu gehören Merkmale, wie zum Beispiel, ob ein Schüler gewissenhaft arbeitet, das Feedback gründlich liest und reflektiert, oder ob er denkt, es wäre nicht nötig, sich damit auseinanderzusetzen. Auch wie offen ein Lerner auf Hilfeangebote reagiert, könnte die Effekte beeinflussen.

Im Rahmen der oben beschriebenen äußeren Bedingungen (Unterrichtseinheit Biologie, Sekundarstufe, schulischer Kontext, elektronische Leistungstests, Variation des Feedbacks, etc.) möchte ich folgende Hypothese aufstellen und auf ihre Richtigkeit hin überprüfen:

Schüler, die während der formativen Leistungstests ein ausführliches Feedback erhalten und dieses auch nutzen, schneiden beim Post- und Behaltenstest bei der Beantwortung von Fragen zum

a) Fakten- und

b) Konzeptwissen

signifikant besser ab, als Schüler, die kein elaboriertes Feedback erhalten, oder die angebotene Rückmeldung nicht nutzen.

Da Bangert-Drowns et al. (1991) herausgefunden haben, dass der Informationsgehalt der Rückmeldung von Bedeutung ist, wenn es um Effektstärken geht und es höhere Effekte gab, wenn den Schülern ein elaboriertes Feedback zur Verfügung gestellt wurde, möchte ich den Schülern der Treatmentgruppe eine Rückmeldung anbieten, die nach Bangert-Drowns (1991) in die Kategorie *„instruction based elaboration“* fällt, also den im Unterricht behandelten Lösungsweg für die Aufgabe noch einmal aufgreift und erläutert. Dabei achte ich darauf, dass die richtigen und falschen Antworten als solche kenntlich gemacht werden und zusätzlich zurückgemeldet wird, wie die richtige Lösung ausgesehen hätte und warum die gegebene Antwort falsch / korrekt war. Das Feedback selbst wird sowohl direkt nach der Beantwortung der Frage, als auch nach Beendigung des gesamten Tests bereitgestellt.

Eine ausführliche Beschreibung des gesamten Treatments findet man in Kapitel 5.3.2.3.

4 Organisatorische und inhaltliche Vorbereitung der Studie

4.1 Vorüberlegungen und Zielsetzung

Ziel der Studie ist die empirische Überprüfung der kognitiven und motivationalen Effekte einer Feedbackstrategie innerhalb eines FLM-Konzeptes im Biologieunterricht der Sekundarstufe 1:

Die Hypothesen sollen dabei innerhalb einer randomisierten Feldstudie mit Hilfe einer zeitlich und thematisch vorstrukturierten Unterrichtseinheit zum Vogelflug überprüft werden. Da die quasi-experimentelle Feldstudie von Yin et al. (2008) gezeigt hat, dass Lehrkräfte unterschiedliche Methoden und Unterrichtstechniken einsetzten sollen die Rückmeldungen innerhalb einer Klasse variiert werden. Die Lehrkräfte dürfen allerdings ihre methodisch-didaktische Vorgehensweise beibehalten (konkrete Methoden, Sozialformen, Aufgaben, Experimente, etc.).
„Die Gültigkeit der Hypothesenprüfung hängt sowohl von der internen als auch der externen Validität der Untersuchung ab (Shadish, Cook & Campbell 2002). Die externe Validität ist aufgrund der Durchführung des Experiments im Schulunterricht bereits relativ hoch.

Die interne Validität wird gewährleistet, durch die zufällige Zuteilung der Schüler zu den Versuchsgruppen und dadurch, dass jeweils die gleiche Lehrperson alle Treatment- und Kontrollgruppen während der gesamten Unterrichtszeit gemeinsam unterrichtet. Außerdem sind die Treatment- und Kontrollbedingungen genau vorgegeben. (Maier & Randler 2011)

Im Mittelpunkt des Interesses steht die Effektivität unterschiedlicher Rückmeldeformate auf die Schülerleistung und -motivation und die unterschiedlichen Möglichkeiten der technischen Realisierung elektronischer, formativer Leistungstests.

4.2 Verlauf der Studie

Zu Beginn der Studie, wurde ein Arbeitsplan erstellt, den es nach und nach abzuarbeiten galt. Allerdings stellte sich an unterschiedlichen Punkten heraus, dass Änderungen notwendig werden, um die Hypothesen der Studie vernünftig überprüfen zu können.

Zum Beispiel zeigte sich während der ersten Vorstudie sehr schnell, dass eine sinnvolle Variation des Feedbacks nur mit Hilfe elektronischer Leistungstests möglich ist (siehe Punkt 4.2.2).

Im Folgenden werden die verschiedenen Schritte vorgestellt, die für einen erfolgreichen Verlauf der Evaluation notwendig waren:

- Fachdidaktische Literaturrecherche
- Erstellung einer Unterrichtseinheit zum Thema „Vögel und deren Anpassung an das Fliegen“. Dabei galt es darauf zu achten, biologische Themen mit diagnostischem Potenzial zu finden, die sowohl Faktenwissen, als auch konzeptuelles Wissen beinhalten. Zusammenstellung der Unterrichtsmaterialien (Experimentierhefte, Materialien für den Lernzirkel, etc.). Die Unterrichtseinheit wurde teilweise veröffentlicht und kann dort eingesehen werden (Maier, U., Wolf, N., Randler, C. 2014).
- Entwicklung der diagnostischen Aufgaben (Paper-Pencil Tests) und der Rückmeldeformate für die Unterrichtseinheit. Um konzeptuelles Wissen besser abprüfen zu können, wurde ein Aufgabenformat entwickelt, das eine Mischung aus Multiple-/Single-Choice und Lückentext darstellt (siehe 4.2.2).
- Durchführung der Vorstudie 1 in einer M7 an einer Mittelschule in Fürth: Erprobung der Diagnoseaufgaben und der Unterrichtseinheit; teilnehmende Beobachtung und Sammlung von Schülerdokumenten (Aufgabenlösungen).
    - Beim ersten Durchgang der Vorstudie stellte sich heraus, dass es mit Paper - Pencil Tests, sehr schwierig ist, dem Schüler ein individuelles Feedback hinsichtlich seines Lernstandes zu geben – eine Variation des Feedbacks war unmöglich, da ausführlichere Rückmeldungen auf einer Kontrollfolie keinen Platz gefunden hätten. Daher wurde nach der Vorstudie 1 beschlossen, die formativen Leistungstests in elektronischer Form auf Tablet PCs durchführen zu lassen.

      Dies führte zu den nächsten Schritten:
- Präzisierung verschiedener Rückmeldestrategien (für die elektronischen Tests) in Abhängigkeit der Aufgabenkomplexität und der Lernziele; Diskussion der Materialien mit der Lehrkraft der Vorstudie und anschließende Überarbeitung der Unterrichtseinheit und Testaufgaben.

- Suche nach geeigneten elektronischen Endgeräten und Autorensystemen zur Erstellung der Testaufgaben und einer geeigneten Lernplattform zur Speicherung der Schülerdaten.
- Entwicklung und Programmierung der elektronischen Fragen, Einbindung der Testaufgaben in die Lernplattform Moodle, Installation eines Laptops, der als Server fungieren kann, Anschaffung und Konfiguration der Tablet PCs für den Einsatz im Unterricht (siehe 5.4). Zwischenevaluation der Testfragen und Überprüfung der Technik in einer 7. Klasse (iPad Klasse) einer Realschule in Erlangen.
- Durchführung der Vorstudie 2 in einer M7 einer Mittelschule in Fürth und genaue zusätzliche Überprüfung der Diagnoseaufgaben aus dem Bereich „Evolutionäre Anpassung“ an einer Realschule in Erlangen: Testdurchlauf für die Technik und die Testfragen.
- Optimierung und Weiterentwicklung der Diagnoseaufgaben und deren Rückmeldungen.
- Zusammenstellung der Lehrermaterialien (auf einer CD), Vorbereitung und Durchführung des Workshops mit Lehrkräften aus Mittelschulen des Großraumes Erlangen – Fürth – Nürnberg, Akquise weiterer interessierter Lehrkräfte aus Realschulen und Gymnasien.
- Durchführung des Feldexperiment während des gesamten Schuljahres 2012/2013 an 10 Klassen unterschiedlicher Schularten in den Jahrgangsstufen 6 und 7, Betreuung der Datenerhebung
- Dateneingabe und Datenaufbereitung, Korrektur des Behaltenstests
- Auswertung der quantitativen Daten, Rückmeldung an interessierte Lehrkräfte
- Präsentation der Forschungsbefunde auf nationalen und internationalen Tagungen; Fertigstellung von Publikationen

Eine ausführliche Beschreibung der oben angeführten Punkte, befindet sich in den nachfolgenden Kapiteln.

4.2.1 Entwicklung prototypischer Diagnoseaufgaben und Feedbackformate

Die modellkonforme Umsetzung der Versuchsbedingung ist Voraussetzung für die Prüfung der Hypothesen und wird durch die Erarbeitung des FLM-Konzeptes und die genaue Planung der Rückmeldestrategien gewährleistet. Zunächst wurde mit Hilfe von biologischer Fachliteratur, Kinderbüchern zum Thema „Vogelflug“ und verschiedensten Schulbüchern aller Schularten, eine Unterrichtssequenz zum Thema „Vögel und deren Anpassung an das Fliegen“ entworfen. Dabei war besonders wichtig, zu Beginn der Sequenz alle wichtigen Sachverhalte zum Thema „evolutionäre Anpassung“ im Allgemeinen und Anpassungsmerkmale der Vögel im Speziellen mit den Schülern zu erarbeiten, um das Erlernte später auf den eigentlichen Schwerpunkt, den Vogelflug anwenden zu können. In einem letzten Unterrichtsteil werden die verschiedenen Flugarten und deren Einsatz im Detail beleuchtet. Der Vogelzug, der vor allem mit Hilfe der Thermik stattfindet, beendete die Sequenz. Eine ausführliche Beschreibung der Unterrichtseinheit findet man unter 5.2. Anhand der vorliegenden Unterrichtssequenz wurden im Anschluss prototypische Diagnoseaufgaben für die zentralen Lernziele entwickelt. Dabei war wichtig, einen Aufgabentyp zu wählen, der es nicht nur erlaubt, Faktenwissen abzufragen, sondern auch Zusammenhänge (konzeptuelles Wissen), also ein tieferes Verständnis komplexerer Konstrukte (Auftrieb, Evolutionäre Anpassung) zu erfassen. Diese Prototypen wurden in den beiden Vorstudien mehrmals überprüft und überarbeitet. Dafür wurden die Schüler bei der Bearbeitung der Aufgabenstellungen beobachtet, sie zur Verständlichkeit der Aufgabenstellung befragt und die elektronisch vorliegenden Testergebnisse aus der zweiten Vorstudie näher analysiert und daraus Möglichkeiten für die Optimierung der Diagnoseaufgaben und der Rückmeldeformate abgeleitet.

Auf die detaillierten Schritte bei der Erarbeitung der Diagnoseaufgaben, wird in Punkt 4.2.2 und auf das Design und die Wahl der Aufgabentypen in Punkt 5.3 (Aufgaben- und Testdesign) näher eingegangen. Die Rückmeldestrategien und das Feedbackdesign werden unter Punkt 5.3.3 genau beschrieben.

4.2.2 Durchführung der Vorstudien

Vor der Durchführung der ersten Vorstudie, stand die Erarbeitung der Unterrichtssequenz (siehe 5.2) und die inhaltlichen Entwicklung von prototypischen Diagnoseaufgaben im Vordergrund. Dabei wurden mit Hilfe eines Brainstormings Aufgaben unterschiedlichen Typs (siehe

5.3.1) zu allen Lernzielen entworfen und dann diejenigen ausgewählt, die am präzisesten, das vermittelte Wissen abfragen konnten.

Um die Durchdringung des Lerninhalts und insbesondere der beiden Konzepte „Auftrieb“ und „Evolutionäre Anpassung“ besser abprüfen zu können, wurde ein Aufgabentyp entwickelt, der eine Mischung aus Multiple-Choice Format und Lückentext darstellt:

Tabelle 7: Aufgabenformat aus den schriftlichen formativen Leistungstests der Vorstudie

Die Schwanz-federn des Vogels	**A**		**E**	dienen als Schutz vor Kälte oder Hitze.	Um diese Aufgaben besonders gut erfüllen zu können, müssen die Schwungfedern folgen-dermaßen aufgebaut sein:	**K**	Ihre Federfahne ist luftundurchlässig.
	B		**F**	schützen vor Verletzungen. Außerdem verleihen sie dem Vogel seine glatte Oberfläche.		**L**	Ihre Federfahne ist luftdurchlässig.
			G	werden bei der Balz einge-setzt, um das Vogelweib-chen zu beeindrucken.			
	C		**H**	bilden die Tragflächen für den Flug.		**M**	Ihre Federfahne ist rau.
	D		**J**	helfen beim Steuern und Bremsen beim Flug.		**N**	Ihre Federfahne ist manchmal bunt und auffällig.

Dadurch sind die Schüler gezwungen, sich nicht nur bei den „Multiple Choice Teilen“ der Aufgaben für die richtige Auswahl zu entscheiden, sondern diese auch noch in die richtige Beziehung zu den Aussagen, der vorhergehenden und nachfolgenden „Lücken“ des Textes zu bringen. Die Schüler kreuzen den Buchstaben an, den sie für richtig halten. Der dicke Strich zeigt den Beginn eines neuen Satzes an.

Auch die Erarbeitung von Möglichkeiten einer Selbstkontrolle durch die Schüler bei den Testaufgaben und des damit einhergehendes Feedbacks an die Schüler war ein wichtiger Teil dieser Projektphase:

Vor allem das Design der Selbstkontrollbögen, stellte sich als sehr schwierig heraus, denn diese sollten übersichtlich bleiben und die Möglichkeit bieten, individuelles Feedback zu geben, das über die Rückmeldung „richtig“ oder „falsch“ hinausgeht. Letzteres stellte sich während der

Vorstudie als unmöglich heraus: Es wurde klar, dass die Lösung in elektronischen Leistungstests zu suchen war (siehe hierzu 5.4.1).

Ob die Testaufgaben verständlich formuliert wurden, der Aufgabentyp von den Schülern gut angenommen und verstanden wird und wie gut die Kontrollbögen funktionieren, wurde in einer M8 an einer Mittelschule in Fürth getestet. Nach mehreren Gesprächen mit den Schülern und der unterrichtenden Lehrkraft, wurden einzelne Formulierungen geändert und angepasst. Der Aufgabentyp wurde von allen Schülern, als gut und leicht verständlich bezeichnet, doch die Arbeit mit dem Selbstkontrollbogen stellte sich als sehr zeitaufwändig und unübersichtlich heraus – hier ein kleiner Auszug aus dem Kontrollbogen, damit man sich ein besseres Bild machen kann:

Tabelle 8: Kontrollbogen Version 1 zum FLT - Vorstudie

Aufgabe	Thema	Richtige Antwort		Richtig	Falsch
1		**A, C, D, E, F**	Vögel verwenden ihren Schnabel unter anderem zum Transport von Nistmaterial, zum Füttern der Jungen, zur Pflege ihres Gefieders, zum Festhalten ihrer Beute und als Waffe.	○	○
2		**D, G, J, O, Z**	Der Mäusebussard ist ein Greifvogel und ernährt sich vor allem von Mäusen. Dabei verwendet er seinen Schnabel wie einen Seitenschneider. Um seine Nahrung festhalten zu können, benötigt er kräftige Füße mit 3 nach vorne gerichteten und einer nach hinten gerichteten Zehe mit scharfen Krallen, die sich beim Festhalten in die Beute bohren lassen.	○	○
3		**A, H, K, P**	Die Deckfedern des Vogels befinden sich über den Daunenfedern und schützen vor Verletzungen und Kälte. Außerdem geben sie dem Vogel seine glatte Oberfläche. Um diese Aufgabe besonders gut erfüllen zu können, müssen die Deckfedern folgendermaßen aufgebaut sein: Ihre Oberfläche ist glatt und stabil	○	○

Folgende Probleme traten bei der Arbeit mit diesem Kontrollbogen in Erscheinung:

- Die Schüler sehen auf dem Bogen die Buchstaben der richtigen Antworten und müssen immer zwischen ihrem Test und der Kontrollinstanz hin und her springen, um zu erarbeiten, welche ihrer Antworten richtig, bzw. falsch waren und diese dann entsprechend kennzeichnen.

- Es ist nur eine Möglichkeit für Feedback möglich: Man schreibt die Antwort hin, die richtig gewesen wäre. Es ist nicht möglich auf Falschantworten und die darin enthaltenen Denkfehler und Verständnisprobleme detailliert und individuell einzugehen.
- Es ist wenig motivierend, wenn die Schüler zwar mehrere Teilaufgaben richtig beantwortet haben, die Antwort insgesamt aber als falsch gewertet wird, wenn auch nur eine Teilantwort nicht richtig war.

Deshalb wurden an der ersten Version mehrere Änderungen vorgenommen: Es wurde eine genaue Punktvergabe eingeführt und der Selbstkontrollbogen selbst als Folie konzipiert (Auszug – siehe unten), die man direkt über die jeweilige Testseite legen konnte. Dadurch sollten die Schüler die Antworten übersichtlicher und schneller auf ihre Richtigkeit hin überprüfen können, und gerechter bewertet werden:

Tabelle 9: Kontrollbogen Version 2 zum FLT - Vorstudie

Die Schwanzfedern des Vogels	**A**		**E**	dienen als Schutz vor Kälte oder Hitze.	Um diese Aufgaben besonders gut erfüllen zu können, müssen die Schwungfedern folgendermaßen aufgebaut sein:	**K**	Ihre Federfahne ist luftundurchlässig.
	B		**F**	schützen vor Verletzungen. Außerdem verleihen sie dem Vogel seine glatte Oberfläche.		**L**	Ihre Federfahne ist luftdurchlässig.
			G	werden bei der Balz eingesetzt, um das Vogelweibchen zu beeindrucken.			
	C		**H**	bilden die Tragflächen für den Flug.		**M**	Ihre Federfahne ist rau.
	D		**J**	helfen beim Steuern und Bremsen beim Flug.		**N**	Ihre Federfahne ist manchmal bunt und auffällig.

Die Folie wird direkt über den Test gelegt und die grau hinterlegten Kästchen zeigen die richtigen Antworten an. In einer Auswertebox können die Schüler die Anzahl ihrer falsch angekreuzten (weißen) Kästchen und ihrer richtig angekreuztes (grauen) Kästchen eintragen und dann die „falschen Punkte“ von den „richtigen Punkten“ abziehen – dadurch erhalten sie ihre Punktzahl für die Aufgabe.

Tabelle 10: Auswertung- Kontrollbogen

Frage		Punkte
3		

Doch auch diese Version war viel zu kompliziert und zeitaufwändig und auch hier gab es keine Möglichkeit, ein individuelles Feedback für die einzelnen Schüler einzubinden.

Eine weitere Idee, um auch Rückmeldung zu den Falschantworten geben zu können, war das Anbringen einer Leiste am hinteren Rand der Tests, wo alle ankreuzbaren Buchstaben der Frage auftauchen und einen dazugehörigen Kontrollbogen, wo diese Leiste links auftaucht, um sie an die entsprechende Leiste im Test anzulegen. Hier wurden die richtigen Antworten grau hinterlegt und eine an die Leiste angeschlossene Tabelle bot genügend Platz, um auch bei falschen Antworten ein Feedback geben zu können:

Tabelle 11: Kontrollbogen Version 3 zum FLT - Vorstudie

A	Die Deckfeder besteht aus einem Daunenenteil und einer festen Federfahne. Sie schützen vor Verletzungen und glätten den Vogelkörper.
B	Die Daunenfeder lässt sich aufplustern. Die dadurch entstehende Luftschicht schützt vor Kälte.
C	Die Schwungfedern haben eine asymmetrische, luftundurchlässige Federfahne. Sie bilden die Tragflächen für den Flug.
D	Typische Schwanzfeder mit symmetrischer Federfahne.
E	Platz für Kommentare, Anregungen, Verbesserungen
F	
G	Die Schwanzfedern können sehr kunstvoll aussehen (Pfau) und sollen bei der Balz die Weibchen beeindrucken.
H	
J	Die Schwanzfedern übernehmen beim Flug das Steuern und Bremsen.
K	Ihre Federfahne ist luftundurchlässig.
L	
M	
N	Manchmal sind die Schwanzfedern auch bunt und auffällig.

Doch auch diese Version stellte sich als nicht wirklich befriedigend heraus, da die Leisten und Tabellen zu viel Platz benötigten und dadurch die Test und auch die Kontrollbögen jeweils ca. 10 Seiten in Anspruch nehmen würden – ein Papierchaos wäre vorprogrammiert gewesen.

Natürlich hätte die Möglichkeit bestanden, die Tests einzusammeln, diese zu korrigieren, mit dem entsprechenden Feedback zu versehen und am Anfang der darauffolgenden Stunde auszuteilen, aber dadurch wären andere Nachteile zum Tragen gekommen: Der hohe Korrekturaufwand wäre hinzugekommen und die Möglichkeit der Selbstreflexion auf Schülerseite wäre

weggefallen. Außerdem wäre eine durchgängig gleiche Variation des Feedbacks für die verschiedenen Schülergruppen so gut wie unmöglich geworden. Die Rückmeldetexte müssen ja innerhalb der einzelnen Gruppen identisch bleiben und dürfen nicht variieren, sonst ist eine Untersuchung der einzelnen Feedbackvarianten nicht durchführbar.

Da es bei der Studie aber genau um die Variation des Feedbacks gehen sollte, zeichnete sich schnell ab, dass man elektronische Tests einsetzen muss, um die Forschungsfrage hinreichend beantworten zu können.

In dieser ersten Vorstudie wurden natürlich auch die Unterrichtssequenz und die für die Durchführung der Unterrichtseinheit benötigten Materialien geprüft.

Für die Arbeit im Lernzirkel stand den Lehrern ein Experimentierkasten zur Verfügung. Dies sollte es für die Lehrkräfte leichter machen und deren Akzeptanz erhöhen, mit Experimenten zu arbeiten. Außerdem wurde eine Materialliste zusammengestellt, auf der alle benötigten Materialien für die einzelnen Unterrichtsstunden aufgeführt waren (bei Interesse bitte anfordern), damit die Lehrkräfte einen besseren Überblick hatten, was vor der jeweiligen Stunde bereitgestellt werden musste.

Für die Schüler wurden zwei verschiedene Schülerordner erstellt, die folgende Materialien beinhalten sollten:

Tabelle 12: Übersicht über den Inhalt der Schülerordner aus der Vorstudie

Treatmentgruppe	**Kontrollgruppe**
Deckblatt	
Erstellung des Kennwortes	
Lernziele	
Alle Unterrichtsmaterialien zum Thema Vogelflug (kommen in der Unterrichtseinheit nach und nach dazu)	
Dokumentation des Lernerfolgs (nach Pretest)	

Ein Ordner sollte mit allen **Unterrichtsmaterialien**, wie Hefteinträgen, Experimentierheften und Arbeitsblättern gefüllt werden und so am Ende die gesamten Lerninhalte der Unterrichtseinheit aufnehmen. In der Einführungsstunde, wurde jedem Schüler einen Schnellhefter mit den Materialien ausgeteilt und diese einzeln mit ihnen besprochen.

Zunächst mussten alle Lerner ein Kennwort erstellen, das sich aus einer Kombination von Buchstaben (aus ihrem Namen) und Zahlen (aus ihrem Geburtsdatum) zusammensetze und für alle Tests verwendet wurde:

Gib den ersten und den letzten Buchstaben deines Vornamens an:

1.Buchstabe ☐ Letzter Buchstabe ☐

Dazu den Geburtstag (nur den Tag) und m/w. Dies hilft uns, die Zuordnung zu erleichtern.

Geburtstag: ☐☐

Kreuze bitte an: **w**eiblich = **w** / **m**ännlich = **m**.

w ☐ ***m*** ☐

Gesamtes Kennwort: ☐☐☐☐☐

Beispiel:
Wäre dein Name „Thomas“ (= männlich) und dein Geburtsdatum 04.2.2002 würde das Kennwort TS04m lauten.

Abbildung 4: Erstellung des Schülerkennworts für die Vorstudie

Die „Dokumentation des Lernerfolgs“ (siehe Abbildung 22), erhielten nur die Schüler der Treatmentgruppe direkt nach dem Pretest ausgehändigt – hier konnten sie ihre erreichten Punkte in den verschiedenen Lernkategorien eintragen und von Test zu Test ihren Lernfortschritt festhalten und verfolgen. Meine Beobachtungen zeigten, dass es für die Schüler sehr motivierend und spannend war, ihre Ergebnisse einzutragen und mit ihren Mitschülern zu vergleichen.

Der zweite Ordner diente als Testordner und wurde farblich zwischen Treatmentgruppe (rot) und Kontrollgruppe (blau) unterschieden. Die Schüler wurden zufällig den beiden Gruppen zugeteilt und bekamen die Ordner nur zum Abheften ihrer Tests ausgehändigt, ansonsten verblieben diese bei der Biologielehrkraft. So hatte ich immer Zugriff auf die Tests und konnte nachkontrollieren, wie die Tests ausgefüllt waren und die Arbeit mit den Kontrollfolien lief.

Tabelle 13: Übersicht über den Inhalt der Testordner aus der Vorstudie

Treatmentgruppe	Kontrollgruppe
Deckblatt	
Pretest und Motivationstest 1	
Kurzskala 1 (MOT)	
FLT 1 und Kontrollbogen	KG 1
Kurzskala 2 (MOT)	
FLT 2 und Kontrollbogen	KG 2
Kurzskala 3 (MOT)	
FLT 3 und Kontrollbogen	KG 3
Posttest 1 und Motivationstest 2	
Behaltenstest (Posttest 2)	

Während der Testphasen wurden die Klassen getrennt. Die Biologielehrkraft schrieb die Tests mit der Kontrollgruppe während ich mit der Treatmentgruppe beschäftigt war und das Ausfüllen der Frage- und Kontrollbögen beobachtet habe, sowie Hilfestellung dabei leistete.

Der Pre- und Posttest der Vorstudie sollte den Wissenstand der Schüler zum gesamten Lerninhalt der Unterrichtssequenz abbilden und bestand deshalb, wie herkömmliche Probearbeiten, aus verschiedenen Aufgabetypen (auch halboffenen und offenen Fragen). Dadurch brachten sie einen hohen Korrekturaufwand mit sich. Auch die Tests der Kontrollgruppe waren an die Struktur herkömmlicher Probearbeiten angepasst, enthielten aber nur Fragen zum vorangehenden Unterrichtsblock und dessen Wissensgebiet.

Die formativen Leistungstests der Treatmentgruppe beinhalteten hingegen Fragen zu allen Lerninhalten der Unterrichtseinheit und ausschließlich, den von mir entwickelten Aufgabentyp und wurden gleich nach der Beantwortung von den Schülern selbst mit Hilfe des Kontrollbogens korrigiert.

Vorstudie 1:

Die erste Vorstudie fand in einer M7 an einer Mittelschule in Fürth statt. Dort wurde die Unterrichtseinheit (teilweise auch von mir) gehalten und im Anschluss gemeinsam mit der unterrichteten Lehrkraft besprochen und überarbeitet. Auch die Tests wurden alle, wie geplant, durchgeführt und danach ausgewertet. Als Korrekturhilfe der FLTs kam bei dieser Vorstudie die oben beschriebene Kontrollfolie zum Einsatz.

Folgender Verlaufsplan lag diesem ersten Durchgang zu Grunde:

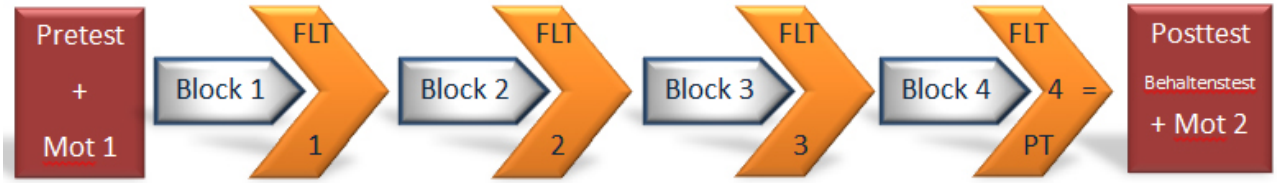


Abbildung 5: Verlaufsplan Vorstudie

Angedacht waren vorerst 3 Unterrichtsblöcke mit jeweils anschließenden formativen Leistungstests (FLT 1 – 3). Der Unterrichtsblock 4 beinhaltete nur eine Stunde zum Vogelschutz, der als Puffer dienen sollte, damit die unterrichtende Lehrkraft vor dem eigentlichen Test (FLT4 = Posttest) nicht in Versuchung geraten konnte, mit der gesamten Klasse Lerninhalte zu wiederholen und somit mögliche Effekte zu zerstören. Der Pre- und Posttest waren identisch und beinhalteten, anders als die formativen Leistungstests (spezielles Aufgabenformat – siehe oben), ähnlich wie bei herkömmlichen Leistungstests, offene und halboffene Fragen. Nach einer Zeit von ca. 6 bis 8 Wochen sollte der Posttest wiederholt werden, um die Behaltensleistung der einzelnen Gruppen zu überprüfen. Zu den Testzeitpunkten war die Klasse, wie bereits beschrieben in eine Treatmentgruppe und eine Kontrollgruppe unterteilt. Zunächst ging es vor allem darum, den Schwierigkeitsgrad der Aufgaben festzustellen, die Formulierungen der Aufgabentexte zu überarbeiten, den Umgang mit den Selbstkontrollbögen zu beobachten und die Schwächen in der Unterrichtseinheit ausfindig zu machen und diese zu verbessern.

Ergebnisse der ersten Vorstudie:

Der erste Entwurf der **Unterrichtseinheit** war etwas ausführlicher, als der, der letztendlich bei der eigentlichen Studie eingesetzt wurde und beinhaltete folgende zusätzliche Punkte, die vor allem aus Zeitgründen und aus Gründen der inhaltlichen und curricularen Passung gestrichen wurden:

- Vogelbestimmung
- Flugunfähige Vögel
- Flugfähige Wirbeltiere
- Bionik
- Vogelschutz (Pufferstunde vor dem Posttest)

Dafür wurde die Sequenz um den Punkt „Evolutionäre Anpassung“ erweitert, weil dieser Inhalt eine zentrale Rolle in der Biologie darstellt, Teil eines größeren Konstrukts ist (Evolution – Selektion, Mutation von Genen, ökologische Nische, Vererbung, etc.) und dadurch in vielen Jahrgangsstufen in irgendeiner Form (siehe 5.2.2) wieder aufgegriffen wird. Neben dem Begriff des Auftriebs (siehe Concept Map - Abbildung 21), stellt die evolutionäre Anpassung also eine zweite Möglichkeit dar, konzeptuelles Wissen in die Unterrichtseinheit einzubinden. Interessant ist dies vor allem, weil es bei diesem Thema sehr viele falsche Vorstellungen bei den Schülern gibt, die sich zum Teil, bis zum Studium halten (genauere Ausführungen hierzu finden sich unter 2.3.1.1) und die mit Hilfe des Unterrichts reduziert werden sollen.
Gemeinsam mit der Lehrkraft aus der Vorstudie wurde die Unterrichtseinheit gekürzt und Schwachstellen ausgebessert. Die neu eingebundene Sequenz über den allgemeinen Anpassungsbegriff und die dazugehörigen Testfragen mit Rückmeldungen, wurden im Anschluss an die erste Vorstudie, sowohl im zweiten Durchlauf an der zweiten M7 in Fürth, als auch gesondert in zwei iPad Klassen (7. Jahrgangsstufe) einer Realschule in Erlangen getestet.

Danach ergaben sich einige Änderungen:

Einzelne **Fragen** wurden, weil sie zu einfach und somit nicht aussagekräftig waren, gestrichen. Einige zu schwere Fragen, umformuliert.

Dadurch, dass sich sehr schnell herausgestellt hat, dass die Arbeit mit dem **Selbstkontrollbogen** nicht wirklich gut funktioniert (beim Durchsehen der Schülerordner habe ich einige Fehler bei der Korrektur durch die Schüler gefunden), und diese die wichtigste Aufgabe, individuelles Feedback zu den Schülerantworten zu geben, nicht leisten konnte, gab es an diesem Punkt die weitreichendste Veränderung im Projekt:
Alle Paper-/Pencil- Tests wurden auf **elektronische Leistungstests** umgestellt, was große Veränderungen im Testdesign, in der Wahl der Aufgabentypen und in der Formulierung der Rückmeldungen nach sich zog. Auch musste überlegt werden, wie die Testergebnisse der einzelnen Schüler gespeichert (Lernplattform) werden könnten und welche Endgeräte für die Bereitstellung der elektronischen Leistungstests am besten geeignet wären, ohne das Unterrichtsgeschehen in großem Maße zu stören (zum Beispiel durch einen Wechsel in den Computerraum,

etc.). Natürlich wäre es auch nicht einfach, sich in das Schulnetz (falls vorhanden) aller Testschulen einzuwählen. Wie die genaue Umsetzung der elektronischen Lösung aussieht, kann man in Kapitel 5.4 genauer nachlesen. Dort findet man auch Überlegungen, welche Software für die Erstellung der Testfragen in Frage kommt.

Ein letzter Punkt, der mit der Kürzung der Unterrichtseinheit einher ging, war die **Reduzierung der Testzeitpunkte** – es stellte sich heraus, dass 3 Erhebungszeitpunkte sowohl in zeitlicher, als auch motivationaler Hinsicht zu viel sind. Daher wurde die neue Unterrichtseinheit in zwei Lernblöcke und eine Pufferstunde untergliedert (siehe 5.2) und es wurden nur noch 2 formative Leistungstests eingeplant (siehe 5.3.2.3).

Viele der Materialien, die für die Durchführung der ersten Vorstudie entwickelt wurden, können bei der PH Schwäbisch Gmünd angefordert werden, dazu gehören:

- Alle schriftlichen Leistungstests (Pre-/Posttest und die formativen Leistungstests 1,2 und 3 mit den dazugehörigen Kontrollfolien).
- Die Motivationstests 1 und 2 in ihrer ursprünglichen Form.

Die zusätzlichen Lerninhalte, die nach der Vorstudie gestrichen wurden, um die Gesamtzeit der Studie zu verkürzen, die Testzeitpunkte zu verringern und die Unterrichtssequenz inhaltlich besser abzurunden (neuer Schwerpunkt: „evolutionäre Anpassung“), sind nicht extra mit aufgeführt.

<u>Erprobung der Technik und der ersten elektronischen Testfragen:</u>

Nach einer ausführlichen Suche nach Möglichkeiten der technischen Umsetzung (die genaue Beschreibung, der Recherche und der Überlegungen zum Thema elektronische Leistungstestes befindet sich unter Punkt 5.4), wurde Moodle als Lernplattform zur Speicherung der Schülerdaten und auch als Werkzeug für die Erstellung der elektronischen Testfragen herangezogen.
In einem zweiten Schritt musste elektronisches Bild- und Filmmaterial zum Thema der Unterrichtssequenz zusammengetragen und in Moodle mit Hilfe der vorliegenden Fragensammlung aus der Vorstudie und den bereits entwickelten Paper-Pencil Tests ein großer Fragepool angelegt werden, aus dem die unterschiedlichen Leistungstests der Studie zusammengestellt wer-

den konnten. Genaueres zu den ausgewählten Testfragen, den verwendeten Aufgabentypen und dem Testdesign findet man unter 5.3.

Bevor die zweiten Vorstudie in der Parallelklasse der M7 an der Mittelschule in Fürth durchgeführt werden konnte, wurde zuerst überprüft, ob das technische Konzept (gewählte Hard- und Softwarezusammenstellung) beim Einsatz in der Klasse funktionierte und inwieweit die Fragen zu den neu in die Unterrichtseinheit aufgenommenen Lerninhalten der „Evolutionären Anpassung“ geeignet sind. Diese Überprüfung fand in zwei iPad Klassen einer Realschule in Erlangen statt. Dies war ein großer Vorteil, denn zu diesem Zeitpunkt der Studie waren noch keine Endgeräte (Tablet PCs) vorhanden. Zwei Biologielehrkräfte waren bereit eine kurze Lerneinheit über die evolutionäre Anpassung und den darauffolgenden Lernzirkel zum Thema „Anpassung der Vögel ans Fliegen“ zu halten und die dazugehörigen Tests in ihren Klassen durchführen zu lassen.

Der erste Versuch, den Pretest mit einer iPad Klasse durchzuführen, schlug aufgrund technischer Probleme fehl: Es war egal, ob 20, oder nur 2 Schüler versuchten, auf den Server (Laptop der Versuchsleiterin) zuzugreifen - das ganze System brach sofort zusammen. Die darauffolgende Fehlersuche, die vor allem im Installieren und De-installieren der Serversoftware und der dazugehörenden Datenbank und der Veränderung der Einstellungen bestand, zeigte, dass die von Moodle selbst empfohlene Kombination von Server und hinterlegter Datenbank nicht funktionierte. Erst, als der wamp Server durch den xamp Server ersetzt wurde, lief das System fehlerfrei und stabil und konnte im Unterricht eingesetzt werden.

Der weitere Einsatz in den iPad Klassen lief ohne weitere Probleme und so konnten am Ende dieses Zwischentests, die neu entwickelten Testfragen in der zweiten Vorstudie auf ihre Tauglichkeit hin untersucht und gegebenenfalls anpasst werden.

Auch die Motivationstests, die als Formulare im Feedback Modul von Moodle erstellt wurden, funktionierten einwandfrei. So war eine komplette technische Erfassung und Auswertung aller in der Studie entstehenden Daten möglich.

<u>Vorstudie 2</u>

Wie bereits erwähnt, fand die zweite Vorstudie ebenfalls an der Mittelschule in Fürth statt, wo bereits die erste Vorstudie lief. Bei der gleichen Lehrkraft und in der Parallelklasse der M7.

Hier war es möglich einen Durchlauf der gesamten Unterrichtseinheit und aller dazugehörigen Tests durchzuführen, die Diagnoseaufgaben nochmal genauer zu testen, etwas Routine in das Handling der Technik zu bekommen und den Umgang der Schüler mit dem erteilten Feedback zu beobachten. Für die zweiten Vorstudie wurde der zweiten Motivationstest geändert – einige Items wurden entfernt, dafür wurden einige selbstentwickelte Items zur Bewertung der Unterrichtseinheit, zum Tablethandling und zur Feedbacknutzung hinzugefügt, um eine Rückmeldung der Schüler zu den Themen „Technik“ und „Feedback“ zu erhalten. Die genaue Beschreibung der beiden Motivationstests und der Kurzskalen findet man unter 5.3.2.5.

Ergebnisse der zweiten Vorstudie:

Durch die Auswertung, der Fehler, die bei der Beantwortung der Diagnoseaufgaben gemacht wurden, konnten zu schwere Fragetexte isoliert und umformuliert werden.

Man konnte feststellen, dass es einige Dinge gab, auf die alle Schüler hingewiesen werden müssen, bevor die Arbeit mit den Tablets und den elektronischen Tests beginnen kann. Diese Punkte wurden daraufhin auf einer Folie, als eine Art Anleitung festgehalten, die bei der Vorbesprechung mit den Schülern während zu Beginn der Erhebung herangezogen werden konnte:

- Es ist nicht erlaubt, etwas an den Einstellungen zu verändern
- Die Tablets sollte man im Querformat halten, sonst werden manche Antworten nicht ganz angezeigt (dieses Problem ist von Tablet zu Tablet unterschiedlich stark ausgeprägt: Samsung schneidet teilweise selbst im Querformat ab, Lenovo nur im Hochformat und Apple gar nicht)
- Wenn es mehrere Antwortmöglichkeiten in Drop Down Menüs gibt, muss man oft scrollen, um alle angezeigt zu bekommen (Problem nur bei Samsung und Lenovo, also den kleinen Tablets)
- Man muss erklären, dass es einen Unterschied zwischen Single- und Multiple Choice Aufgaben gibt (Um welchen Aufgabentyp es sich handelt, wird durch die Form des Icons angezeigt: rund = eine Antwort ist richtig, eckig: es gibt mehrere richtige Lösungen)
- Ganz wichtig: Die Schüler müssen nach der Beantwortung jeder Frage, erst ihr Ergebnis „prüfen“, damit die Punkte in der Datenbank abgespeichert werden. Punkte nicht geprüfter Aufgaben gehen verloren.

Zusätzlich zu den Eigenheiten der Endgeräte, gab es während der zweiten Vorstudie kleine technische Aussetzer:

So wurden bei einem User die Bilder in den Textfragen und den Rückmeldescreens nicht angezeigt – das Problem konnte leider nicht gelöst werden, weder durch Veränderungen im Nutzerprofil, in dem es soundso keinen Unterschied zu den anderen Nutzern gab, noch in den Einstellungen des Tablets. Auch in der Lernplattform selbst gab es keine Möglichkeit, externe Medien zu- oder abzuschalten. Erst durch die Nutzung eines neuen Nutzprofils war dieses Problem gelöst.

Teilweise war der Text für das kleine Display zu lang. Trotz der Optimierung der Darstellung für das Endgerät, innerhalb von Moodle, wurde teilweise Text abgeschnitten, so dass eine Umformulierung einzelner Aufgabentexte nötig wurde.
Es stellte sich heraus, dass es sinnvoll ist, immer nur die Tests auf der Kursseite (Startseite) von Moodle anzeigen zu lassen, der gerade benötigt wird, sonst verwirrt das die Schüler.

Weiterhin wurde in der zweiten Vorstudie überprüft, wann es am sinnvollsten ist, das Feedback anzuzeigen und in welcher Form es am effektivsten für den Schüler ist. Über den Einsatz und das Design der Rückmeldungen wird ausführlich in Kapitel 5.3.3 berichtet.

4.2.3 Vorbereitung der Lehrkräfte auf das Feldexperiment

Als Teilnehmer für das Feldexperiment sollten erfahrene Lehrkräfte gewonnen werden (keine Referendare oder Praktikanten), damit grundlegende Schwierigkeiten im Umgang mit Schülern nicht zu verzerrenden Effekten führen konnten. Dafür wurde eine Ausschreibung zu einer Fort- und Weiterbildung für Lehrer an Mittelschulen im Fach PCB ausgeschrieben. Darüber hinaus wurden über persönliche und berufliche Kontakte, Lehrkräfte aus unterschiedlichen Realschulen und einem Gymnasium rekrutiert. Um diese Lehrkräfte zu gewinnen und über das Projekt zu informieren, wurde ein Informationsflyer entworfen, um an den Schulen Werbung machen zu können. Lediglich die Lehrkräfte aus den Mittelschulen, wurden in einem Workshop auf das Projekt vorbereitet, die zusätzlich gewonnen Lehrkräfte aus den Realschulen und dem Gymnasium, wurden in einem persönlichen Gespräch auf das Feldexperiment vorbereitet.

Der Workshop fand am Ende des Schuljahres 2011/2012 statt, damit es den teilnehmenden Lehrkräften möglich war, gemeinsam mit ihrer Schulleitung das folgende Schuljahr zwecks Klassen- und Fächerzuteilung, für die Studie zu planen.

Themen des Workshops waren:

- Kurze, generelle Einführung in die FLM-Theorie sowie fachdidaktische und motivationstheoretische Grundlagen der Unterrichtseinheit.
- Informationen über den Verlauf des geplanten Projekts (Organisatorisches, Arbeitsplan).
- Vorstellung des Unterrichtsmaterials und dessen Aufbau: Alle Lehrkräfte erhielten eine CD Rom, die alle Lerninhalte der Unterrichtssequenz (als PDF und als Word – Datei, damit die Stunden noch individuell an die jeweiligen Klassen angepasst, bzw. nur Teile der Materialien übernommen werden konnten), sowie einen Ablaufplan der Tests und eine Materialliste beinhaltete.

  Alle erstellten Unterrichtsmaterialien sind folgendermaßen aufgebaut: Zu Beginn jedes Themas findet sich immer ein ausführlicher fachtheoretischer Hintergrund, dann kommt ein Stundenvorschlag und danach alle Materialien, die man in der Stunde benötigt (Arbeitsblätter, Experimentierhefte, Folien, Lösungen für die Lehrkraft, Übungsaufgaben, etc.). Die Lehrkräfte bekamen als Vorgabe, dass lediglich die Lernziele der Unterrichtseinheiten hinreichend bei den Schülern ankommen müssen, wie sie das methodisch umsetzen, wurde hingegen nicht vorgegeben. So hätten sie den vorgeschlagenen Lernzirkel (ausschließlich Schülerexperimente in Gruppenarbeit) auch als Lehrerexperimente (vorführend), oder auch im Frontalunterricht (nur inhaltlich darbietend) durchführen können.
- Demonstration der Technik, bei der die Lehrkräfte auf mitgebrachten Tablets oder den eigenen Smartphones, die Tests probeweise selbst bearbeiten konnten, um dadurch einen kleinen Einblick in die Aufgaben zu bekommen.

Am Anfang des Schuljahres 2012/2013 setzte ich mich mit jeder Lehrkraft in Verbindung, um abzufragen, welche Klassen mit wie vielen Schülern (Anzahl der benötigten Logins) sie erhalten hatten und wann sie mit dem Projekt beginnen wollten. Bei diesem Gespräch (und während die Schüler die Pretests bearbeiteten) hatten die Lehrkräfte auch die Möglichkeit noch Fragen zu

den Unterrichtsmaterialien zu stellen. Es wurde versucht, an den verschiedenen Schulen zu unterschiedlichen Zeitpunkten zu starten, damit man die Versuchsmaterialien für den Lernzirkel jeder Lehrkraft zur Verfügung stellen konnte, ohne mehrere Experimentierkisten zu füllen - dies hätte sich etwas schwieriger gestaltet, weil manche Versuchsmaterialien (Fell, präparierte und haltbar gemachte Straußen- und Säugetier- und Vogelknochen), nicht so einfach zu beschaffen waren.

Während der Evaluationsphase war ich ständig für die Lehrkräfte erreichbar und mit ihnen in Kontakt. Nach Beendigung der Evaluation und nach der Auswertung der Daten, wurden die Lehrkräfte zu einem abschließenden Gespräch eingeladen, in dem die Ergebnisse vorgestellt wurden.

4.2.4 Ablauf des Feldexperiments

11 Klassen mit insgesamt 288 Schülern und 9 Lehrkräften aus Mittelschulen, Realschulen und einem Gymnasium beteiligten sich zu Beginn an der Feldstudie. Leider schied eine Realschulklasse mit 27 Schülern nach dem Pretest wegen unterschiedlicher anderer Projekte und dem damit einhergehenden zeitlichen Problemen aus, so dass am Ende noch 261 Schüler übrig blieben, die an der gesamten Studie teilnahmen. Genaueres zur Stichprobengröße und deren Eigenschaften kann man unter dem Punkt 6.2.1 nachlesen.

Das Feldexperiment fand während des gesamten Schuljahres 2012/2013 und ausschließlich in den Jahrgangsstufen 6 und 7 statt. Es wurde darauf geachtet, dass die Klassen gestaffelt mit der Unterrichtseinheit begannen, damit die Unterrichtsmaterialien (Experimentierkiste für den Lernzirkel, Federn für das Mikroskopieren) allen Klassen zur Verfügung stehen konnten. Außerdem wurde so gut es ging auf die zeitlichen Wünsche der Lehrkräfte und deren Stoffverteilungspläne eingegangen. Der zeitliche Ablauf gestaltete sich problemlos und es gab keine organisatorischen Probleme, weil sich alle teilnehmenden Lehrer voll ins Projekt eingebracht haben. Sogar der Versand der Experimentierkiste an die nächste Schule wurde bereitwillig von den Lehrern übernommen, wenn es zu zeitlichen Engpässen kam.

Bevor die Studie starten konnte, mussten sehr viele Aufgaben erledigt werden, die an verschiedenen Punkten dieser Arbeit näher erläutert werden:

Die Erstellung der Unterrichtseinheit (Punkt 5.2) und der Diagnoseaufgaben (Kapitel 5.3), sowie die Zusammenstellung der Tests (5.3.2) und des dazugehörigen Feedbacks (5.3.3). Außerdem

musste die technische Umsetzung geklärt werden: Welche Lernplattform bietet sich an und mit welchem Autorensystem können die Testaufgaben am effektivsten erstellt werden? Welche Endgeräte wählt man zur Bearbeitung der Tests? Welche Logins erstellt man am besten für die einzelnen Schüler, um jederzeit die jeweilige Schule und Klasse daraus ablesen zu können? Wie umgeht man das Problem, das manche Schulen kein WLAN haben oder wenn doch, man das Schulnetz nicht nutzen darf? Diese und weitere technische Details werden unter Punkt 5.4 näher diskutiert.

Organisation und Ablauf

Wie bereits in 4.2.2 beschrieben, gab es während der Vorstudien Änderungen bei den Inhalten der Unterrichtseinheit und der Anzahl der Testzeitpunkte, so dass der Erhebung in den einzelnen Klassen letztendlich folgender Ablauf zugrunde lag:

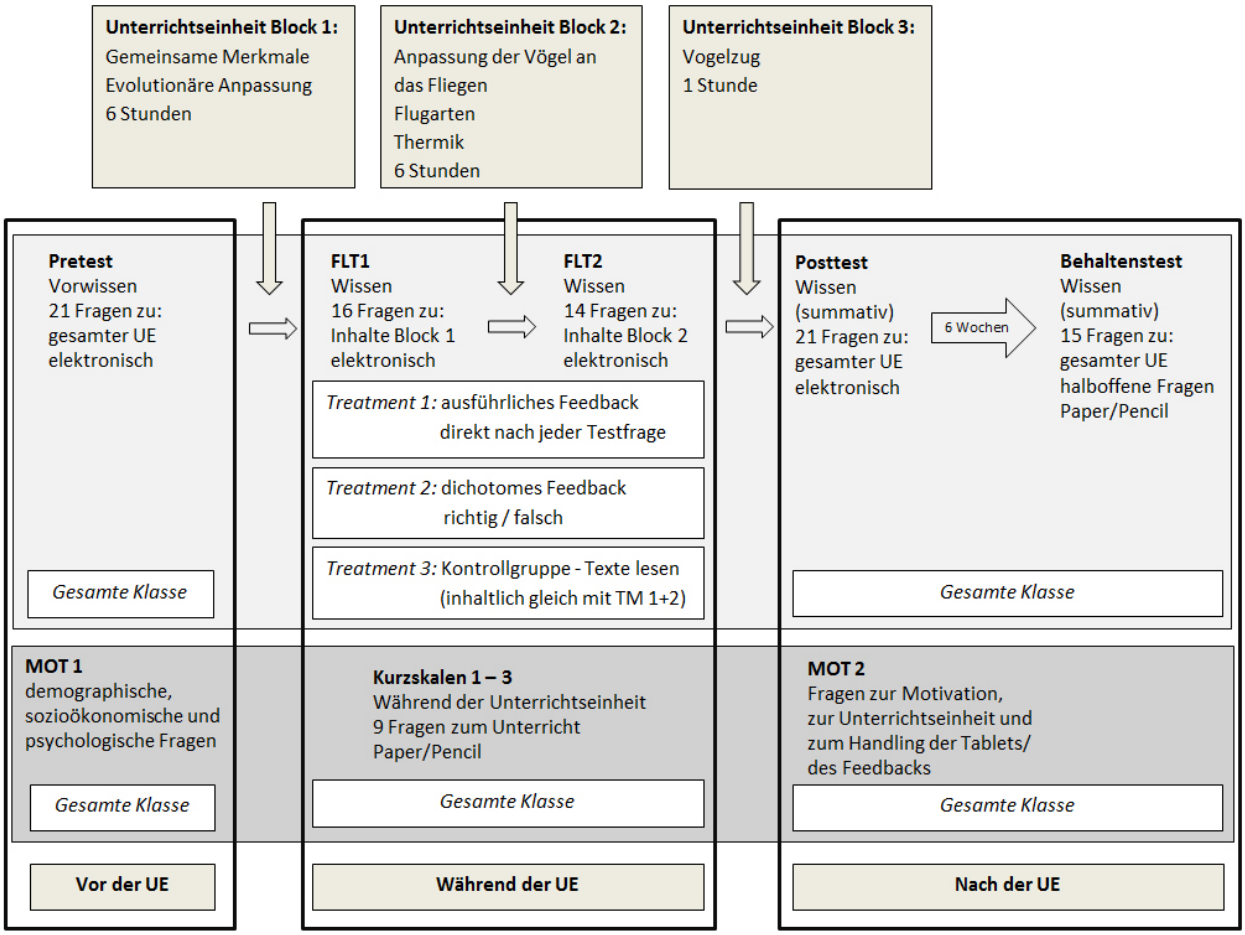


Abbildung 6: Ablauf der randomisierten Feldstudie

Die Erhebung begann immer mit einer Doppelstunde, in der den Schülern das Projekt erläutert und ihnen kurz erklärt wurde, was in den nächsten Wochen auf sie zukommt. Die Schüler erhielten ihren Benutzernamen und ihr Passwort, um sich in Moodle einloggen zu können und eine Einführung in die Handhabung der Tablets und die Bearbeitung der elektronischen Tests. Im weiteren Verlauf der Stunde wurde mit der Klasse der Pre- und der Motivationstest 1 durchgeführt.

Während die Schüler die Tests bearbeiteten gab es viele Dinge zu erledigen und zu beachten:

- Klärung technischer und inhaltlicher Fragen,
- Unterstützung bei der Abgabe des Tests (sonst sind die Ergebnisse und die Bearbeitungszeit nicht in der Datenbank erfasst),
- Hilfe beim Öffnen des Motivationstests (jedem Schüler wurde noch einmal kurz erläutert, auf was sie beim Ausfüllen des Formulars achten mussten),
- Speichern der Einträge des Motivationstests - diese müssen am Ende gesondert abgespeichert werden, sonst gehen sie verloren,
- Kontrolle, ob sich alle Schüler richtig aus Moodle ausloggen und
- Abschalten der Tablets (um Energie zu sparen).

Abbildung 7: Schülerinnen bei der Arbeit mit den Tablets

In der darauffolgenden Biologiestunde startete die jeweilige Lehrkraft mit ihrer Klasse direkt mit der Unterrichtseinheit. Während der Unterrichtsblöcke, wurden an drei Zeitpunkten von den Schülern Kurzskalen zur Motivation ausgefüllt.

Wann die Kurzskalen und die anderen Tests genau stattfinden, kann man folgender Tabelle entnehmen:

Tabelle 14: Überblick über die Lerninhalte und Messzeitpunkte der Studie

Thema	***Stundenzahl***	***Tests***
	1,5	*Pretest und Motivationstest 1*
Gemeinsame Merkmale Schnabel- / Fußform	*1-2 (bei Durchführung der Versuche)*	
Gemeinsame Merkmale Federn	*1-2 (beim Mikroskopieren)*	*Nach der Stunde: Kurzskala 1*
Specht als Beispiel für Anpassung	*1*	
Allgem. Anpassungsbegriff	*2*	
Block 1 (gesamt)	***5 - 7***	
	1	*FLT 1*
Lernzirkel	*2 Stunden Lernzirkel + 1 Stunde Besprechung*	*Nach der Stunde: Kurzskala 2*
Arten des Vogelflugs Thermik	*2*	*Nach der Stunde: Kurzskala 3*
Block 2 (gesamt)	***5***	
	1	*FLT 2*
Vogelzug	*1*	
Block 3 (gesamt)	***1***	
	1,5	*Posttest und Motivationstest 2*
Unterrichtseinheit gesamt	***11****-13*	
Stunden Einführung + TESTS	***5***	
Stunden (gesamte Erhebung)	**16 - 18**	

Die Kurzskalen sind immer gleich und beinhalten Fragen zur vorangegangenen Stunde, die die Schüler auf einer 5-stufigen Skala eingruppieren sollen (genaueres zu den Kurzskalen findet man unter 5.3.2.5).

Die FLTs, die nach den einzelnen Unterrichtsblöcken geschrieben werden beziehen sich direkt auf die Lerninhalte des vorhergehenden Unterrichtsblockes. Für die FLTs werden die Lernenden innerhalb einer Klasse per Zufall (Durchmischung der Logins und der damit einhergehenden Gruppenzugehörigkeit) den unterschiedlichen Treatmentgruppen zugeordnet. Dadurch gehört ungefähr ein Drittel aller Teilnehmer je einer der 3 Treatmentgruppen an.

Durch die zufällige Zuordnung sollten sich die Schüler zwischen den Versuchsbedingungen hinsichtlich leistungsrelevanter Prädiktoren nicht unterscheiden.
Dennoch soll diese Voraussetzung durch die im Motivationstest erhobenen Variablen (Vorwissen, Motivation, Interesse), die Schulnoten und die Geschlechterverteilung zusätzlich geprüft werden. Notfalls werden stichprobenverzerrende Faktoren als Kovariaten berücksichtigt.

Die Rückmeldungen der unterschiedlichen Gruppen sehen folgendermaßen aus:

Treatment 1: Diese Gruppe schreibt einen Test und erhält direkt nach der Beantwortung der Testfragen ein ausführliches Feedback. Darin wird erläutert, was der Schüler korrekt und was er falsch geantwortet hat, und zusätzlich eine detaillierte Rückmeldung zu ihrer Antwort.
Treatment 2: Schreibt den gleichen Tests, wie die erste Treatmentgruppe, erhält aber lediglich eine kurze Rückmeldung, welche der Antworten falsch und welche richtig war.
Treatment 3 = Kontrollgruppe: Behandelt inhaltlich die gleichen Themen, wie die Treatmentgruppen 1 und 2, indem sie ausführliche Texte zu den Lerninhalten lesen. Dabei haben die Schüler dieser Gruppe genauso viel Zeit, sich mit den Inhalten zu beschäftigen, wie die anderen beiden Treatmentgruppen.

Der Pre- und der Posttest hingegen, beinhalten Fragen zur gesamten Unterrichtssequenz und sollen den Wissenstand zum Thema an zwei verschiedenen Zeitpunkten festhalten (einmal zu Beginn der Studie zur Erfassung des Vorwissens und einmal am Ende, um den Lernzuwachs festzustellen) und dokumentieren.
Der Block 3 dient als Pufferstunde, damit die Lehrkraft vor dem Posttest nicht in Versuchung gerät, mit der gesamten Klasse alles Gelernte zu wiederholen und dadurch mögliche Effekte zu zerstören.
Der Behaltenstest findet ungefähr 6-8 Wochen nach der Evaluation statt und soll mit Hilfe halboffener und offener Fragen überprüfen, inwieweit die Schüler der unterschiedlichen Gruppen die Lerninhalte behalten und durchdrungen haben.
Unter 5.3.2 findet man alles über den Aufbau und die Inhalte der verschiedenen Tests, sowie deren Aufgabenformate. Sämtliche Datenerhebungen (Vor- und Nachtests, Interessen- und

Motivationstests) wurden von mir durchgeführt, um die Lehrkräfte nicht zusätzlich zu belasten und die Versuchsbedingungen konstant zu halten.

Was muss vor und nach jedem Testzeitpunkt bedacht werden?

Direkt vor jedem Testzeitpunkt sind folgende Aufgaben zu erledigen:

- Zu Hause: Aufladen und Reinigen der Tablets: Es wurden immer ca. 6 Tablets gleichzeitig an einer Mehrfachsteckdose geladen – mehr sollte man nicht an das Hausnetz anschließen, weil das zu Problemen führen kann. Die Touchscreens waren nach jeder Nutzung durch die Schüler ziemlich verschmiert, so dass sie mit Desinfektionsspray und einem weichen Tuch gereinigt werden mussten.

Abbildung 8: Wartung der Tablet PCs

- Die Moodleeinstellungen mussten für den nächsten Testzeitpunkt verändert werden: Die benötigten Tests mussten sichtbar, andere unsichtbar gemacht werden, denn eine zu große Auswahl an Tests hätte die Schüler nur verwirrt.
- In der Schule: Einführung der Schüler in die Handhabung der Tests (es gab Unterschiede bei der Bearbeitung der verschiedenen Testformate, auf die die Schüler gesondert hingewiesen werden mussten). Eventuell musste bei einzelnen Schülern das WLAN wieder ein-, bzw. der Ton wieder ausgeschaltet werden, weil irgendwer doch wieder einzelne Einstellungen des Tablets verändert hatte. Außerdem musste man dafür sorgen, dass zwei Steckdosen im Klassenzimmer vorhanden waren, die für den Rechner und den Access – Point nötig waren.

Vor dem ersten Testzeitpunkt mussten zusätzlich folgende Schritte durchgeführt werden:

- Anlegen der neuen Kursteilnehmer in Moodle (nach einem technischen Ausfall bei einem User, wurden immer pro Klasse noch zwei Ersatzteilnehmer angelegt, falls es bei einem Login wieder zu Problemen kommt. Nach der Erhebung, wenn alles gut gelaufen war, wurden diese wieder aus Moodle gelöscht).

- Ausdrucken der Login Daten für die Schüler auf kleine Kärtchen, Zuordnung der Teilnehmer zu den einzelnen Treatmentgruppen, bzw. zur Kontrollgruppe. All diese Informationen (Logins, Gruppenzugehörigkeit, etc.) wurden gesondert in einer Exceldatei festgehalten, weil häufig einzelne Schüler ihre Logins verloren haben und deren Gruppenzugehörigkeit geklärt werden musste. Dies wiederum wurde nur möglich durch das
- Anlegen einer Login – Liste durch die unterrichtende Lehrkraft: Bei der Durchführung der Studie in der ersten teilnehmenden Klasse hat sich herausgestellt, dass die Schüler ihre Logins sehr oft vergessen oder verlieren, obwohl sie darauf hingewiesen wurden, wie wichtig diese Informationen sind. Eigentlich sollten sie ihre Daten in ein Heft übertragen, das sie immer bei sich haben, oder den Zettel ins Hausaufgabenheft kleben. Irgendwie schafften es immer welche, die Logins nicht dabei zu haben. Darum, und weil die Lehrkraft zur Benotung des Posttests (dieser wurde immer als Probearbeit gewertet) die Zuordnung der Logins zu den Schülern soundso brauchte, wurde in den darauffolgenden Durchgängen immer zu Beginn eine Klassenliste anlegen, auf der die Schüler ihre Logins eintragen mussten. Diese Informationen verblieben bei der Lehrkraft, so dass die Anonymität der Kursteilnehmer weiterhin gewahrt wurde. Dieses Vorgehen erleichterte Vieles und verhinderte viele fehlende Daten.
- Einführung der Lerner in den Umgang mit den Tablets (Querformat; keine Änderungen in den Einstellungen; Möglicherweise: Scrollen bei den Auswahlmenüs, weil sonst nicht alle Auswahlmöglichkeiten angezeigt werden, etc.)
- Erklärung, was bei der Bearbeitung der Tests zu beachten ist (Unterscheidung von Multiple- und Single- Choice Aufgaben; nach der Beantwortung jeder Frage bei den FLTs erst „Prüfen“ drücken, damit die Ergebnisse auf dem Laptop gesichert und die Antwort auf ihre Richtigkeit hin überprüft wird, dann erst auf „Weiter“ klicken; Springen zwischen den einzelnen Fragen ist in der Navigation links oben möglich; bei den Motivationstests am Ende unbedingt „Einträge speichern“ drücken, um das Formular zu sichern).

Nach den Testzeitpunkten:

- Wurden alle neuen Daten gesichert: Moodle bietet die Möglichkeit, alle Daten in Excellisten zu exportieren. Diese Option wurde nach jedem Erhebungszeitpunkt zur Datensicherung genutzt. Um die Daten zusätzlich zu schützen, wurden sie auf einer externen

Festplatte gespeichert, da der Computer, der als Server fungiert keine Firewall haben darf, damit man von außen auf ihn zugreifen kann.

- Mussten die Daten der Motivationstests per Copy and Paste in eine Excelliste übertragen werden – leider gab es in Moodle keine Funktion, um diesen Arbeitsschritt zu automatisieren. Auch die Ergebnisse der Kurzskalen musste per Hand in eine elektronische Form gebracht werden.
- Wurden die zuvor angelegten Ersatzteilnehmer wieder aus Moodle entfernt.
- Musste man immer in den aufgezeichneten Testergebnissen in Moodle kontrollieren, ob alle Schüler ihren Test auch abgegeben hatten (nur dann wurde die Bearbeitungszeit gestoppt), sich die Schüler ausgeloggt hatten, um die Sitzung zu beenden und ob alle Tablets ausgeschaltet waren, um nicht unnötig Energie zu verschwenden.
- Wurde überprüft, welche Schüler an den jeweiligen Tests an- bzw. abwesend waren und entsprechend eigentragen, um einen späteren Abgleich bei der Endauswertung zu ermöglichen.

Verlauf der Testzeitpunkte

Schnell hat sich herausgestellt, dass es immer einige Zeit dauert, bis sich die Tablets mit dem WLAN des Laptops verbunden haben, daher lies ich nach den ersten Tests die Tablets immer zuerst austeilen und anschalten, bevor ich mit den Erläuterungen zum folgenden Test begann.

Leider gab es bei den vorbereitenden Schritten immer wieder folgende Probleme:

- Die Samsung Tablets haben sich gar nicht von selbst mit dem WLAN verbunden: Man musste das WLAN jedes Mal manuell trennen und wieder verbinden, bevor irgendetwas funktionierte.
- Die iPad Klassen mussten das WLAN immer manuell anwählen, um sich mit dem Server zu verbinden und es stellte sich heraus, dass man Cookies generell erlauben musste, damit die Schüler mit dem iPad auf der Moodleseite arbeiten konnten. Außerdem konnte man bei den iPads Moodle nicht als Startseite einstellen, so dass die Schüler durch Eingabe der Adresse „192.168.0.124/moodle“ die entsprechende Internetseite aufrufen mussten und dabei haben sich die Schüler regelmäßig vertippt, so dass erst mal nichts funktionierte.

- Beim Einloggen verwechselten Schüler teilweise in ihrer Benutzerkennung das „O“ mit der „0“ (Null), oder missachteten beim Passwort die Groß- und Kleinschreibung.
- Nach dem Einloggen zeigte Moodle unerklärlicherweise die Startseite in zwei unterschiedlichen Variationen an, deshalb mussten beim Start der Tests immer zwei unterschiedliche Vorgehensweisen erläutern werden.

Dadurch dauerte es immer einige Zeit, bis jeder mit der Bearbeitung seines Tests beschäftigt war. Da die Studie nicht unter Laborbedingungen stattfand, sondern im ganz normalen Unterrichtsverlauf eingebettet war, gab es während der Testphasen immer wieder kleine Ablenkungen und Unterbrechungen.

5 Evaluationsdesign / Methodik

Die Stichprobe umfasste 261 Schüler (6. und 7. Jahrgangsstufe) aus 10 Klassen bayerischer Mittel- und Realschulen, sowie einem Gymnasium. Biologielehrkräfte unterrichteten eine vom Zeitumfang und Inhalt festgelegte Unterrichtseinheit zum Thema Vogelflug. In einem Pretest mit 21 Items wurde das Vorwissen der Schüler zu allen Einzelthemen der Unterrichtseinheit erhoben und die gleichen 21 Fragen wurden im Posttest wieder aufgegriffen, um den Lernzuwachs überprüfen zu können. Für die formativen Leistungstests (beinhalten Testfragen zum vorhergehenden Unterricht) wurden die Klassen in 3 Gruppen unterteilt und die Schüler erhielten über Tablet-PCs, in die Lernplattform „moodle“ eingebettete diagnostische Aufgaben, bzw. Texte zur vorher behandelten Thematik. Die Effekte des Treatments auf Fachwissen, Interesse und Motivation wurden mit elektronischen Vor- und Nachtests geprüft. Ein Behaltenstest in Schriftform (6 Wochen nach Abschluss der Einheit) sollte durch die Verwendung halboffener Fragen zeigen, wie gut die Lerninhalte und insbesondere die Konzepte „Auftrieb“ und „Evolutionäre Anpassung“ von den drei Schülergruppen verinnerlicht wurden.

Bei der Evaluation handelte es sich um eine randomisierte Feldstudie.

Randomisiert deshalb, weil die Zuordnung der Schüler zu den einzelnen Versuchsgruppen nach dem Zufallsprinzip erfolgte. Durch die Randomisierung wurde eine gleichmäßige Verteilung von bekannten und nicht bekannten Einflussfaktoren auf alle Gruppen sichergestellt.

Dadurch, dass sich alle Versuchsgruppen in der gleichen Klasse befanden, kann eine unterschiedliche Beeinflussung der Gruppen und deren Testergebnisse durch die Variable „Lehrer“ ausgeschlossen werden.

Feldstudie, weil die Studie während des gesamten Erhebungszeitraumes im realen Schulalltag eingebettet war. Es gab keine Laborbedingungen, bei denen man bestimmte äußere Faktoren ausschließen konnte, sondern musste alle Störfaktoren, die der Schulalltag mit sich bringt in Kauf nehmen.
In folgenden Abschnitten wird auf alle Einzelheiten des Studiendesigns ausführlich eingegangen.

5.1 Zielgruppe

Dass die Studie an allen Schularten Bayerns durchgeführt wurde, hatte zwei Gründe:

Erstens ergaben sich dadurch weitere Fragestellungen bei der Datenauswertung (Gibt es Zusammenhänge zwischen dem Bildungsniveau und der Nutzung des Feedbacks? Bringen bestimmte Arten von Feedback leistungsschwächeren Schülern mehr, als leistungsstärkeren, oder ist es genau umgekehrt? …) und zweitens war die Größe der Stichprobe nur durch die Einbeziehung aller Schularten (Mittelschule, Realschule, Gymnasium) möglich.

Die Verankerung des Themas der Unterrichtseinheit in den Lehrplänen aller Schularten legitimiert dieses Vorgehen. Die Erhebung fand ausschließlich in der 6. und 7. Jahrgangsstufe statt, weil dort der Schwerpunkt der Unterrichtseinheit zu finden war. Natürlich sind Teilaspekte der Lerninhalte, wie genetische Mutationen, Vererbung, etc. erst in höheren Klassen zu finden und können nur dort im Detail verstanden und durchdrungen werden, aber die groben Zusammenhänge sollten auch die Schüler niedrigerer Klassenstufen in einer vereinfachten Form (mit Hilfe einer didaktischen Reduktion) nachvollziehen können.

5.2 Die Unterrichtseinheit

Im Folgenden möchte ich eine kurze Zusammenfassung zu den Inhalten und der Aufbereitung der Unterrichtseinheit und den dazugehörigen Materialien geben. Die Unterrichtssequenz selbst befindet sich im Anhang und kann dort eingesehen werden, um sich einen tieferen Einblick zu verschaffen.

5.2.1 Thema und Inhalt

Da in der Studie nicht nur der individuelle Lernfortschritt der Schüler bei der Aneignung von Faktenwissen überprüft, sondern auch untersucht werden sollte, inwieweit man durch individuelles Feedback und geeignete Fragestellungen, das tiefere Durchdringen von biologischen Zusammenhängen (konzeptuelles Wissen) ermöglichen kann, wurden zwei biologische Konzepte ausgewählt, die in die Unterrichtseinheit mit eingezogen wurden: Das Konzept der „evolutionären Anpassung“ und die physikalischen Zusammenhänge beim „Auftrieb“.

Dabei bildeten die „Vögel und deren Anpassung an das Fliegen“ das übergeordnete Thema der Unterrichtseinheit - diese beinhaltet verschiedene Schwerpunkte, die in unterschiedlichen **Lernblöcken** zusammengefasst wurden:

Der Lernblock 1 beinhaltet alles, was zu einem besseren Verständnis des Konzepts „evolutionäre Anpassung“ führt. So beginnen die Schüler mit dem Herausarbeiten der gemeinsamen Merkmale von Vögeln, wie den Schnabel- und Fußformen und den unterschiedlichen Federtypen und deren Aufgaben. Schon hier können die Kinder erkennen dass im Laufe der Evolution eine Spezialisierung für und eine Anpassung an den jeweiligen Lebensraum der unterschiedlichen Vogelarten stattgefunden hat. Am Beispiel des Spechts erarbeiten sie zusammenfassend dessen Anpassung an seinen Lebensraum Baum in Hinsicht auf Fortbewegung (Kletterfüße, Stützschwanz), Nahrungsaufnahme (Meißelschnabel und Schleuderzunge) und Nestbau.
Vom Beispiel des Spechts ausgehend soll das Gesamtkonzept der evolutionären Anpassung erarbeitet werden. Dies ist nicht ganz einfach, weil es hier sehr viele Fehlvorstellungen gibt, die es erst einmal zu berichtigen gilt (siehe hierzu Punkt 2.3.1.1 – hier werden die Fehlvorstellungen der Schüler genauer erläutert).

Im zweiten Lernblock steht die Anpassung der Vögel an das Fliegen im Mittelpunkt:
Die Schüler erfahren anhand von Experimenten in einem Lernzirkel, warum Vögel fliegen können und wir Menschen nicht.
Dabei spielt das geringe Körpergewicht der Vögel eine große Rolle: Sie sind leicht, aufgrund ihrer hohlen Knochen, ihrer Federn und ihrer hohlen, zahnlosen Schnäbel. Außerdem wiegt das Federkleid um einiges weniger, als das Fell anderer Tiere oder die Haut der Menschen. Weiterhin spart der Vogel Gewicht, indem er seine Eier ablegt und seine Nachkommen nicht mit sich herumträgt.
Auch der Körperbau der Vögel trägt zu ihrer Flugfähigkeit bei: Der stromlinienförmige Körperbau erlaubt ein gleiten durch die Luft, ohne großen Luftwiderstand und somit ohne großen Energieverbrauch, am großen Brustbein setzt die starke Brustmuskulatur an, die es dem Vogel erlaubt, im Ruderflug zu fliegen (die Brustmuskulatur der Menschen müsste ca. 1,5 m dick sein, damit eine Fortbewegung im Ruderflug möglich wäre). Und die gewölbte Flügelform sorgt für den nötigen Auftrieb (die physikalischen Zusammenhänge des Auftriebs sind das zweite Konzept innerhalb der Lerneinheit). Die Eigenschaften der Schwungfeder (wasserabweisend, biegsam und durch Bogen- und Hakenstrahlen jederzeit reparierbar) sind ebenfalls Thema des Lernzirkels.

Der Lernblock 2 wird abgerundet durch die Erarbeitung der unterschiedlichen Flugarten und des Phänomens der Thermik, die sich Vögel zu Nutze machen, um lange Strecken zu gleiten.
Im dritten und letzten Lernblock, der nur eine Stunde umfasst, geht es um den Vogelzug am Beispiel der Störche und wie diese sich die Thermik zu Nutze machen, um in ihr Winterquartier zu gelangen. Anhand einer Atlasarbeit stellen die Schüler erstaunt fest, dass die Störche einen großen Umweg fliegen, um in den Süden zu gelangen. Anstatt den direkten Weg zu nehmen, der auch über dem Wasser verläuft, wählen sie einen Umweg, der ausschließlich über Land geht. Die Erklärung: Wasser wärmt sich nicht genügend auf, dadurch gibt es keine Thermik, über dem Land hingegen, können die Vögel weite Strecken zurücklegen, ohne mit den Flügeln schlagen zu müssen und sparen dadurch sehr viel Energie.
Der dritte Lernblock dient als eine Art Puffer, damit die unterrichtende Lehrkraft nicht in Versuchung gerät, mit der gesamten Klasse vor dem Posttest eine gemeinsame Wiederholung durchzuführen, die mögliche Effekte zerstören würde.

Die Aufteilung in Lernblöcke war notwendig, um die Unterrichtseinheit sinnvoll zu gliedern und um die Testzeitpunkte für die Durchführung der formativen Leistungstests einheitlich festzulegen.

Eine ausführlichere Beschreibung der Tests und deren zeitlichen Einsatzes findet sich unter 5.3.2. Die genaue Aufbereitung der Lehrermaterialien wird unter 4.2.3 beschrieben.

5.2.2 Verankerung in den bayerischen Lehrplänen

In den Lehrplänen aller Schularten, sind die Themen Vogelflug, Vögel, Evolution, Genetik, Selektion, Ökologische Nische und evolutionäre Anpassung verankert. Vor allem in den Jahrgangsstufen 6 und 7 finden sich die meisten Themen, die die Unterrichtseinheit betreffen und deren Schwerpunkt. Deshalb wird die Studie auch in diesen Jahrgängen durchgeführt. Alle angrenzenden Themenbereiche zur evolutionären Anpassung verteilen sich auf das gesamte Schulleben der Schüler, was zeigt, dass dieses Konstrukt immer wieder aufgegriffen wird und in der Biologie von zentraler Bedeutung ist.
Im Folgenden soll kurz aufgezeigt werden, wo sich die verwendeten Lerninhalte in den jeweiligen Curricula der Schularten befinden:

Mittelschule – Schwerpunkt in Jahrgangsstufe 7:

7.1 Luft – Lebensgrundlage und Lebensraum

7.1.4. Lebensraum für Vögel

Die Unterrichtseinheit bezieht sich auf das Lernziel 7.1 des bayerischen Mittelschullehrplans im Fach PCB „Luft – Lebensgrundlage und Lebensraum“. Im Unterpunkt 7.1.4 „Luft – Lebensraum für Vögel“ geht es unter anderem um die Anpassung der Vögel an den Lebensraum Luft: Körperbau, Federkleid und Auftrieb erlauben den Vogelflug – diese Merkmale haben Bedeutung für die moderne Luftfahrt.

Für die Unterrichtseinheit relevante Themen:

6.2.2 Angepasstheit von Lebewesen an den Lebensraum Wasser

9.3 Blick in den Mikrokosmos

9.3.1 Zellen – Bausteine der Lebewesen

9.3.2 Träger der Erbinformation

9.4 Entwicklung des Menschen

9.4.2 Evolution des Menschen

10.3 Blick in den Mikrokosmos

10.3.1 Zellkern im Blickpunkt der Forschung

10.4 Entwicklung der Lebewesen

10.4.1 Stammesgeschichte und Evolution

10.4.2 Evolution des Menschen

Realschule – Schwerpunkt in Jahrgangsstufe 6:

B 6.2 Stammesgeschichtliche Entwicklung

- Formenvielfalt und Verwandtschaft
- gemeinsame Merkmale der Wirbeltiere; Verwandtschaft
- das System der Wirbeltiere im Überblick

6.2.1 Indizien für eine stammesgeschichtliche Entwicklung

6.2.2 Erklärungen für die Artenvielfalt

- Beispiele für das Zusammenwirken von Veränderung und Auslese; die Veränderungen eines Individuums von denen einer Art unterscheiden
- Artenvielfalt; Verantwortung des Menschen

B 6.3 Aufbau von Lebewesen aus Zellen

B 6.5 Lebensgemeinschaft Wald oder Wiese

Für die Unterrichtseinheit relevante Themen:

B 5.2 Lebewesen - Biologie, die Naturwissenschaft von den Lebewesen; Teilbereiche (z. B. Zoologie, Botanik, Genetik)

B 10.2 Genetik

B10.4 Biologische und kulturelle Evolution des Menschen

Gymnasium – Schwerpunkt in Jahrgangsstufe 6:

NT 6.1.1 Jahrgangsstufe: Wirbeltiere in verschiedenen Lebensräumen

Fische, Amphibien, Reptilien, Vögel

Mindestens zwei der vier Klassen werden unter Auswahl geeigneter Aspekte aus folgender Liste behandelt, wobei zur vergleichenden Betrachtung auch Säugetiere herangezogen werden können:

- Körperbau
- Körpertemperatur und Atmung
- Fortbewegung
- Ernährungsstrategien und Nahrungsbeziehungen
- Fortpflanzung und Entwicklung
- Lebensweise im Jahreslauf
- Verhalten
- Gefährdung und Schutz
- Nutzung durch den Menschen
- Bionik: Vergleich von Strukturen und Funktionen mit technischen Anwendungen

Für die Unterrichtseinheit relevante Themen:

NT 5.2.1 Biologie – Lehre von den Lebewesen

B 8.3 Evolutionstheorie: eine naturwissenschaftliche Erklärung zur Entstehung der Arten

B9.3 Grundlagen der Genetik

B 10.3 Grundlegende Wechselbeziehungen zwischen Lebewesen (Ökosysteme)

B 11.1 Strukturelle und energetische Grundlagen des Lebens

B 11.2 Genetik und Gentechnik

B 12.1 Evolution

Wie man sehen kann, sind die Themen „Evolution“ und vor allem „Genetik und Vererbung“ in allen Schularten in den höheren Klassen angesiedelt. Erst dort ist eine tiefere Durchdringung der Begriffe „genetische Variabilität“, „Mutation von Genen“ und „Evolution“ möglich. In der Unterrichtseinheit der Studie kann man eher von einer Metaebene aus, also etwas abstrakter mit den Schülern besprechen, was es heißt, wenn ein Gen mutiert und nicht bis ins Detail gehen. Dennoch sollte es bewusst, neben dem Auftrieb einen zweiten Punkt geben, wo es um die Erarbeitung von Zusammenhängen und damit um konzeptuelles Wissen und nicht um die reine Abfrage von Faktenwissen geht.

5.3 Aufgaben- und Testdesign

5.3.1 Aufgabentypen

Insgesamt gibt es eine Vielzahl an unterschiedlichen Aufgabentypen, die für verschiedene Zwecke (Abfragen von Faktenwissen oder konzeptuellem Wissen) und verschiedene Darreichungsformen (Paper-Pencil Tests, mündliche Leistungserhebung oder elektronische Leistungstests) unterschiedlich gut geeignet sind. Im Folgenden soll mit Hilfe einer Tabelle eine Auswahl der wichtigsten Aufgabentypen aufgelistet und sowohl deren Eignung, als auch deren Vor- und Nachteile kurz erläutert werden:

Tabelle 15: Aufgabentypen und ihre Eignung

Aufgabentyp	Umsetzung am PC / Tablet PC	Herkömmlicher Test in Schriftform (Paper - Pencil)	Zweck / Eignung	Anmerkungen
Multiple Choice/ Single Choice	Leichte Erstellung und Auswertung. Einbindung versch. Medien möglich (Animationen, Filme, Bilder, etc.)	Einfache, aber nicht automatisierte Auswertung. Allerdings ist nur die Einbeziehung von Text und unbewegten Bildern möglich.	Besonders gut, um Faktenwissen abzuprüfen, aber bei geschickter Formulierung der Frage und der dazugehörigen Antwortmöglichkeiten auch geeignet, um Wissenslücken bei Zusammenhängen aufzudecken.	Genaue Formulierung der Fragen sehr wichtig, um eventuelle Denkfehler der Schüler aufdecken zu können.
Offene Fragen / freier Text	Nicht automatisch auswertbar (man kann den Text nur nach bestimmten Begriffen absuchen, aber auch nicht werten, ob sie im richtigen Zusammenhang genutzt werden).	Auswertung möglich, aber mit hohem Korrekturaufwand verbunden.	Sehr gut geeignet, um Zusammenhänge und Verständnis zu überprüfen: Die Schüler müssen die Antwort selbst formulieren und daher genau reflektieren, was sie aussagen möchten.	Ungenaue Antworten machen eindeutige Bewertung schwierig. Punktvergabe ist manchmal Auslegesache und könnte bei zwei verschiedenen Korrektoren schon unterschiedlich ausfallen → Könnte zu Ungenauigkeiten in der Evaluation führen
Experiment (Versuchsaufbau)	Als Drag and Drop Aufgabe denkbar, allerdings nur mit hohem Programmieraufwand umzusetzen. In einer abschließenden Animation könnte der Versuchsablauf und das Ergebnis gezeigt werden. (Was passiert, wenn man …)	Auf dem Papier als Zuordnungsaufgabe möglich: Versuchsaufbau ist vorgezeichnet, richtige Geräte/Zutaten werden durch die richtige Nummer an den dafür vorgesehen Platz „gestellt“.	Dient dazu, eine Kernkompetenz des naturwissenschaftlichen Arbeitens abzuprüfen. Ursache und Wirkung, sowie konzeptuelles Wissen stehen im Vordergrund.	Eigentlich nur als mündliche Leistungserhebung während des Unterrichts ohne zu großen Aufwand machbar: Eine eindeutige Bewertung ist schwierig.
Lückentext	Sowohl in elektronischer Form, als auch in herkömmlichen Tests üblich und leicht umsetzbar. Es gibt zwei Möglichkeiten: Einige Antworten werden vorformuliert und dem Lerner zur Auswahl angeboten, oder der Schüler füllt die Lücken selbst aus (dann als Paper-Pencil Test wegen der manuellen Auswertung besser geeignet).		Kann sowohl reines Faktenwissen, als auch Zusammenhänge abfragen, vor allem, wenn die Lücken zueinander in Beziehung gesetzt werden müssen.	Dadurch, dass man die Lücken vorgeben oder selbst formulieren lassen kann, ist eine Differenzierung im Schwierigkeitsgrad sehr gut umsetzbar.

Aufgabentyp	Umsetzung am PC / Tablet PC	Herkömmlicher Test in Schriftform (Paper - Pencil)	Zweck / Eignung	Anmerkungen
Kreuzworträtsel	Kann man mit Hilfe von open Source Software kostenfrei und schnell erstellen und abspeichern. Eigentlich soll man diese Quizzes auch in Moodle importieren können, was aber leider nicht funktioniert hat.	Es gibt Programme (auf CD Rom), mit denen man automatisch Kreuzworträtsel erstellen, in die Unterrichtsmaterialien integrieren und dann ausdrucken kann.	Spielerische Möglichkeit, um Daten und Fakten abzufragen.	
Sortier-/ Zuordnungsaufgaben	Sowohl mit Drag and Drop, als auch mit Hilfe von Drop Down Menüs (bei Moodle) umsetzbar. Auch bei diesem Aufgabentyp kann man Medien, wie Filme einbinden.	Auf dem Papier kann man die richtigen Antworten durch die Zuordnung der richtigen Zahl/ des richtigen Buchstaben oder das Verbinden von zwei Elementen realisieren.	Dieser Aufgabentyp eignet sich vor allem, um Bildern die richtigen Begriffe zuzuordnen und zusammengehörige Begriffe miteinander in Verbindung zu bringen.	Die Zuordnungsaufgaben, die ich in Moodle erstellt habe, dienen vor allem der Abfrage von Faktenwissen.
Zeichnung erstellen	Die Erstellung einer Skizze ist auf Tablet PCs durch den Touchscreen ohne Probleme möglich, eine automatisierte Auswertung leider nicht.	Da herkömmliche Tests immer manuell ausgewertet werden, ist die Auswertung von Zeichnungen hier unproblematisch.	Besonders gut geeignet, um abzuprüfen, ob die Schüler genau beobachtet haben und eine detailgetreue Wiedergabe schaffen.	Prima, wenn es um kleine Details und ihre Darstellung geht (Mikroskopierergebnisse).
Suchbilder mit falschen Elementen	Mit „Hot Spots“ leicht umsetzbar, eine Einbeziehung von externen Medien ist denkbar (klickt man auf ein falsches Element, geht ein neues Fenster mit zusätzlichen Informationen, etc. auf).	Durch Ankreuzen der falschen Elemente werden diese isoliert. Eine direkte Rückmeldung zur getroffen Auswahl ist nicht möglich).	Durch die Beschriftung von Bildern oder das Isolieren von falschen Elementen in einer Zeichnung, müssen die Schüler Ordnung in die fachlichen Begriffe bringen und diese mit ihrer bildlichen Vorstellung in Einklang bringen.	Auf einem Bild kann man auch sehr schön einen Zusammenhang zeigen, der durch das Entfernen einzelner Elemente, richtig gestellt werden muss.
Bilder richtig beschriften	Drag and Drop: Begriffe an die richtige Stelle ziehen, oder manuell ausfüllen lassen (mögliche Rechtschreibfehler müssen bei der Auswertung der Begriffe beachtet werden).	Leicht realisierbare Aufgabenstellung		Auch hier ist eine Differenzierung im Schwierigkeitsgrad denkbar (siehe Lückentext).

Da nach der Vorstudie entschieden wurde, bei der Studie elektronische Tests einzusetzen, gab es bei der Auswahl an Aufgabentypen, eine von der Technik (genauer: von Moodle) eingeschränkte Auswahl. Je nach Autorensystem, sind unterschiedlich viele Testformen möglich. Dabei gilt: Je mehr Programmieraufwand man zu investieren bereit ist, desto größer ist die Auswahl an Möglichkeiten. Da die Programmierkünste der Lehrkräfte bei Moodle nicht sehr ausgeprägt sein müssen, kann man nur auf eine kleinere Auswahl von Aufgabentypen zurückgreifen. Aus dieser Auswahl wurden wiederrum diejenigen ausgewählt, die für die Lerninhalte und das Projekt am besten geeignet waren. Auf diese soll im nächsten Abschnitt genauer eingegangen werden:

5.3.1.1 Multiple Choice / Single Choice

Multiple Choice bzw. Single Choice Fragen sind sehr beliebt, wenn es darum geht, viel Wissen innerhalb kurzer Zeit abzufragen und vor allem leicht auszuwerten. Dieser leichten Auswertbarkeit fallen allerdings auch Erkenntnisse zum Opfer, die man aus einer selbst formulierten Antwort des Lerners bei der Beantwortung einer offenen Frage ziehen könnte. Die vorformulierten Antwortmöglichkeiten einer Multiple Choice Aufgabe müssen demnach sehr gut durchdacht sein, wenn man aufdecken möchte, wo die Denkfehler der Schüler innerhalb des Problemlöseprozesses liegen. Daher sieht es zwar oberflächlich betrachtet sehr einfach aus, Multiple Choice Aufgaben zu erstellen, schaut man genauer hin, stellt man jedoch schnell fest, dass es beim Design der Testfragen und der dazugehörigen Antwortmöglichkeiten sehr viel zu beachten gibt, wenn die Auswertung der Ergebnisse aussagekräftig in Hinblick auf den Lernzuwachs der Schüler und dessen Durchdringung des Lerninhaltes sein sollen.

In den Tests wurden sowohl Multiple, als auch Single Choice Aufgaben verwendet, die sich für die Schüler optisch unterschieden (siehe Beispiele unten).

Ist nur eine Antwort richtig, ist die anklickbare Form vor den Antwortmöglichkeiten rund:

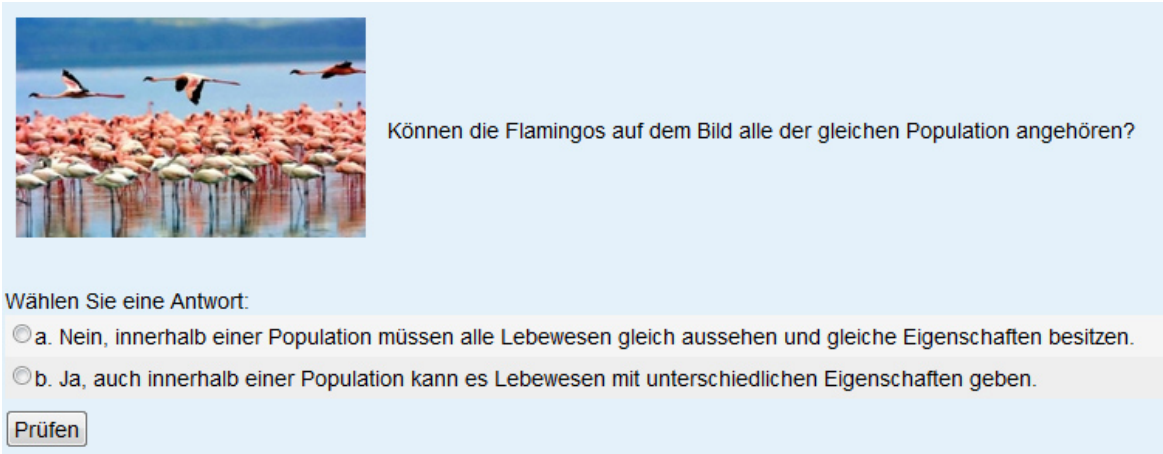


Abbildung 9: Beispiel einer Single Choice Aufgabe in Moodle

Sind mehrere Antworten richtig, ist das Kästchen zum Ankreuzen viereckig.

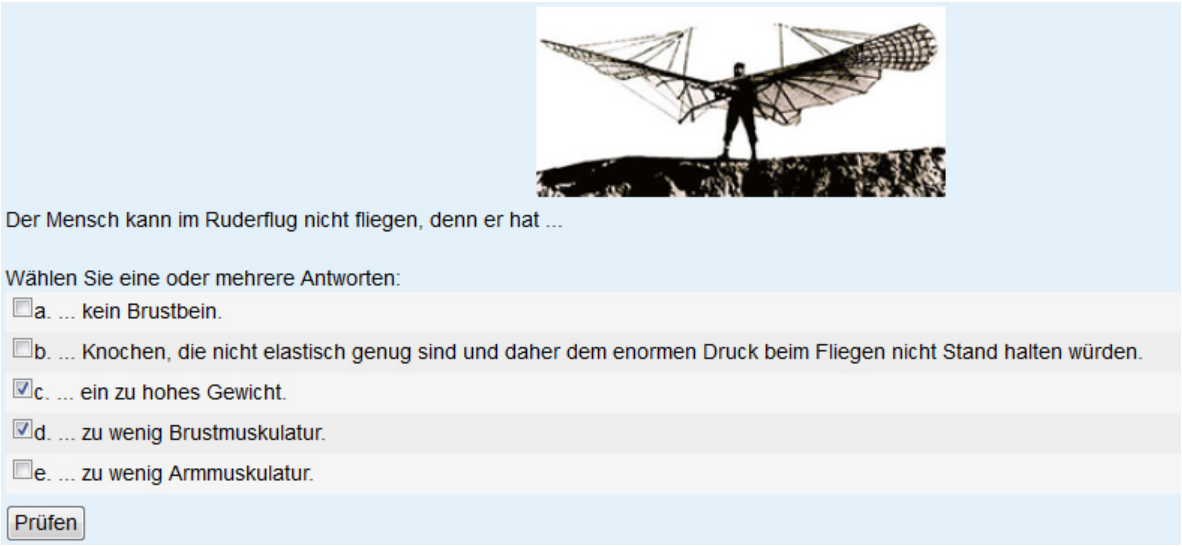


Abbildung 10: Beispiel einer Multiple Choice Aufgabe in Moodle

In der Literatur findet man eine Vielzahl von Empfehlungen, an die man sich bei der Formulierung von MC Fragen und deren Antwortitems halten sollte. Einige davon wurden beim Design der Aufgaben für die Studie beachtet, wie zum Beispiel, dass alle Wahlantworten grammatikalisch zum Stamm passen sollten, oder dass alle Antwortmöglichkeiten inhaltlich aus dem gleichen Themengebiet stammen müssen. Viele Empfehlung waren eher für den Einsatz im Hochschulbereich gedacht, wie zum Beispiel, dass vor der eigentlichen Frage eine realistische (lebensnahe) und ausführliche Vignette formuliert werden soll, die dem Leser ein Problem plas-

tisch vor Augen führt. Da ich es mit Schülern zu tun habe, die teilweise (vor allem in den Hauptschulklassen), oft noch Verständnisschwierigkeiten bei langen, ausgefeilten Texten haben, wurde auf ausschweifende und dadurch eventuell verwirrende Vorgeschichten verzichtet. Die MC Fragen wurden, neben den Zuordnungsfragen, vor allem beim Abfragen von Faktenwissen eingesetzt, teilweise wurde aber auch versucht, durch kurze, genau formulierte Fragen und die dazugehörigen Antwortmöglichkeiten, die Defizite der Lerner zu einem bestimmten Teilschritt im größeren Wissenszusammenhang (Konzept der evolutionären Anpassung) zu isolieren.

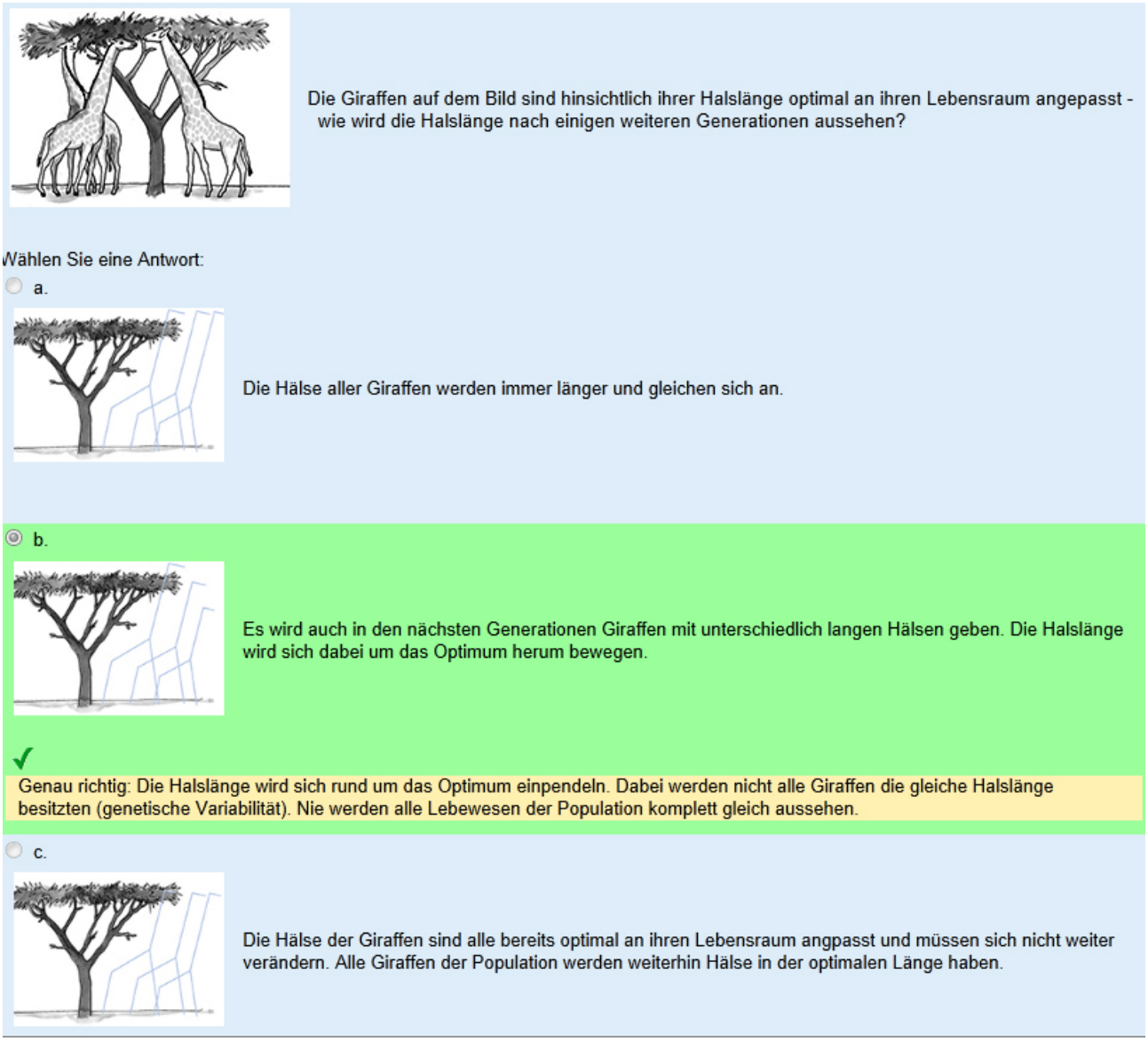


Abbildung 11: Beispiel einer Single Choice Aufgabe zur Abfrage konzeptuellen Wissens

Das Beispiel zeigt eine Aufgabe zur Problematik der genetischen Variabilität innerhalb des Konzepts der evolutionären Anpassung.
Im Feedback wurde versucht, genau auf dieses isolierte Teilproblem einzugehen, indem die Merkmale der genetischen Variabilität noch einmal ausführlich aufgegriffen und erläutert wurden.
Sollten mehrere Zusammenhänge innerhalb eines größeren Kontextes abfragt werden, wurden vor allem auf Lückentextaufgaben gewählt (siehe Punkt 5.3.1.2).

Dennoch wurde versucht, neben den bereits oben angeführten beachteten Kriterien, einige allgemeingültige Dinge bei der Konzeption der Aufgaben zu berücksichtigen:

- Die Fragen sind kurz, prägnant und eindeutig formuliert, da es um das Abprüfen des Problemlöseverständnisses und nicht um das Sprachverständnis geht.
- Die Schüler könnten die Frage auch beantworten, ohne die Antwortmöglichkeiten zu kennen.
- Jede Wahlantwort enthält nur eine Aussage und diese ist so kurz wie möglich gehalten.
- Alle absoluten Begriffe, wie „immer, nie, ausschließlich“ (in der Regel ist die Aussage dann falsch) und vage Begriffe, wie „meist, selten, oft“ (Aussage meistens richtig) wurden vermieden.
- Doppelte Verneinungen wurden ebenfalls nicht verwendet, weil das nur zu Verwirrungen geführt hätte.
- Alle Antwortmöglichkeiten sollten gleich plausibel sein, um die Ratewahrscheinlichkeit möglichst gering zu halten.
- Inhalte aus vorhergehenden Fragen sollten nicht bei der Beantwortung späterer Aufgaben helfen, oder diese Vorwegnehmen.

Damit sich die Schüler bei den Aufgaben nicht die Anordnung der falschen und richtigen Antworten einprägen konnten, wurde Moodle so eingestellt, dass die Antworten bei jedem Test durchmischt und neu angeordnet wurden.

5.3.1.2 Lückentext

Wie bereits an früherer Stelle angeführt, entspricht das Lückentextformat am meisten, dem für die Vorstudie entwickelten, Paper – Pencil – Aufgabentyp, mit dem Zusammenhänge innerhalb größerer Wissenskonstrukte (Auftrieb, evolutionäre Anpassung, Thermik, etc.) abgefragt werden sollte. Dies wurde möglich, da der Aufgabentext von den auszuwählenden Textbausteinen der unterschiedlichen Lücken unterbrochen wurde und durch die Wahl der richtigen Lücke in eine wahre Aussage verwandelt werden musste. Die Lücken und ihre Aussage mussten also nicht nur für sich richtig sein, sondern auch noch richtig in Beziehung zueinander gesetzt werden.

Als Beispiel soll eine Frage zur evolutionären Anpassung herangezogen werden, die gleich mehrere mögliche Denkfehler auf Schülerseite aufdecken kann, denn in der Frage werden unterschiedliche Teilbereiche der evolutionären Anpassung berücksichtigt:

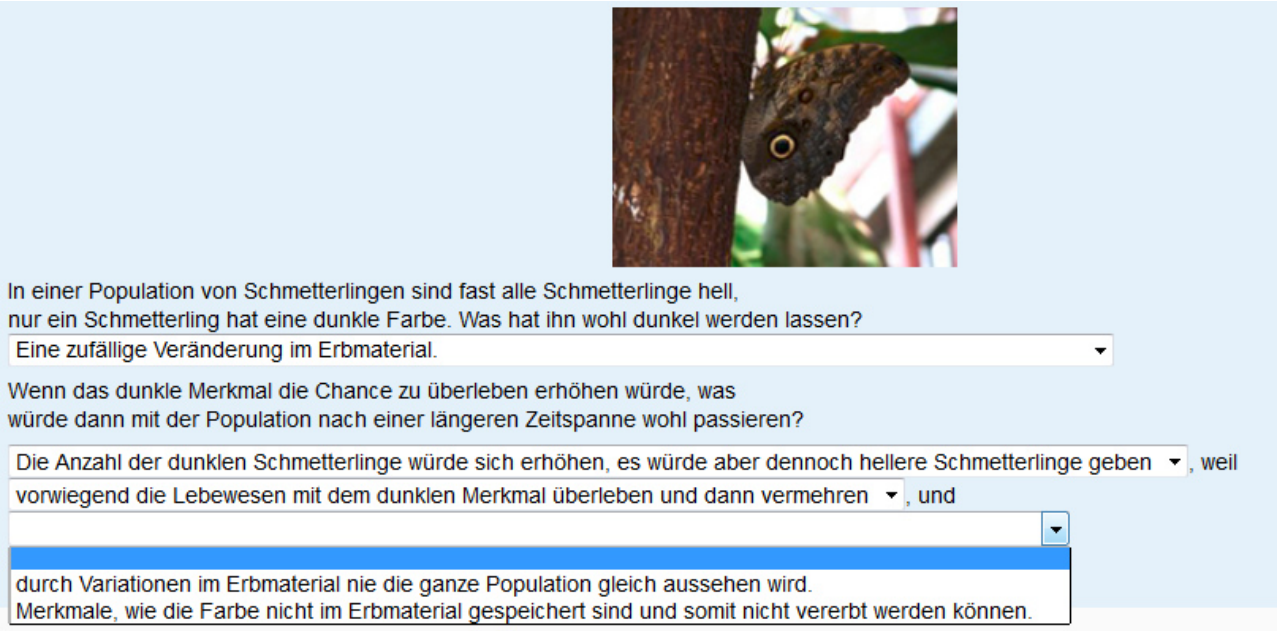


Abbildung 12: Beispiel einer Lückentextaufgabe in Moodle

Nach Beantwortung der Frage, erhalten die Schüler der ersten Treatmentgruppe folgende Rückmeldung:

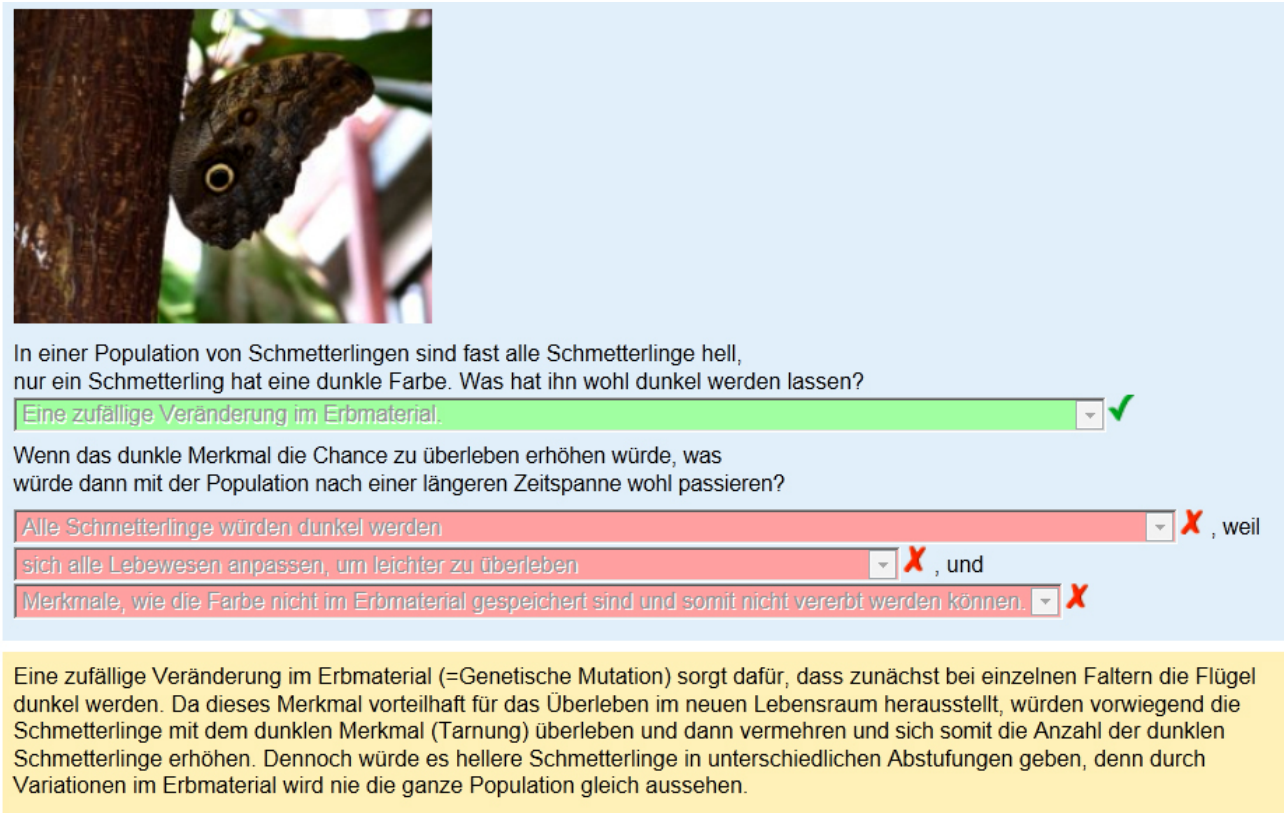


Abbildung 13: Ausführliches Feedback - Treatmentgruppe 1

Im Feedback wird immer auf alle Teilaspekte eingegangen, die in der Frage angesprochen werden, egal, welche der Antwortmöglichkeiten vom Lerner innerhalb der Lücken richtig oder falsch gewählt wurden. Bei Multiple Choice Aufgaben ist es häufig so, dass es zu jeder einzelnen Teilantwort eine spezifische Rückmeldung gibt.

5.3.1.3 Zuordnungsaufgaben

Zuordnungsaufgaben bieten sich vor allem bei der Abfrage von Faktenwissen an. Dabei eignen sich sowohl Begriffe, als auch Bilder oder Videos, um diese bestimmten Fachbegriffen oder Gegebenheiten zuzuordnen. In den Tests der Studie werden nur Begriffe und Bilder verwendet, da es in Moodle Probleme beim gleichzeitigen Abspielen mehrerer Filme und dem Abschalten der Tonspur gab.

Um zu zeigen, wie die Testumgebung aussieht, in der die Schüler arbeiten, wird das Beispiel der Zuordnungsaufgabe, eingebettet in die Nutzoberfläche von Moodle gezeigt. Eine genauere Beschreibung von Moodle und dessen Möglichkeiten findet man im Kapitel 5.4.2.

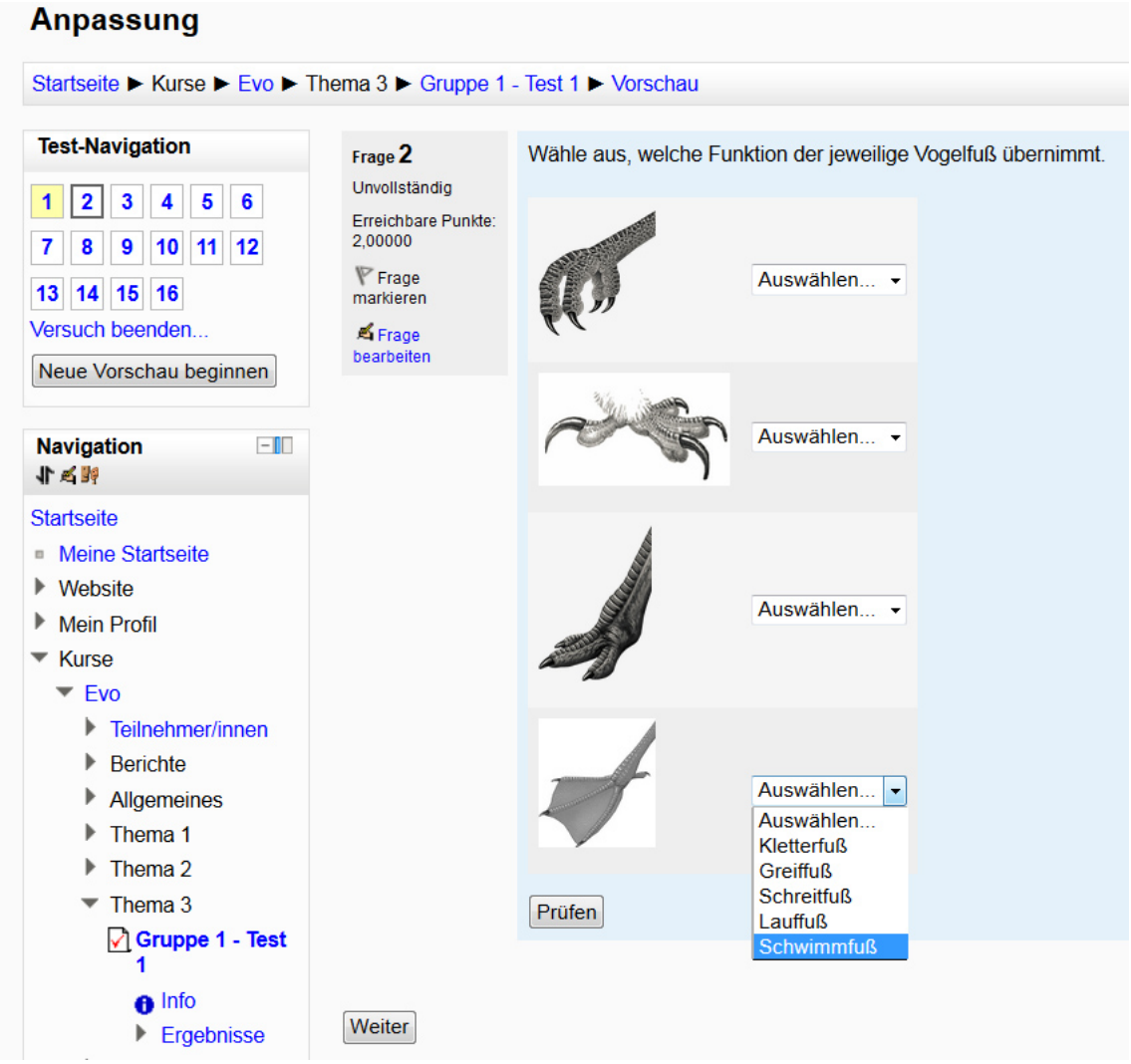


Abbildung 14: Zuordnungsaufgabe in der Testumgebung von Moodle

Eine elegantere und umfassendere Möglichkeit der Zuordnung sind Drag and Drop Aufgaben: Hier kann man gleich mehrere richtige Komponenten einem kompletten Szenario zuordnen, wie zum Beispiel, Vögel in den richtigen Lebensraum setzen und mit den passenden Schnäbeln und Füßen ausstatten. Allerdings sind in Moodle Drag & Drop Aufgaben und der Import extern generierter Aufgaben nicht vorgesehen, bzw. funktionierte nicht, so dass ich auf die vorhandenen einfachen Zuordnungsaufgaben zurückgreifen musste.

In Verbindung mit den kleinen, 7 Zoll – Tablets gestalteten sich die Zuordnungsaufgaben ein wenig problematisch, da beim Ausklappen der Drop Down Menüs aufgrund des Platzmangels nicht alle Auswahlmöglichkeiten sichtbar waren – um alle Möglichkeiten ansehen zu können, mussten die Schüler im Menü hin und her scrollen. Dieses Problem gab es bei den größeren iPads nicht, auch wurden dort alle Antwortsätze vollständig angezeigt und nicht teilweise abgeschnitten.

5.3.2 Messinstrumente

5.3.2.1 Einleitung

Ein Test soll die Leistungsfähigkeit eines Lerners ermitteln und dabei das zu prüfende Themengebiet breit abdecken. Zunächst muss man sich folgende Fragen stellen:

Was möchte ich prüfen? Hier stehen die Lernziele im Mittelpunkt des Interesses. Es ist zu entscheiden, mit welcher Frage und welchem Aufgabentyp, man am besten abprüfen kann, ob der Schüler die Lerninhalte auch wirklich verstanden hat.

Auf welchem Level möchte ich dies prüfen? Es gibt verschiedene Taxonomiestufen (Wissenslevel), denen die Lernziele angehören können (nach Bloom 1956: Wissen, Verstehen, Anwenden, Analyse, Synthese, Urteilen).

- Die „niedrigste“ Taxonomiestufe ist wohl die Reproduktion von Faktenwissen, also die Wiedergabe von erworbenen Daten, Fakten und Begrifflichkeiten.
- Komplexer ist die theoretische Anwendung des Gelernten, also eine theoretische Kompetenz, die es erlaubt, das angeeignete Wissen zu reorganisieren und in einem nächste Schritt auch
- praktisch anzuwenden, das heißt auf andere Situationen zu transferieren. Man kann also, wenn man das Erlernte tiefer durchdrungen und verstanden hat, dieses Wissen auf neue Gegebenheit anwenden (praktische Kompetenz) und später sogar
- die erworbenen Kenntnisse beim eigenen Handeln im Alltag mit einbeziehen. (problemlösendes Denken – Synthese).

Außerdem sollte man zusätzlich folgende Anforderungen an seine Prüfung stellen:

Sie sollte vorwärts gerichtet konzipiert werden, das heißt für die Zukunft (die weitere Unterrichtspraxis) sinnvoll sein und nur wichtige Dinge beinhalten.
Dabei sollte sie alles Wesentliche abprüfen (Qualitätskontrolle) und objektiv (verschiedene Prüfer erzielen das gleiche Ergebnis), reliabel (zuverlässig – Genauigkeit, mit der ein Merkmal gemessen wird) und valide (Gültigkeit der Prüfungsergebnisse – wurde wirklich das abgeprüft, was getestet werden sollte?) sein. Die Machbarkeit und Nachvollziehbarkeit ist ebenfalls ein wichtiger Aspekt der Prüfung und zu guter Letzt sollte sie das Lernverhalten der Schüler positiv beeinflussen.

In der Studie wurden verschiedene elektronische Leistungstests eingebunden, die in den folgenden Abschnitten näher erläutert werden.
Einen Überblick über alle, in den elektronischen Leistungstests (Pre-, Posttest und formative Leistungstests) auftauchenden Testfragen, deren Antwortmöglichkeiten und Rückmeldetexte kann man sich anhand der Excelliste „Fragensammlung“ (bitte bei der PH Schwäbisch Gmünd anfordern) verschaffen. Dort sieht man auch, um welchen Aufgabentyp es sich bei der jeweiligen Frage handelt.
Unter dem Punkt 4.2.4 „Durchführung des Feldexperiments“ befindet sich eine genauere Übersicht, zu welchen Zeitpunkten der Unterrichtssequenz die einzelnen Tests durchgeführt wurden.

Bei allen Tests wurde eingestellt, dass die Schüler immer nur eine Chance haben, die Fragen zu beantworten, denn lässt man eine mehrmalige Antwortmöglichkeit zu, können die Schüler so lange raten, bis sie das richtige Ergebnis gefunden haben. Einerseits wäre eine ausführlichere Beschäftigung mit einer Frage und deren Antworten natürlich vorteilhaft, wenn die Lerner dabei genau auf die Lerninhalte schauen und reflektieren, welche Antwort falsch bzw. richtig war und warum, denn dadurch würde ja bereits wieder in Lernprozess angestoßen. Allerdings könnte es auch sein, dass einfach, ohne nachzudenken, alle Antwortmöglichkeiten durchprobiert werden, bis die Vollpunktzahl erscheint. In diesem Fall wäre der Wissenszuwachs sehr gering oder gar nicht vorhanden.

Im folgenden Abschnitt werden, die im Feldexperiment verwendeten Tests kurz angeführt:

Der Pre- bzw. Vortest beinhaltet Fragen zu allen Lerninhalten der Unterrichtseinheit "Evolutionäre Anpassung" und „Anpassung der Vögel an das Fliegen (21 Items) und soll das Vorwissen der Schüler erheben. Genaueres über den Pretest findet sich in Kapitel 5.3.2.2 und im Anhang in der Fragensammlung, wo man alle Testitems und das dazugehörige Feedback einsehen kann.

Zwei in den Lernprozess eingearbeitete formative Leistungstests (FLT 1 = 16 Items, FLT 2 = 14 Items) dienen der Übung des vorhergehenden Unterrichtsstoff und geben den Schülern der Treatmentgruppen Rückmeldung zu ihren Antworten.

Ein Posttest (identisch mit dem Pretest), der als summative Leistungserhebung nach der Unterrichtseinheit dient, überprüft den Lernzuwachs aller Schüler und lässt einen individuellen Vergleich zum Vorwissen der einzelnen Lerner zu.

Die Behaltensleistung wird mit einem follow-up Test zur Unterrichtseinheit ca. 6 – 8 Wochen nach der Unterrichtseinheit geprüft. Anders als die vorhergehenden Tests, wird diese Prüfung nicht elektronisch, sondern auf herkömmlichen Weg, handschriftlich durchgeführt. Offene Fragestellungen sollen aufzeigen, welche Schüler die Lerninhalte und vor allem die Konzepte „Evolutionäre Anpassung“ und „Auftrieb“ durchdrungen haben und eigenständig die richtigen Zusammenhänge herausstellen und diese formulieren können. Alle Items des Behaltenstests sind selbst konstruiert, in einer Vorstudie erprobt und auf ihre Reliabilität hin überprüft worden.

Die lernrelevanten Emotionen (Interesse, Langeweile und Wohlbefinden) während der Unterrichtseinheit werden an drei Zeitpunkten mit Hilfe von Kurzskalen (9 Items) erhoben und beobachtet. (Randler et al. 2011)

Parallel zum Pre- und Posttest gibt es jeweils eine Motivationsskala, die unter dem Gliederungspunkt 5.3.2.5 zu finden sind. Genauere Informationen zu den Skalen befinden sich im Dokument „Skalendokumentation“.
Folgende Kontextfaktoren werden mit Hilfe von selbstkonstruierten Items innerhalb der Motivationstests zur Kontrolle erfasst:
Vornoten in den Fächern Biologie, Deutsch und Mathematik, sozioökonomischer Hintergrund der Schüler (Beruf und Bildungsabschluss der Eltern), Migrationshintergrund sowie lernförderliche Prozesse.

Zusätzlich beinhaltet der zweite Motivationstest selbstkonstruierte Items zur Bewertung der Unterrichtseinheit, und zum Umgang mit den Tablet PCs, den elektronischen Leistungstests und dem Feedback. Auch diese Skalen wurden in der Vorstudie erprobt, einer Faktorenanalyse unterzogen und auf ihre Reliabilität hin überprüft.

Alle Erhebungen wurden mit allen Schülern (allen Treatmentgruppen) der einzelnen Klassen unter Konstanthaltung der Lernzeit und der äußeren Bedingungen (Lehrer, Unterrichtsraum, …) durchgeführt.
Alle Testitems wurden einer Reliabilitätsprüfung unterzogen (siehe Kapitel 6.2.3).
Im weiteren Verlauf sollen die einzelnen Messinstrumente genauer beschrieben werden.

5.3.2.2 Pre- und Posttest

Pre- und Posttest sind inhaltlich identisch, denn durch den Vergleich dieser beiden Tests, soll der Wissenszuwachs jedes Schülers während der Unterrichtseinheit erfasst werden.
Der Pretest findet ganz am Anfang der Erhebungsphase statt und soll das Vorwissen der Schüler zum Thema „Vogelflug“ und „Anpassung“ festhalten. Da die Fragen sehr anspruchsvoll sind, wurde den Schülern vor dem Test klar gemacht, dass es nicht schlimm ist, wenn sie Vieles noch nicht wissen und sie alles über die abgefragten Lerninhalte in den folgenden Unterrichtsstunden erfahren werden. Sie sollten immer das ankreuzen, was ihnen am wahrscheinlichsten vorkam, wenn sie die richtige Antwort nicht wussten. Auch wurde bei der Abgabe der Tests häufig darauf hingewiesen, dass die erzielte Punktzahl ja gar nicht so schlecht sei, dafür, dass ihnen die

Lerninhalte ja noch nicht vertraut seien. Dies war wichtig, um die Lerner nicht schon bei der ersten Prüfung zu entmutigen.
Der Posttest schließt die Unterrichtseinheit ab und besteht aus den gleichen 21 Items, wie der Pretest, um einen direkten Vergleich zu ermöglichen.
Diese 21 Items sind Fragen zu allen Lerninhalten der gesamten Unterrichtssequenz.
Um eine bessere Vergleichbarkeit bzw. Unterscheidung zwischen den verschiedenen Themen und damit Wissensleveln (Fakten- / Konzeptwissen) besser zu ermöglichen, wurden die einzelnen Fragen für die Auswertung zusätzlich unterschiedlichen Kategorien zugeordnet.

So gibt es folgende Kategorien:

1) Gemeinsame Merkmale (Schnabel- und Fußformen, Federarten und deren Funktionen) – vorwiegend Faktenwissen und daher viele Zuordnungsaufgaben
2) Evolutionäre Anpassung (Konstrukt der Anpassung: Evolution, Natürliche Selektion, Genmutation, Vererbung, ökologische Nische, genetische Variabilität) – vor allem konzeptuelles Wissen
3) Anpassung der Vögel an das Fliegen – Mischform (beinhaltet Fragen, die reine Fakten abprüfen, aber auch Fragen zu den physikalischen Zusammenhängen beim Auftrieb)
4) Flugarten und Thermik (vorwiegend Faktenwissen, mit dem Vogelflug als Anwendung (Transfer) des Thermikwissens)

Zusätzlich wurden in SPSS alle Fragen darauf hin gekennzeichnet, zu welchem Aufgabentyp und zu welchem Wissenslevel sie gehören.
Bei der PH Schwäbisch Gmünd befinden sich zwei Dateien, denen man genauere Informationen zu den einzelnen Aufgaben (Fragensammlung) und der Eingruppierung in SPSS (Skalendokumentation) entnehmen kann.

Den Posttest haben alle Lehrkräfte als bewertete Probearbeit herangezogen, was zur Folge hatte, dass sich die Schüler bei diesem Test besonders angestrengt haben. Die zusätzliche Note hat die Integration der Studie (die ja doch viel Zeit in Anspruch nahm und dazu führte, dass andere Biologiethemen weniger ausführlich behandelt werden konnten) in den eigentlichen Biologieunterricht für die Lehrer erleichtert. Hier kam auch der klare Vorteil der elektronischen Leis-

tungstests zum Vorschein: Mit einem Knopfdruck war alles korrigiert und jeder Schüler hatte seine Note. Die Testdaten der einzelnen Schüler (wie viele Punkte bei jeder Frage erreicht wurden und auf Wunsch: welche Detailantwort bei der jeweiligen Frage gegeben wurde), habe ich in eine Excelliste exportiert und den Lehrkräften zur Verfügung gestellt.

5.3.2.3 Formative Leistungstests / Treatment

Nach der Vorstudie wurden die Testzeitpunkte der formativen Leistungstests von drei auf zwei gekürzt. Gründe hierfür waren erstens ein zeitlicher: Mit drei Testzeitpunkten hätte sich die gesamte Studie noch länger hingezogen (durch die zusätzliche Teststunde und die zusätzlichen Lerninhalte), zweitens ein organisatorischer: Die Tests sind doch sehr aufwändig, denn man muss immer mit dem ganzen technischen Equipment an die unterschiedlichen Schulen fahren und bei 10 Klassen ist die Koordination und Abstimmung der Testzeitpunkte auch nicht immer einfach, und drittens ein motivationaler: es hat sich bereits in der Vorstudie abgezeichnet, dass zu viele Tests demotivierend auf die Schüler wirken. Ich denke, mit 2 formativen Leistungstests wurde ein guter Kompromiss gefunden.

Die Fragen der formativen Leistungstests beziehen sich jeweils auf den unmittelbar vorhergehenden Lerninhalt der Unterrichtssequenz. Der erste formative Leistungstest erfolgte direkt nach dem evolutionären Anpassungsbegriff und enthielt 16 Fragen zu den Kategorien „Gemeinsame Merkmale“ und „Evolutionäre Anpassung“.
Der FLT 2 fand nach den Lernblöcken zu den Kategorien „Anpassung der Vögel an das Fliegen“ und „Flugarten und Thermik“ statt und erstreckte sich über 14 Aufgaben. Nach dem zweiten formativen Test folgte nur noch eine Stunde zum Thema „Vogelzug“, wo das erlernte Wissen über die Thermik Anwendung fand. Diese Stunde diente als eine Art Puffer vor dem Posttest, damit die jeweilige Lehrkraft nicht in Versuchung geriet, den gesamten Unterrichtsstoff vor der Prüfung noch einmal mit der ganzen Klasse zu wiederholen und dadurch mögliche Effekte zu verwischen.
Ursprünglich war angedacht, alle Tests, also Pre-/Posttest und FLT1 und 2 identisch zu gestalten, das heißt überall alle 21 Fragen zur gesamten Unterrichtseinheit einzubeziehen, um den Wissenszuwachs in den unterschiedlichen Kategorien über den gesamten Erhebungszeitraum

besser verfolgen und vergleichen zu können, aber das hätte mehrere Nachteile mit sich gebracht:
Man hätte für alle Testzeitpunkte mehr als eine Unterrichtsstunde benötigt, was eine riesige Umorganisation in den Schulen (Stundentausch, Lehrertausch, Raumauslastung, etc.) nach sich gezogen hätte. Mit den 14/16 Fragen kam man mit der jeweiligen Einführung zu den Neuerungen beim Test (Gruppeneinteilung, etc.) gerade so hin.
Außerdem hätten die Schüler immer einen mehr oder weniger großen Anteil an Fragen gehabt, über den sie noch nichts in Erfahrung gebracht haben. Es ist auf Dauer nicht motivierend, wenn man beim Anwachsen der Gesamtpunktzahl der Tests nur wenig Fortschritt beobachten kann.
Weiter wäre es schwierig für die Schüler, sich in der soundso schon geringen Zeit, auf die Wiederholung und Vertiefung des bisher Gelernten zu konzentrieren, weil sie immer auch Fragen bearbeiten müssten, die sie noch nicht besser lösen könnten, als beim Pretest.

Aus oben genannten Gründen wurden in den FLTs nur Fragen zu den vom vorhergehenden Test bis zum jeweiligen neuen Testzeitpunkt durchgenommenen Lerninhalten gestellt. Dabei wurde darauf geachtet, die in Pre- und Posttest auftauchenden Kernfragen zu den Themen nochmals aufzugreifen und mit weiteren Übungsaufgaben anzureichern, um das jeweilige Wissensgebiet weiter zu vertiefen.
Ein weiterer Unterschied der formativen Leistungstests zum Pre-/Posttest (neben dem Umfang) liegt in der Gruppeneinteilung: Zwar schreiben alle Schüler der Klasse einen identischen Pre- und Posttest, doch während der formativen Tests werden sie in 3 unterschiedliche Gruppen unterteilt (siehe hierzu auch Gliederungspunkt 4.2.4 „Durchführung des Feldexperimentes“). Da diese Einteilung Kernpunkt der Erhebung ist, soll sie hier noch einmal genauer beschrieben werden:

Treatment 1: Diese Gruppe schreibt den formativen Test und erhält nach der Beantwortung der Testfragen ein ausführliches Feedback. Darin wird erläutert, was der Schüler bei der jeweiligen Frage falsch und richtig gemacht hat, und wie die richtige Antwort ausgesehen hätte. Diese erste Treatmentgruppe wurde bei der statistischen Auswertung der Daten in zwei Subgruppen unterteilt, da einige Schüler, nach eigener Aussage das zur Verfügung stehende Feedback gar

nicht nutzten. Für welche Schüler dies zutraf, konnte man über folgende selbstkonstruierten Motivationstestitems herausbekommen:

FB09GRUPPE1	Ich habe das Feedback ausführlich gelesen.
FB10GRUPPE1	Mir war nur wichtig, ob ich richtig, oder falsch geantwortet habe – die genaue Rückmeldung war mir egal.

Diese Items wurden von den Schülern im Selbstreport während des zweiten Motivationstests beantwortet – die Auswertung dieser Items führte letztlich zu einer Unterteilung in folgende Subgruppen:

Treatment 1a: Das zur Verfügung gestellte, ausführliche Feedback wurde von den Schülern genutzt.

Treatment 1b: Ausführliches Feedback als Angebot, aber keine Nutzung der Rückmeldung durch die Schüler.

Treatment 2: Schrieb den gleichen Tests, wie die erste Treatmentgruppe, erhielt aber lediglich eine kurze Rückmeldung, welche der Antworten falsch und welche richtig waren.

Kontrollgruppe: Behandelt inhaltlich die gleichen Themen, wie die Treatmentgruppen, bearbeitet aber keine Testfragen, sondern liest ausführliche Texte. Ein spezifisches Feedback bekommen die Schüler nicht. Um auch bei dieser Gruppe eine einfach Rückmeldung zu erhalten, inwieweit sie mit den Lerninhalten und den bereitgestellten Texten zurecht gekommen sind, wurde nach jeder Textseite die kurze Kontrollfrage gestellt, ob die Lerner die Inhalte verstanden haben, oder noch Hilfe von der Lehrkraft benötigen würden. So konnte man einen ganz guten Einblick bekommen, inwieweit die unterschiedlichen Kategorien verstanden wurden.

Ich habe mich bei der Kontrollgruppe für das „Texte lesen“ entschieden, weil die Schüler so die Möglichkeit hatten, sich zeitlich und inhaltlich genauso ausgiebig zu beschäftigen, wie die anderen Gruppen. Es hätte sehr viele andere Designmöglichkeiten für die Kontrollgruppe gegeben.

Zum Beispiel hätte die KG den gleichen formativen Leistungstest, wie Treatment 1 und 2 schreiben, aber eine weitere Feedbackvariation erhalten können - zum Beispiel Rückmeldungen mit

Hilfe von Filmmaterial und Animationen anstelle von Text- und Bildmaterial, oder einfach gar kein Feedback.

Von diesen Möglichkeiten wurde Abstand genommen, weil man so (bei Bedarf) bei der Auswertung zusätzlich überprüfen könnte, ob der Test an sich schon etwas bringt, oder ausschließlich die Ausführlichkeit der Rückmeldungen einen Effekt auf den Leistungszuwachs hat. Alle Texte der Kontrollgruppe befinden sich im Dokument „Texte Kontrollgruppe“ (bei Interesse bei der PH Schwäbisch Gmünd anfordern).

Die Einteilung in die jeweiligen Gruppen erfolgte rein zufällig - zu Beginn der Studie erhielt jeder Schüler einen Zettel, der die Benutzerkennung, das Passwort und dessen Gruppe zeigte:

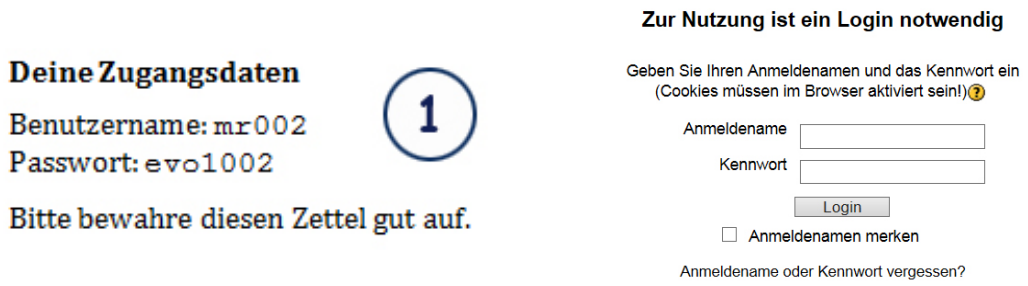


Abbildung 15: Zugangsdaten der Schüler und LogIn Screen von Moodle im Vergleich

Aus meiner Erfahrung heraus, sollte man bei den LogIn Informationen für die Schüler die gleiche Terminologie, wie beim Login der Moodle Seite verwenden, dadurch kann man die ab und zu auftretende Verwirrung auf Schülerseite, an welcher Stelle, welche Information eingetragen werden muss, von vornherein verhindern.

In einer gesonderten Excelliste wurden alle Kennungen der Studienteilnehmer aufgenommen und mit allen wichtigen Informationen (Passwort, Schule, Klasse, Gruppe, An-/Abwesenheit bei den unterschiedlichen Tests, etc.) versehen, damit die Möglichkeit bestand jederzeit (vor allem bei Verlust des Logins auf Schülerseite) nachsehen zu können, welches Login welcher Gruppe zugeteilt war.

Alle Testfragen und ihre dazugehörigen Antworten und Rückmeldungen wurden in den beiden Vorstudien und während des Testlaufs in den beiden Realschulklassen in Erlangen auf ihren Schwierigkeitsgrad, ihre Reliabilität und Validität hin überprüft und gegebenenfalls umformuliert, ersetzt, oder gestrichen.

5.3.2.4 Posttest 2 – Behaltenstest

Wie bereits angemerkt, wurde der Behaltenstest, ca. 6-8 Wochen nach Beendigung der Unterrichtssequenz in Paper - Pencil Form mit halboffenen Fragen durchgeführt.

Die halboffenen Fragen deshalb, weil man nur dadurch sehen kann, inwieweit die Schüler (vor allem die Konzepte „Evolutionäre Anpassung“ und „Auftrieb“) durchdrungen und verstanden haben. Sie müssen ihre Antworten selbst formulieren und durchdenken, fachliche Begrifflichkeiten in die richtige Beziehung zueinander setzen, also erlernte Zusammenhänge darstellen (Shavelson et al. 2008). Dies stellt eine ungleich größere Herausforderung dar, als lediglich die richtige, von der Lehrkraft vorformulierte Antwort im elektronischen Leistungstest herauszusuchen.

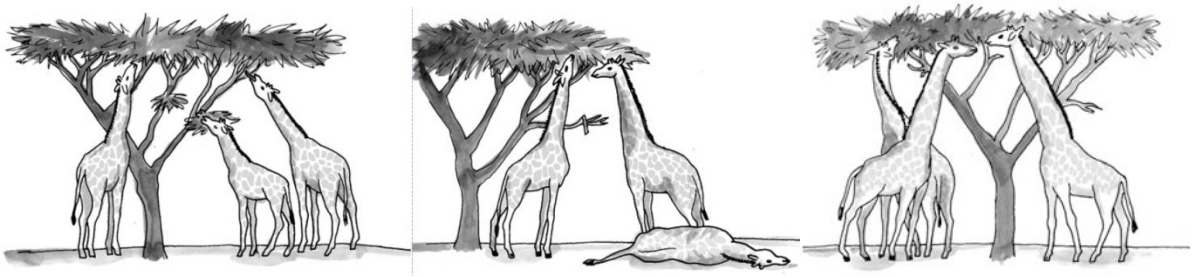
Natürlich erhöht sich hier, bei über 260 teilnehmenden Schülern der Korrekturaufwand immens, aber dafür bekommt man auch einen guten Überblick, wie lange und wie detailliert sich die Konzepte bei den Lernern im Gedächtnis verankert haben.

Die größte Schwierigkeit bei der Durchführung des follow-up Tests war die Motivation der Schüler. Es galt eine Möglichkeit zu finden, die Schüler so weit zu motivieren, dass sie sich überhaupt auf den Test einließen und auch vernünftige Antworten gaben. Schließlich ging es für die Lerner zu diesem Zeitpunkt um nichts mehr. Deshalb wurden für die besten Schüler der Klasse ein kleiner Preis und eine zusätzliche gute Biologienote in Aussicht gestellt. Dennoch gab es leider einige Arbeiten, bei denen nur Quatschantworten gegeben wurden oder gar nichts stand.

Damit sich die Schüler bei der Formulierung der Antworten etwas leichter taten, wurden kleine Übergänge in den Testbogen eingearbeitet. So hatten die Schüler bei der Beantwortung der Fragen einen kleinen Leitfaden. Ein kleines Beispiel dieser halboffenen Fragestellungen wird an dieser Stelle angeführt, den gesamten Behaltenstest kann man im elektronischen Anhang finden:

B) Evolutionäre Anpassung:

5. a) Beschreibe anhand der Bilder genau, wie evolutionäre Anpassung vor sich geht:



Bildmaterial aus Kattmann et al. (2009)

Evolutionäre Anpassung verläuft in folgenden Schritten:

_____

_____

_____

_____

_____

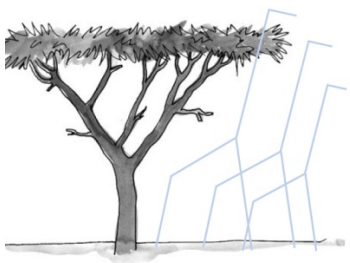
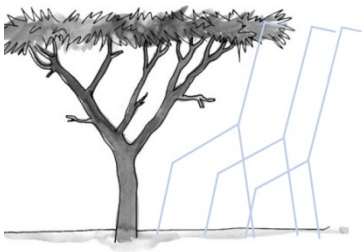
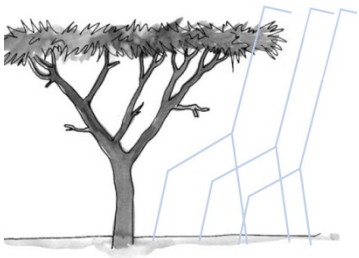
_____.

Evolution betrifft _____.

Anpassungsvorgänge dauern _____ und enden _____

_____.

b) Wie werden die Giraffen nach einigen weiteren Generationen aussehen? Begründe!

1	2	3
		

Bildmaterial aus Kattmann et al. (2009)

Nach vielen weiteren Generationen werden die Giraffen so ähnlich aussehen, wie auf Bild ____, weil _____

_____

_____.

Dies nennt man _____ - sie ist dafür verantwortlich, dass es

_____.

5.3.2.5 Motivationstests und Kurzskalen

In der Studie werden unterschiedliche psychologische Aspekte mit Hilfe der verschiedenen motivationalen Messinstrumente untersucht. So werden in den Kurzskalen die kurzzeitigen situativen emotionalen Zustände näher beleuchtet (State) und die überdauernden emotionalen Persönlichkeitsdispositionen („personality traits“), wie das Selbstkonzept, in den Motivationstests 1 und 2 erhoben. Die Unterscheidung zwischen „state“ und „personality traits“ stammt aus der Emotionspsychologie (vergleiche Hummel 2012).

Bevor näher auf die einzelnen Messinstrumente eingegangen wird, soll festgehalten werden, dass alle Motivationstests und Kurzskalen Likert - skaliert sind, d.h. sie bieten Ankreuzalternativen von 1 bis 5, wobei 1 für „trifft gar nicht zu“ und 5 für „trifft völlig zu“ steht. Bei der Beschreibung der Kurzskalen, wird der Fragebogen selbst dargestellt, um zu zeigen, wie man sich dessen Design vorzustellen hat (siehe unten).

Die **Kurzskalen** dienten dazu, die Qualität der Unterrichtssequenz und die Motivation der Schüler während einzelner Unterrichtsstunden festzuhalten. Um einen Überblick zu bekommen, ob die Lerner die Inhalte interessant und die methodische Umsetzung ansprechend fanden, wurden drei Stunden ausgewählt, die unterschiedlich komplexe Lerninhalte und damit einhergehend verschiedene methodische Unterrichtsformen beinhaltete. So fanden die 3 Kurzskalen nach einer Mikroskopierstunde zum Thema „Federn“, nach dem Lernzirkel mit vielen Schülerexperimenten zum Thema „Warum können Vögel fliegen“ und nach einer eher Lehrerzentrierten Stunde über die Thermik statt. Einen genauen Testzeitplan kann man unter dem Gliederungspunkt 4.2.4 „Durchführung des Feldexperiments“, finden.

Dabei stellte die jeweils in der Klasse unterrichtende Lehrkraft die Kurzskalen am Ende der entsprechenden Stunde zur Verfügung (Randler et al. 2011). Diese wurden von den Lernern ausgefüllt und abgegeben.

Die Kurzskala sah folgendermaßen aus:

Tabelle 16: Kurzskala

		trifft gar nicht zu				trifft völlig zu
01	Die Stunde hat mir Freude gemacht.	○	○	○	○	○
02	Ich fand das Thema wichtig.	○	○	○	○	○
03	Ich habe mich gelangweilt.	○	○	○	○	○
04	Ich war mit der Stunde zufrieden.	○	○	○	○	○
05	Ich war mit den Gedanken heute öfter woanders.	○	○	○	○	○
06	Der Unterricht hat mir Spaß gemacht.	○	○	○	○	○
07	Was ich über das Thema erfahren habe, bringt mir was.	○	○	○	○	○
08	Ich möchte mehr über das Thema erfahren.	○	○	○	○	○
09	Die heutige Stunde war zum Einschlafen.	○	○	○	○	○

Anhand der Befragungsergebnisse der drei Testzeitpunkte wurden für die Auswertung jeweils die Mittelwerte von Interesse, Langeweile und Wohlbefinden berechnet und zu folgenden Variablen zusammengefasst:

- State_Interesse
- State_Wohlbefinden
- State_Langeweile

(Randler et al. 2011)

Der **erste Motivationstest** fand, genau wie der Pretest, am ersten Erhebungszeitpunkt statt und beinhaltete vor allem Fragen zur Motivation und zum Selbstkonzept der Schüler.

Im folgenden Abschnitt befinden sich die Fragen, die für die Auswertung der Daten herangezogen wurden:

SKTA01	Ich traue mir im Fach Biologie viel zu.
SKTA02	Das Fach Biologie fällt mir nicht schwer.
SKTA03	Ich denke, dass ich im Fach Biologie immer alles schaffen werde.
SKTA04	Das Fach Biologie liegt mir nicht.
SKTA05	Für das Fach Biologie habe ich einfach keine Begabung.
SKTA06	Das Fach Biologie werde ich nie richtig verstehen.
SKTA07	Ich glaube, dass ich das Fach Biologie nie durchschauen werde.
SKTA08	Das Fach Biologie kann ich einfach nicht verstehen.

(Helmke 1992; verändert in Gläser-Zikuda et al. 2005)

Im Fach Biologie lerne und beteilige ich mich am Unterricht, ...

1Mo01	weil die Unterrichtsinhalte meinen Neigungen entsprechen
1Mo02	weil ich ein großes Interesse an den angebotenen Inhalten habe.
1Mo03	weil mir die Beschäftigung mit den Inhalten des Unterrichts Spaß macht.
1Mo04	weil es mir wichtig ist, ein umfangreiches Fachwissen zu besitzen.
1Mo05	weil es mir wichtig ist, meine fachlichen Fähigkeiten immer mehr zu erweitern.
1Mo06	weil ich ein hohes Maß an Kompetenz (allgemein anerkannter Sachverstand) erreichen möchte.

(Aus: Wild, Krapp, Schiefele, Lewalter, Schreyer 1995; verändert in Gläser–Zikuda et al. 2005)

Denke an das Fach Biologie im Allgemeinen:

INA01	Ich finde die Themen im Fach Biologie wichtig.
INA02	Die Themen im Fach Biologie bringen mir was.
INA03	Der Unterricht im Fach Biologie macht mir Angst.
INA04	Ich möchte mehr über die Themen im Fach Biologie erfahren.
INA05	Der Unterricht im Fach Biologie macht mir Spaß.
INA06	Der Unterricht im Fach Biologie macht mich nervös.
INA07	Ich bin mit dem Unterricht im Fach Biologie zufrieden.
INA08	Ich fühle mich im Biologieunterricht unter Druck.
INA09	Den Unterricht im Fach Biologie finde ich gut.
INA10	Der Unterricht im Fach Biologie macht mir Freude.
INA11	Der Unterricht im Fach Biologie interessiert mich.
INA12	Im Biologieunterricht beunruhigen mich einige Dinge.

(Aus: Gläser–Zikuda et al. 2005)

Diese Items lassen sich zu folgenden Variablen zusammenfassen, die vor Beginn der Datenauswertung daraufhin überprüft wurden, ob sie einen Einfluss auf das Treatment haben (siehe dazu Kapitel 6.2.2 und 6.2.5):

- Positives Selbstkonzept – SKTA01 bis SKTA03 und
- Negatives Selbstkonzept – SKTA04 bis SKTA08 (Helmke 1992)
- Intrinsische Motivation (Interesse) – 1Mo01 bis 1Mo03 und
- Intrinsische Motivation (Kompetenzorientiert) – 1Mo04 bis 1Mo06 (Wild et al. 1995)
- Instruktionsbezogene Angst (Trait) – INA03, INA06, INA08, INA12,
- Instruktionsbezogenes Interesse (Trait) - INA01, INA02, INA04, INA11 und
- Instruktionsbezogenes Wohlbefinden (Trait) - INA05, INA07, INA09, INA10 (Gläser-Zikuda et al. 2005)

Zusätzlich wurden Fragen zum sozioökonomischen Hintergrund der Schüler und zu deren schulischen Leistungen hinzugefügt (selbstkonstruierte Items), deren Auswertung unter 6.2.1 zu finden sind. Dabei mussten die Schüler kurze Angaben manuell eintippen, wie zum Beispiel ihr Geburtsjahr oder ihre Vornoten der Fächer Mathematik, Biologie und Deutsch, oder aus einem Drop Down Menü die passende Antwort auswählen.

Beim Beruf der Vater und der Mutter* wurde eine umfassende Liste an Berufen vorgegeben, die bei der Auswertung zu den Kategorien „nichtakademische“, „akademische“ Berufe und „Hausfrau/-mann, bzw. nicht berufstätig“ zusammengefasst wurden.

Tabelle 17: Fragen zum sozioökonomischen Hintergrund der Schüler (Motivationstest 1)

Bitte gib dein Geburtsjahr an:	
Geschlecht:	weiblich
	männlich
Bitte gib deine Note in Mathematik aus dem vergangenen Schuljahr an: (1 - 6)	
Bitte gib deine Note in Biologie/PCB aus dem vergangenen Schuljahr an: (1 - 6)	
Bitte gib deine Note in Deutsch aus dem vergangenen Schuljahr an: (1 - 6)	
Wurde dein Vater in Deutschland geboren?	ja
	nein
Wurde deine Mutter in Deutschland geboren?	ja
	nein
Was macht dein Vater beruflich?	*
Was macht deine Mutter beruflich?	
Welche Sprache wird bei dir zu Hause gesprochen?	nur deutsch
	deutsch und eine andere Sprache
	nur eine andere Sprache

Wie viele Bücher liest du pro Monat?	keines
	1 Buch
	Bis zu 3 Bücher
	mehr als 3 Bücher
Was liest du gerne?	Sachbücher
	Comics
	Spannendes (Abenteuer)
	Science Fiction / Fantasy
	Romantisches
Wie viele Bücher habt ihr zu Hause?	Keine
	Nur wenige
	Ca. ein Regalbrett voll
	Ein ganzes Regal voll
	Mehr als ein Regal voll

Im **Motivationstest 2**, der direkt nach dem Posttest durchgeführt wurde, blieben lediglich die ersten 6 Fragen zur Beteiligung im Biologieunterricht erhalten, um zu überprüfen, ob sich im Laufe der Studie irgendetwas in dieser Hinsicht bei den einzelnen Lernern verändert hat.

Er beinhaltete folglich die Erhebung folgender Persönlichkeitsmerkmale:

- Intrinsische Motivation (Interesse) – 2Mo01 bis 2Mo03
- Intrinsische Motivation (Kompetenzorientiert) – 2Mo04 bis 2Mo06

Diese Merkmale wurden mit Hilfe der gleichen Motivationsitems überprüft, wie im ersten Motivationstest. Weiter oben sind die einzelnen Items (1Mo01 bis 1 Mo06) genauer aufgeführt (Wild, Krapp, Schiefele, Lewalter, Schreyer 1995; zitiert in Gläser–Zikuda et al. 2005).

Alle anderen Fragen wurden durch neue, selbstkonstruierte und für die Studie interessante Fragestellungen zur Schülerpersönlichkeit, zur Unterrichtseinheit, zu den Tests, zum Umgang mit den Tablets und zum Feedback, ersetzt. Alle Items wurden von den Schülern im Selbstreport ausgefüllt.

Dabei wurde die Persönlichkeit der Schüler mithilfe des Konzeptes des Big Five (Costa & McCrae 1992) überprüft. Verwendet wurde eine 10-Item Kurzversion der Skala von Rammstedt & John (2007):

Tabelle 18: Big Five

E01	...bin eher zurückhaltend, reserviert.
A01	...schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.
C01	...bin bequem, neige zur Faulheit.
N01	...bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.
O01	...habe nur wenig künstlerisches Interesse.
E02	...gehe aus mir heraus, bin gesellig.
A02	...neige dazu, andere zu kritisieren.
C02	...erledige Aufgaben gründlich.
N02	...werde leicht nervös und unsicher.
O02	...habe eine aktive Vorstellungskraft, bin phantasievoll.

Die Skala selbst enthält je ein positives und ein negatives Item für jedes dieser Persönlichkeitsmerkmale:

- Offenheit – O01 und O02
- Extraversion – E01 und E02
- Gewissenhaftigkeit – C01 und C02
- Verträglichkeit – A01 und A02
- Neurotizismus - N01 und N02

Beim Big Five handelt es sich um ein stabiles Persönlichkeitskonzept (Rammstedt & John 2007), so dass der Erhebungszeitpunkt irrelevant ist, da sich innerhalb eines halben Jahres so gut wie keine Änderungen ergeben.

Die folgenden selbstkonstruierten Items sind ebenfalls Bestandteil des zweiten Motivationstests, wurden von den Schülern im Selbstreport beantwortet und einer Faktorenanalyse unterzogen (siehe 6.2.7 und 6.2.8):

Items, die die Unterrichtseinheit betreffen:

UE01	Die Inhalte fand ich spannend.
UE02	Ich habe mich gerne mit diesem Thema beschäftigt.
UE03	Auch zu Hause habe ich das Eine oder Andere zu diesem Thema nachgelesen.
UE04	Das Thema interessiert mich nicht.
UE05	Ich habe gerne etwas Neues zu diesem Thema erfahren.
UE06	Ich fand die Inhalte langweilig.

Items, die die Tabletnutzung betreffen:

FB01	Ich bin mit der Bearbeitung der Tests gut klar gekommen.
FB02	Der Umgang mit den Tablet - PCs hat mir Spass gemacht.
FB03	Ich hätte die Tests lieber auf herkömmlichem Weg mit Papier und Stift bearbeitet.
FB04	Ich hatte Schwierigkeiten, bei der Arbeit mit den Tablets.
FB05	Ich wusste nach den Tests, wo meine Wissenslücken sind.
FB06	Ich habe versucht, diese Wissenslücken zu schließen.
FB07	Durch die Wiederholung des Lerninhaltes war ich besser auf die Probe vorbereitet.
FB08	Die Vielzahl der Tests hat mich genervt.

Durch die Aufnahme dieser Items in den zweiten Motivationstest wurde versucht, etwas mehr von den Schülern zu erfahren, zum Beispiel, wie sie mit der Technik und den Tablets zurecht gekommen sind (fanden sie es motivierend, schwierig, etc. die Tests in elektronischer Form zu bearbeiten) und wie Interessant sie die Lerninhalte und die Fragestellungen vom Schwierigkeitsrad her einschätzten. Außerdem war von Interesse, ob und wie ausführlich sie das erhaltene Feedback gelesen haben und vor allem, ob sie die Tests zum Anlass genommen haben, auftauchende Wissenslücken, zu Hause zu schließen.

Leider war eine genaue Beobachtung der Schüler während der Bearbeitung der Tests selten möglich, weil ich ständig damit beschäftigt war, technische und inhaltliche Fragen zu klären.
Dennoch wurden einige Beobachtungen gemacht, auf die in Kapitel 6.2.9 näher eingegangen wird.

5.3.3 Feedback

Bevor man sich darüber Gedanken macht, wie das Feedback im Speziellen aussehen soll, muss man sich erst einmal darüber im Klaren sein, welche unterschiedlichen Dinge es beim Einsatz von Rückmeldungen zu bedenken gibt. In der folgenden Aufstellung soll kurz dargestellt werden, dass nicht nur die Formulierung des Rückmeldetextes einiger Überlegungen bedurfte:

- Wann möchte ich das Feedback geben (zeitliche Komponente)?
  Entweder gar nicht, oder direkt nach Beantwortung der jeweiligen Frage, oder am Ende des gesamten Tests oder erst in der nächsten Stunde (bei schriftlichen Leistungstests, wo erst einmal eine Korrektur nötig wird, bevor man eine gezielte Rückmeldung geben kann, gibt es nur diese Möglichkeit).
- Wer soll ein Feedback erhalten?

Der Lernende, der Lehrende, oder beide?

- Auf was soll die Rückmeldung Bezug nehmen?

  Lediglich auf die Aufgabe selbst (inhaltliche Probleme, Denkfehler) und die dazu gegebene Schülerantwort, oder auch in Hinblick auf den Lösungsweg (Kompetenzförderung, vor allem bei der Lösung mathematischer Probleme gut möglich, wo die Aufgabenstellung eines kleinschrittigen Lösungswegs bedarf), oder in Bezug auf die Person (Leistungszuwachs, Arbeitshaltung, Lernfortschritt) und deren Einbindung in die Lerngemeinschaft (wo stehe ich im Vergleich zu den anderen), oder den zukünftigen Unterricht der Lehrkraft betreffend (würde eine adaptive Gestaltung der nächsten Unterrichtseinheit ermöglichen).

- Wie ausführlich soll das Feedback gestaltet sein?

  Es besteht die Möglichkeit, die Antwort lediglich als falsch/richtig einzustufen, oder gleichzeitig aufzuzeigen, was falsch gemacht wurde und wie die richtige Antwort gewesen wäre, darüber hinaus könnte man noch Übungsaufgaben zum spezifischen Denkproblem des einzelnen Schülers geben (zum Beispiel in Mathe), oder sogar noch weiterführende Informationen zu angrenzenden Lerninhalten anbieten.

  Bei Feedback an den Lehrer: Benötigt die Lehrkraft nur eine kurze Statistik, eine genaue Aufstellung der Einzelergebnisse oder darüber hinaus vielleicht sogar eine Unterrichtsempfehlung und eine Bereitstellung weiterer Übungsaufgaben.

- Sollte man ein „Hilfsangebot“ zur Verfügung stellen?

  Schon bei der Beantwortung der Frage, könnte man den Lerner darauf hinweisen, dass er „in die falsche Richtung denkt“ und Hinweise geben. Dabei könnte man den Lerner auch wählen lassen, ob die Hilfe automatisch bei falschen Aktionen eingeblendet werden soll, die Hilfe optional hinzugeschaltet werden kann, oder gar kein Hilfsangebot zur Verfügung stehen soll (interaktives Feedback).

- Bei welcher Gelegenheit gibt man eine Rückmeldung?

  Nur bei falschen Antworten, oder nur bei richtigen Antworten oder generell immer?

- Wie verfügbar sollte das Feedback sein?

  Sollte es immer vollständig eingeblendet werden und somit immer zur Verfügung stehen, sollte man nur einen Hinweis bekommen, wo man die richtige Antwort finden kann, um die Arbeitshaltung und die Lerneffektivität beim Schüler zu steigern, oder sollten die Lerner gar keinen Hinweis erhalten und eigenverantwortlich nachlesen, was ihnen an Wissen zur richtigen Beantwortung der Frage fehlt?

- Welche Art von Rückmeldung ist effektiver?
  Sollte man das Feedback positiv, negativ oder neutral formulieren, und sollte man den Lerner individuell ansprechen oder in der Formulierung eher allgemein bleiben? Es wäre interessant zu untersuchen, ob sich die Schüler eine negativ, oder eine positiv formulierte Rückmeldung besser merken können, bzw. durch welche Art von Feedback sie sich mehr angespornt fühlen. Doch die emotionale Komponente der Formulierung sollte in der Studie nicht untersucht werden, sondern eine Variation im inhaltlichen Bereich.
- In welcher Form sollte das Feedback angeboten werden?
  Neben der schriftlichen Darbietung der Rückmeldungen während der Tests, wäre es noch möglich, den Schülern nach der Auswertung der Arbeiten durch den Lehrer, zusätzlich ein mündliches Feedback zu geben, also die unklaren Punkte mit den Schülern selbst oder häufig auftretende Probleme im Klassenverband zu besprechen. Auch während des normalen Unterrichtsgespräches, Abfragen oder Gruppenarbeiten, kommt es zu mündlichen Rückmeldungen zu Schülerantworten, sowohl von Lehrerseite aus, als auch von Seiten der Mitschüler. Diese Rückmeldungen werden während der Studie nicht gezielt forciert, weil sie die Unterschiede im Treatment verwischen könnten. Sie sind aber auch nicht zu vermeiden, weil es während des Unterrichts natürlich immer wieder Rückmeldungen auf Schülerfragen und –antworten geben wird.

Dadurch, dass bei der Umsetzung und Durchführung der Studie Moodle als Lernplattform und Autorensystem gewählt wurde, gibt es eine, bereits durch die Software vorgegebene Einschränkung, was den Einsatz von Feedback angeht:
Es besteht die Möglichkeit, den Schülern direkt nach Beantwortung der Fragen, eine Rückmeldung zu jeder einzelnen Antwort zu erstellen und zusätzlich anzugeben, wie die komplett richtige Antwort ausgesehen hätte. Zur besseren Veranschaulichung oder Vertiefung des jeweiligen Lerninhalts, kann man auch Bilder bei der Rückmeldung einbeziehen. Die Einbindung anderer externer Medien wie Filme oder Animationen (würde sich vor allem bei der Demonstration des Auftriebs anbieten), ist leider nicht vorgesehen. Auch kann man keine, auf den einzelnen Schüler und dessen Denkfehler zugeschnittenen, weiterführenden Übungsaufgaben adaptiv zuschalten, weil Moodle das nicht anbietet. Ein adaptives System müsste man selbst programmieren und wäre für diese erste Studie zu umfangreich gewesen. Außerdem benötigt die Erarbeitung und Abfrage des Lerninhalts der zugrundeliegenden Unterrichtseinheit überhaupt kein adaptives System.

Anders wäre es, wenn man den Schülern einen mathematischen Sachverhalt, wie zum Beispiel, das Lösen von Gleichungen näherbringen wollen würde: Da kann man die Aufgaben in viele kleine Lösungsschritte unterteilen, die jeder für sich ein kleines Teilproblem des Lösungsweges beinhaltet (Klammern auflösen, Missachtung von Rechenregeln, Vorzeichen, Äquivalenzumformung, etc.). Diese Kleinschrittigkeit würde es erlauben, ein Teilproblem (zum Beispiel „Klammern auflösen“) zu isolieren und dem Schüler vor allem Übungsaufgaben zur Bewältigung dieser Problematik anzubieten.

Die Rückmeldungen, die man für die Antworten formuliert hat, kann man nicht nur direkt nach Beantwortung der jeweiligen Frage einblenden lassen, sondern auch an anderen Testzeitpunkten: Nach der Bearbeitung des gesamten Tests und auch noch (einmal) nach Abgabe des Tests. Allerdings sind dies immer Optionen und kein Muss. Bei den formativen Leistungstests der Studie wurde den Schülern gleich direkt nach Beantwortung der jeweiligen Frage, ein Feedback zu ihren Antworten gegeben.

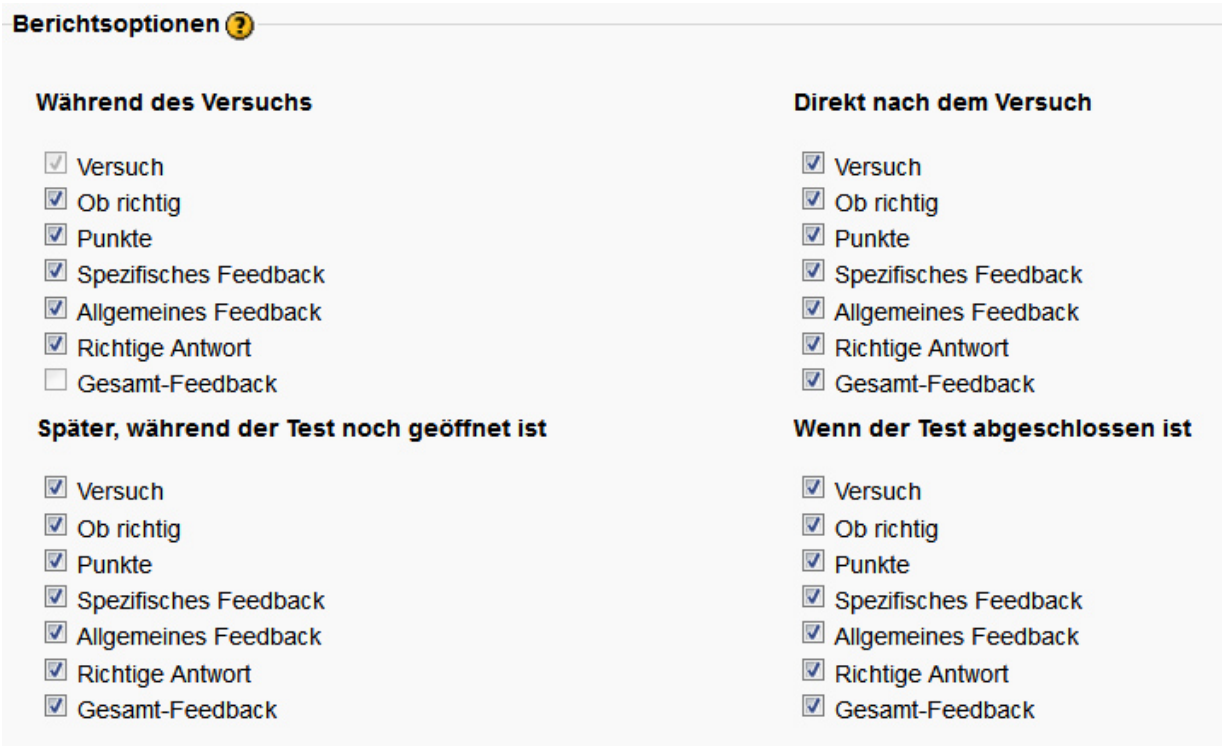


Abbildung 16: Auswahl der Feedbackoptionen in Moodle

Dabei erhielt die Treatmentgruppe 2 nur die kurze Rückmeldung, welche Teilantworten sie richtig, welche sie falsch beantwortet hatten, aber nicht, wie die richtige Antwort gelautet hätte und auch nicht, was sie im Detail falsch gemacht hatten. Die Schüler der Treatmentgruppe 1 hingegen, erfuh-

ren ausführlich, was sie richtig bzw. falsch gemacht hatten, welche einzelnen Antworten richtig gewesen wären und warum, und außerdem ab und zu ein paar Zusatzinformationen. Diese Art des direkten Feedbacks wurde erst durch den Einsatz elektronischer Leistungstests ermöglicht.
Außerdem wurde den Lernern beider Treatmentgruppen die Möglichkeit gegeben, ganz am Ende, nach Abgabe des Tests, noch einmal alle Fragen und deren Feedback einzusehen.
Außerdem bekamen diese beiden Gruppen zu diesem Zeitpunkt zusätzlich eine, an die jeweilige Person gerichtete Rückmeldung, wie gut diese im Test gearbeitet du abgeschnitten hat (dieses Feedback wurde automatisch erzeugt und war abhängig von der erreichten Gesamtpunktzahl des Tests).

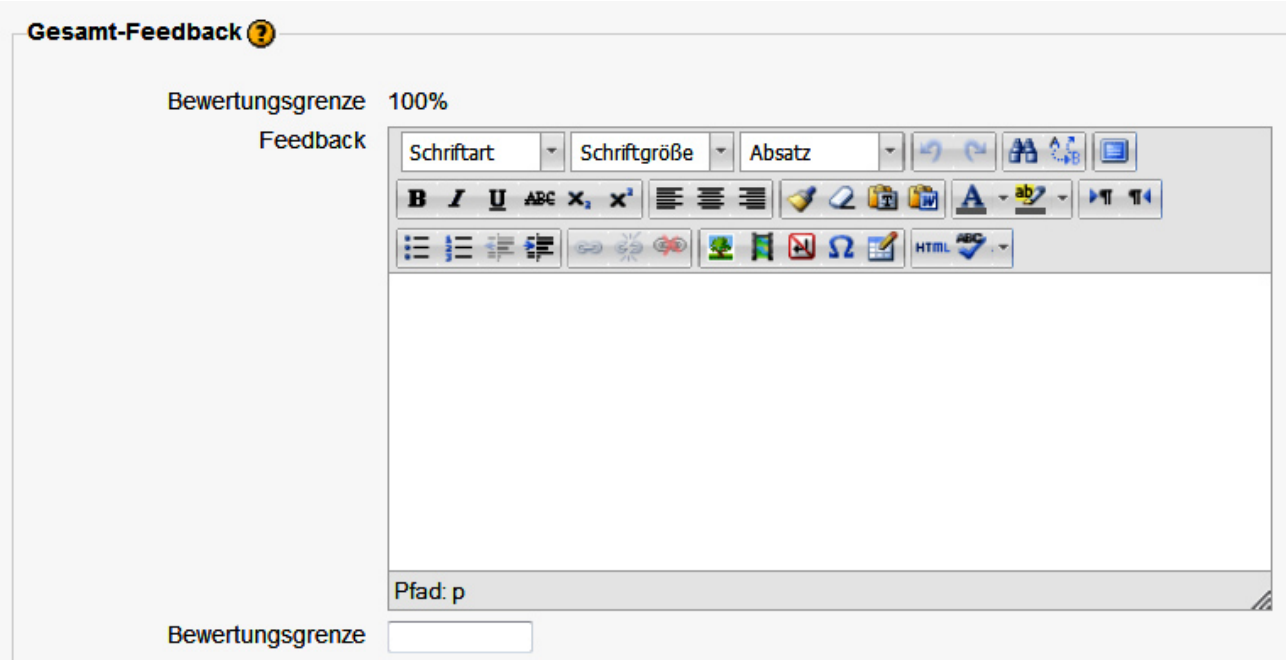


Abbildung 17: Eingabemaske für das Gesamtfeedback in Moodle

Von weiteren Rückmeldungen wurde Abstand genommen, weil es erstens zu verwirrend geworden wäre und sich die Schüler zweitens auf die wichtigsten Rückmeldung konzentrieren sollten (inhaltlich und auch zeitlich).
Leider gibt es in Moodle nicht die Möglichkeit, den Schülern einen Überblick über ihren Lernzuwachs (zum Beispiel anhand einer Lernkurve) zu geben. In der Vorstudie fanden es die Lerner sehr motivierend, ihren Lernverlauf (siehe Abbildung 22: Dokument zum Eintragen des individuellen Lernfortschrittes (Vorstudie)) in ihrem Schülerordner einzutragen. Aber in der Vorstudie wurden auch in allen Tests die gleichen Fragen bearbeitet, so dass ein direkter Vergleich zum jeweils vorhergehenden Testzeitpunkt möglich wurde. Dies ist bei den Tests der eigentlichen Studie nicht

mehr der Fall, so dass es schwierig geworden wäre, eine richtige Bezugsgröße zu finden, nach der man sich bei der Berechnung einer Lernkurve hätte richten können.

Auf den zeitlichen Einsatz der Rückmeldungen wurde bereits eingegangen, in den folgenden Abschnitten sollen die anderen Feedbackentscheidungen näher erläutern werden – etwas über den theoretischen Hintergrund, kann man in Kapitel 2.2.2 nachlesen.

Da Moodle keinen adaptiven Einsatz von Rückmeldungen erlaubt, wurde eine Standardisierung des Feedbacks nötig: Alle Rückmeldungen waren gekoppelt an der Beantwortung der Frage bzw. der erreichten Gesamtpunktzahl.
Der Lernfortschritt der Schüler wurde nicht über die Bearbeitung der Tests hinaus ausführlich mit Hilfe von Selbstevaluationsbögen oder Lerntagebücher begleitet, da das zu komplex geworden wäre und das eigentliche Ziel der Studie, die Variation des Feedbacks, zu untersuchen, aus dem Fokus gedrängt hätte. Mir war wichtig, dass sich die Schüler, die durch die vielen Tests bereits stark eingebunden waren, auf die Lerninhalte konzentrieren.
Eine Kompetenzförderung durch Rückmeldung wurde nicht explizit angestrebt, vielmehr sollte es eine Art „Nebenprodukt“ sein, dass die Schüler ihr Lernverhalten wenigstens in Kleinigkeiten verändern, wenn sie auf die Rückmeldungen eingehen und eventuell erkannte Wissenslücken, selbstständig (aus der Motivation heraus, das nächste Mal besser abzuschneiden) zu schließen versuchen. Derartige Änderungen im Lernverhalten benötigen sehr viel mehr Zeit, als es die Studie zugelassen hat.

In den folgenden Abschnitten soll auf die praktische Umsetzung der Feedbacktheorien (siehe 2.2.2) eingegangen und genauer beschrieben werden, welche Arten von Rückmeldung eingesetzt und wie diese genutzt wurde.

5.3.3.1 Feedback an den Lerner

Das Feedback an den Lerner und dessen Variation steht im Mittelpunkt der Studie. Wie bereits erläutert, erhält die erste Treatmentgruppe eine ausführliche Rückmeldung zu ihren Antworten, egal, ob diese richtig oder falsch beantwortet wurden.

Folgender Screenshot zeigt ein Beispiel, wie dieses ausführliche Feedback aussehen kann:

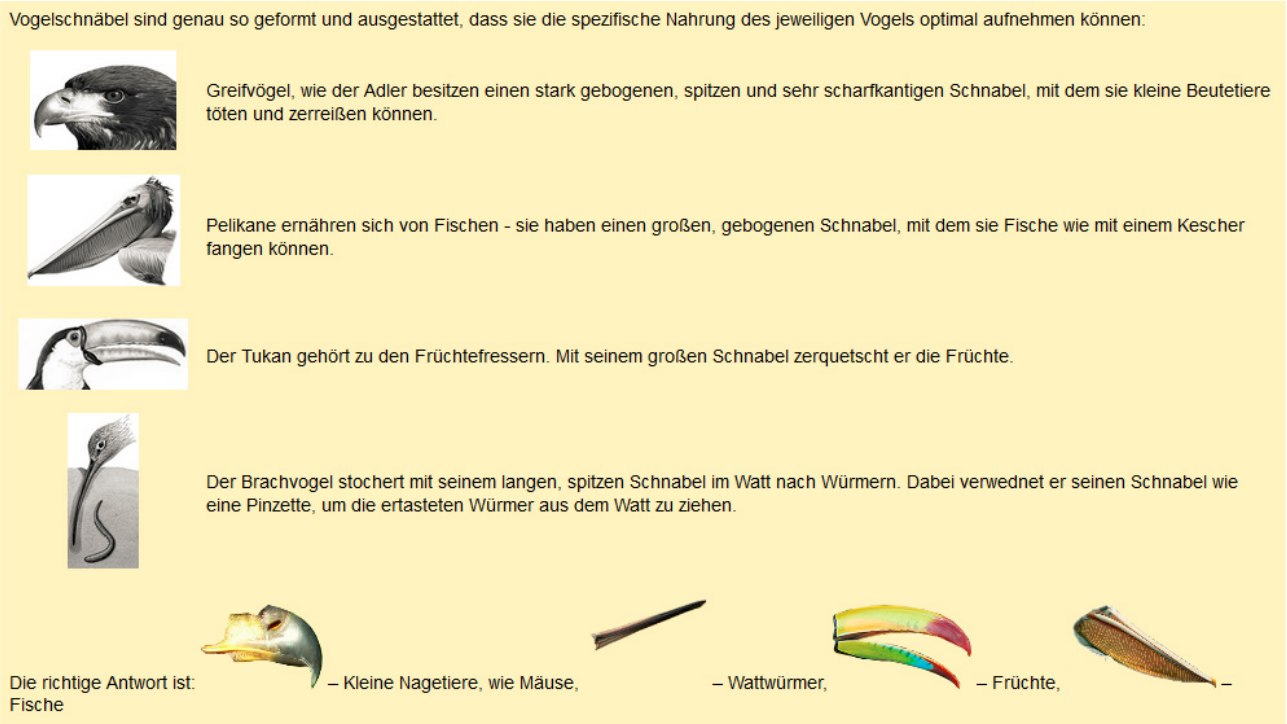


Abbildung 18: Weiteres Beispiel für das ausführliche Feedback der Treatmentgruppe 1

Anmerkung zur Verfügbarkeit des Feedbacks: Das Feedback ist direkt nach der Beantwortung der einzelnen Fragen und noch einmal nach der Abgabe des Gesamttests für jeden Schüler verfügbar und jederzeit (auch mehrmals) einsehbar. So ist es den Lernern überlassen, an welchem Zeitpunkt des Tests und für wie lange sie sich, damit beschäftigen. So soll sichergestellt werden, dass die Rückmeldungen umfassend und klar formuliert sind und alle Schüler der Gruppe nach Bearbeitung des Tests auf dem gleichen Stand sein können. Würden nur Hinweise gegeben werden, wo die nötigen Informationen nachgelesen werden können, würden wahrscheinlich noch viel weniger Schüler die Möglichkeit nutzen, durch Nachlesen ihre Wissenslücken zu schließen. Derartige Veränderungen im Lernverhalten (Wissenslücken erkennen, Informationen zur Beantwortung selbständig recherchieren) müssen lange eingeübt und verinnerlicht werden und sind nicht Ziel der Studie. Hier soll geschaut werden, ob ein ausführliches Feedback eine Wirkung auf die Schülerleistung und den Wissenszuwachs hat, oder nicht.

Hätte man eine Lernplattform, der eine adaptive Software zugrunde liegt, wären die Möglichkeiten, die Rückmeldeformate und noch zusätzlich die angebotenen Fragen zu variieren, ungleich höher: Man könnte bereits direkt nach der Bearbeitung des Pretests auf das individuelle Vorwissen der Schüler eingehen und sogar die Fragen in den darauffolgenden formativen Leistungstests au-

tomatisch so zusammenstellen lassen, dass sie genau die individuellen Denkfehler und inhaltlichen Probleme des vorhergehenden Tests aufgreifen und so, eine auf das Wissen des einzelnen Schülers individuell zugeschnittene Prüfung präsentieren. Allerdings bringt ein derartiges System einen riesigen Programmieraufwand mit sich, den nur ein Softwareentwickler adäquat umsetzen kann. Spannend wäre die Entwicklung einer derartigen Lernplattform, doch auch wenn die Programmierung von einem Experten übernommen würde und in ein Content Management System übertragen würde, so dass es auch für „Laien“ nutzbar würde, bliebe es schwieriger, die Lernplattform zu pflegen und würde bei den Lehrkräften einen höheres Maß an informationstechnischem Wissen voraussetzen.

Ein verbales Feedback während des Unterrichtsgesprächs mit der ganzen Klasse also eine Adaption des Unterrichts durch den Lehrer wird während der Testphase nicht angedacht und umgesetzt, weil es die zu untersuchenden Effekte gefährden kann.

Die Rückmeldetexte wurden neutral formuliert, denn der Fokus der Untersuchung sollte ja auf einem inhaltlichen Aspekt liegen und nicht darauf, inwieweit positive bzw. negative Formulierungen die Leistung und Schülermotivation beeinflussen.
Beim Gesamtfeedback, das nach der Abgabe der Tests angezeigt wurde, wurde hingegen ein persönlicherer Ton gewählt, der die Schüler direkt ansprechen und dadurch motivieren sollte, beim nächsten Mal noch besser abzuschneiden.

5.3.3.2 Feedback in Bezug auf die Aufgabe

Neben dem Lerner stehen die Aufgaben an sich im Fokus des Interesses. Wie bereits mehrfach beschrieben, sind die Klassen während der formativen Leistungstests in 3 Gruppen unterteilt und bekommen nach der Beantwortung der einzelnen Fragen ein unterschiedliches Feedback:

Das Feedback der zweiten Treatmentgruppe, die lediglich weiß, welche der Antworten richtig bzw. falsch beantwortet wurden, aber nicht, wie die richtige Antwort ausgesehen hätte sieht folgendermaßen aus:

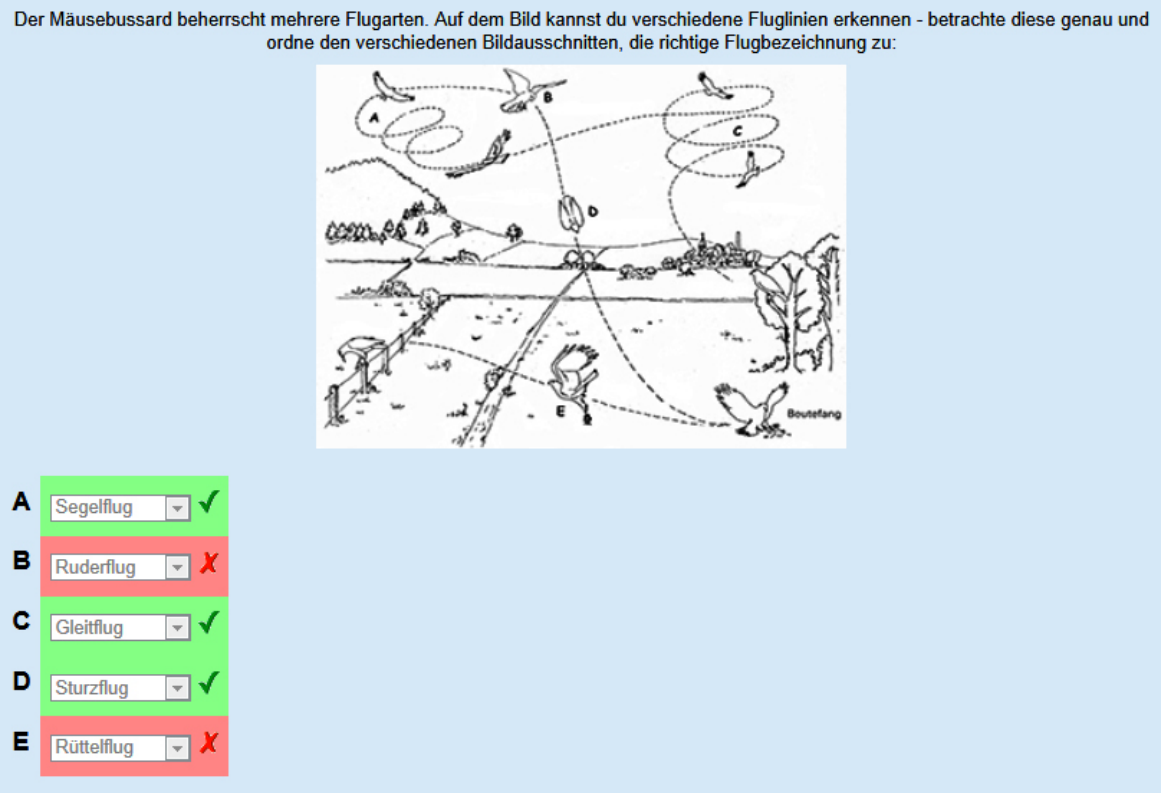


Abbildung 19: Dichotomes Feedback der Treatmentgruppe 2

Die erste Treatmentgruppe erhält neben der Rückmeldung, welche Antwort richtig und welche falsch war, ein ausführliches Feedback: Die komplett richtige Antwort und Zusatzinformationen zu allen Antwortmöglichkeiten. Auch werden Denkfehler gezielt angesprochen und berichtigt. Dies sollte vor allem, wenn es um die Erarbeitung konzeptuellen Wissens geht von Vorteil sein, weil dadurch die richtige Fachterminologie und die Zusammenhänge der Begrifflichkeiten innerhalb des gesamten Konstruktes immer wieder wiederholt und gefestigt werden.

Die Kontrollgruppe wiederholt die Lerninhalte anhand von Texten, die mit einer kurzen Nachfrage zur Selbsteinschätzung und zum Textverständnis abschließen. Die Schüler sollen nach jedem Kapitel reflektieren, ob sie den Text an sich und dessen Inhalt verstanden haben, oder ob es noch Schwierigkeiten beim Verständnis der Lerninhalte gibt. Diese Frage wurde gestellt, damit bei der Endauswertung auch von diesen Schülern eine Auskunft bereitsteht, welche Inhalte sie zum Zeit-

punkt der Tests verstanden wurden, bzw. wo es noch Probleme gab. So kann der Wissensstand dieser Schüler, wenigstens oberflächlich, mit dem der Kinder aus den Treatmentgruppen verglichen werden.

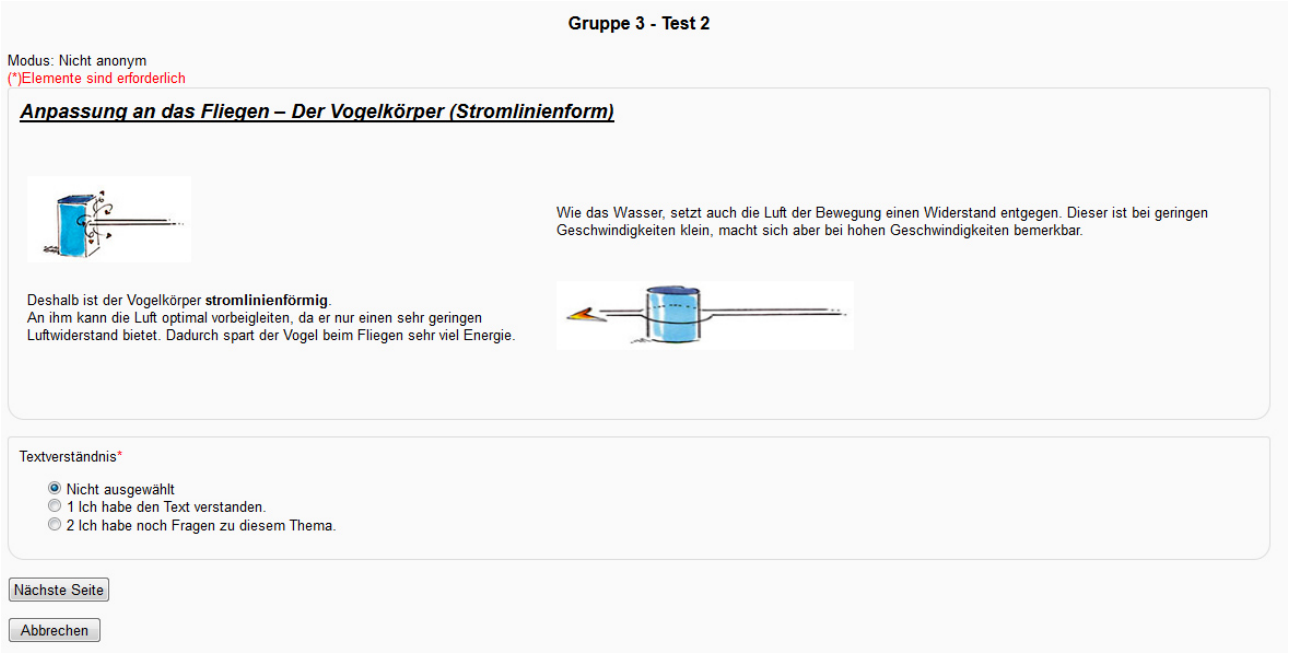


Abbildung 20: Textbeispiel der Kontrollgruppe beim zweiten Messzeitpunkt

Alle Texte der Kontrollgruppe befinden sich im elektronischen Anhang.

5.3.3.3 Feedback in Bezug auf den Lösungsprozess

Wie bereits beschrieben, war es in der Studie schwierig, Rückmeldungen zum Lösungsprozess der Aufgaben zu geben. Dies wurde bedingt durch den Einsatz von Moodle, die Struktur des Faches Biologie und das Thema der Unterrichtssequenz:

Moodle ist leider weder als Lernplattform, noch als Autorensystem, oder Content Management System dafür ausgelegt, adaptive Aufgaben zu generieren, oder zu unterstützen, das heißt, es besteht keine Möglichkeit, Testaufgaben in kleine Teilprozesse zu unterteilen, die sich gegenseitig bedingen, um anhand derer ein Problemlöseverständnis auf Seiten der Schüler zu überprüfen und im Anschluss daran, notwendige Übungsaufgaben zum spezifischen Lösungsproblem für jeden einzelnen Lerner individuell vorzugeben.

Weiter boten sich die Lerninhalte der Unterrichtseinheit nicht besonders gut an, um sie in einzelne Schritte zur Problemlösung zu unterteilen, bzw. immer wiederkehrende Teilprobleme bei der Lösung einer Gesamtaufgabe zu isolieren.

Der Großteil der Inhalte besteht aus Faktenwissen, welches man entweder gelernt hat, oder nicht. Nur ein Teil der Unterrichtssequenz dient der Erarbeitung von konzeptuellen Wissen, das wiederum den Schülern zwar ein genaues Verständnis fachlicher Zusammenhänge näherbringen möchte, aber eben nicht, wie in Mathematik oder Physik, Aufgaben bereit hält, die in kleinen (immer gleichen), aufeinander aufbauenden Schritten zu lösen sind. Dies kann man auch gut an folgender Concept Map zum Konstrukts „Auftrieb“ genauer sehen:

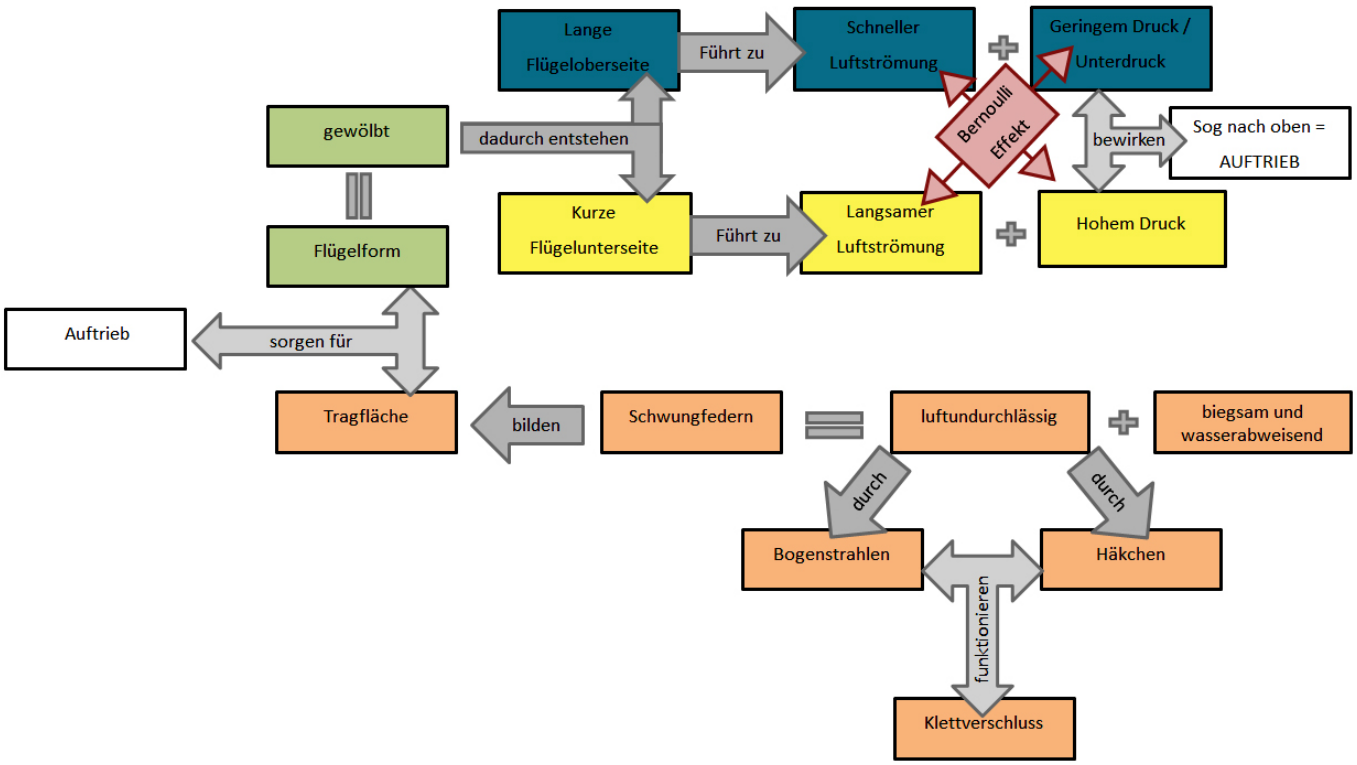


Abbildung 21: Concept Map zum Auftrieb

Man kann beobachten, dass das Zusammenspiel vieler einzelner Komponenten (gewölbte Flügelform, Luftundurchlässigkeit der Tragfläche) nötig ist, damit der Auftrieb funktionieren kann. Die Schüler erlernen während der Unterrichtssequenz, inwieweit sich diese verschiedenen Eigenschaften gegenseitig bedingen und wie sie zusammenarbeiten. Sie erfahren auch, was passieren könnte, wenn einer der Eigenschaften nicht vorhanden wäre, oder plötzlich verändert würde, aber sie berechnen nicht im physikalischen Sinn den Auftrieb, also wie er aufgrund des Neigungswinkels der Tragfläche, oder mit zunehmender Geschwindigkeit des Flugobjektes zu oder abnimmt. Doch genau hier könnte man wieder ansetzen und sehen, inwieweit der Lösungsprozess einer Aufgabe verstanden wurde.

Anders als in diesem Beispiel, gibt es zum Beispiel in Mathematik bei vielen Aufgabenarten eine Systematik, nach der man bei der Lösung dieses Aufgabentyps immer wieder vorgehen kann. Ich versuche dies am Thema „Terme vereinfachen“ kurz zu umreißen:

Tabelle 19: Lösungsprozess einer Mathematikaufgabe zum Thema "Terme vereinfachen"

Zugrundeliegende Rechenregeln	Rechenschritte	Mögliche Fehlerquellen
Klammern zuerst, von innen nach außen.	Schritt 1: Klammern auflösen.	- Vorzeichen oder - Regeln beim Ausmultiplizieren von Klammern werden missachtet.
Potenz, vor Punkt vor Strich.	Schritt 2: Rechenregel beachten.	- Potenzregeln - Schüler führen die Addition oder Subtraktion vor der Multiplikations- oder Divisionsaufgabe durch. - Vorzeichen nicht beachtet
Term vereinfachen	Schritt 3: Gleichnamige Terme erkennen und addieren bzw. subtrahieren	- Auch nicht gleichnamige Terme werden zusammengefasst - Vorzeichen werden missachtet

Wie man der Tabelle entnehmen kann, gibt es genügend Fehlerquellen, die man isolieren kann. Hat ein Schüler immer nur Schwierigkeiten mit den Potenzen, dann braucht er gezielt Unterstützung bei und Übungsaufgaben zu dieser Problematik, denn den Rest der Aufgabe beherrscht er ja. Hier kann man als Lehrer (oder auch das Computerprogramm) zielgerichtet ansetzen und die eigentliche Problematik angehen.

So wäre also denkbar, mit Hilfe eines geeigneten Computerprogrammes und in Fächern wie Mathematik, Physik und teilweise Chemie, bestimmte Aufgabentypen in einzelne Lösungsschritte zu unterteilen und den Schülern individuelle Übungsaufgaben zum jeweiligen problembehafteten Schritt innerhalb des Lösungsprozesses anzubieten.

Auch bei Leuders (2004, 2006), Büchter & Leuders (2005), Blum et al. (2006) und Sjuts (2006, 2007), findet man die Feststellung, dass „Ein informatives Feedback für Lehrkräfte und Schüler nur dann möglich ist, wenn die vorangehende Aufgabe auch Einblicke in den Lern- und Verstehensprozess des Lernenden eröffnet."

Daher denke ich, dass man mit Hilfe eines adaptiven und wirklich individuellen Systems und durch eine gezielte Rückmeldung bei der automatisierten, elektronischen Auswertung an die Lehrkraft und die damit einhergehenden Veränderungen des darauf folgenden Unterrichts die größten Effekte in Hinsicht auf die Verbesserung der Schülerleistungen erzielen könnte.

Und nicht nur in Hinblick auf die Verbesserung der Schülerleistung an sich, wäre ein adaptives System hilfreich, sondern auch in Hinblick auf die Veränderung der Arbeitshaltung einzelner Schüler. Durch weiterführende Links könnte man die Schüler an geeigneter Stelle auf Zusatzangebote aufmerksam machen. Ist dieses Zusatzmaterial durchgängig interessant und veranschaulicht den Lerninhalt nochmals auf vertiefende Art und Weise (Animation eines komplexen Vorgangs), nutzen die Lerner dieses Angebot mit der Zeit gerne und sind vielleicht auch eher gewillt, ihre Arbeit mit dem angebotenen Feedback zu verbessern und dieses öfter zu nutzen.

5.3.3.4 Feedback in Bezug auf die Person

Eigentlich war angedacht, einen Lernverlaufsbogen, in dem die Schüler ihren Wissensstand zu jedem Testzeitpunkt eintragen konnten, zu entwickeln und zur Beobachtung des individuellen Lernfortschritts der Schüler heranzuziehen. In der Vorstudie kam deshalb auch folgendes Dokument zum Einsatz:

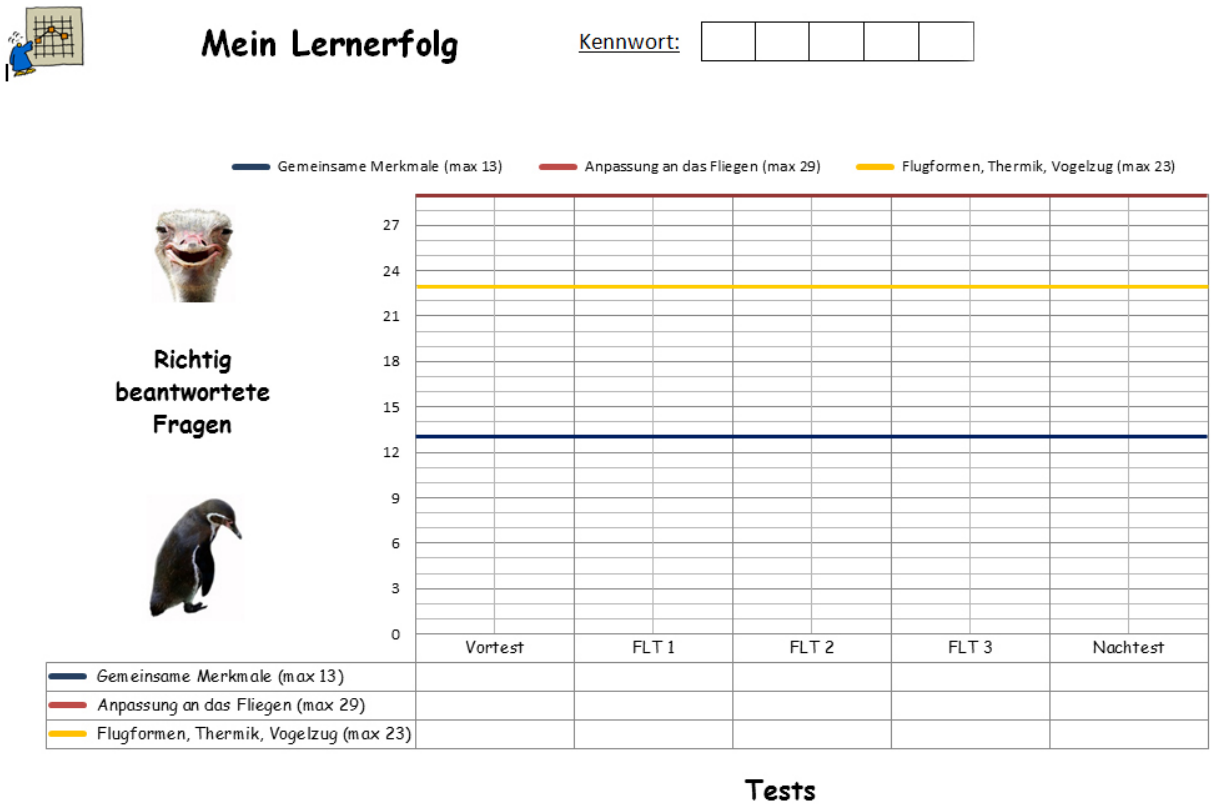


Abbildung 22: Dokument zum Eintragen des individuellen Lernfortschrittes (Vorstudie)

Zu diesem Zeitpunkt der Studie war der Einsatz eines solchen „Lerntagebuchs“ noch möglich, weil alle formativen Leistungstests die gleichen Fragen beinhalteten und auch der Pre- und Posttest durch die gleiche Bepunktung genau auf die formativen Leistungstests abgestimmt waren. So wurde ein direkter Vergleich von Testzeitpunkt zu Testzeitpunkt ermöglicht. Um eine noch differenziertere Aussage zum Lernzuwachs treffen zu können, wurden die Lerninhalte in verschiedene Kategorien unterteilt:

- Gemeinsame Merkmale (Schnabel- und Fußformen, Federarten)
- Anpassung an das Fliegen
- Flugformen, Thermik und Vogelzug

Diese Kategorien wurden nach der Vorstudie an die Änderungen innerhalb der Unterrichtssequenz für die eigentliche Studie neu angepasst. Dort dienten sie auch nicht mehr dazu, den Schülern eine individuelle Rückmeldung zu ihrem Wissenszuwachs zu geben, sondern sollten bei der Gesamtauswertung eine grobe Unterscheidung zwischen Faktenwissen (Gemeinsame Merkmale, Flugarten und Thermik), konzeptuellem Wissen (Evolutionäre Anpassung) und einer Mischform aus beiden (Anpassung an das Fliegen) ermöglichen. In SPSS wurden die Auftriebsfragen (Konstrukt innerhalb der Anpassungsmerkmale) noch einmal gesondert gekennzeichnet.

Ich konnte während der Vorstudie beobachten, dass die Arbeit mit dem Lernverlaufsbogen für die Schüler sehr motivierend war. Sie nahmen den Eintrag der erreichten Punkte nicht nur zum Anlass ihren eigenen Wissenszuwachs zu vergleichen, sondern sich gleichzeitig auch mit ihren Mitschülern zu messen (soziale Bezugsnormorientierung). Leider machte das veränderte Design der Tests innerhalb der Studie (unterschiedliche Fragen, elektronische Leistungserhebungen) den Einsatz eines Lernverlaufbogens schwierig, weil die Umstrukturierung der formativen Leistungstests keinen direkten Vergleich der Testzeitpunkte (und der Kategorien) mehr.

Die Schüler bekamen von Moodle aber angezeigt, wie viele Punkte sie absolut erreicht hatten, und wie viel Prozent diese Punkte in Bezug auf die zu erreichende Gesamtpunktzahl bedeuteten. So war es immerhin auf diesem Weg möglich, den individuellen Wissenszuwachs zu verfolgen und sich mit den Mitschülern zu vergleichen, was auch gerne gemacht wurde.

Denkbar wäre es auch gewesen, den Schülern eine Art Beurteilungsbogen zukommen zu lassen, auf dem gezielt seine inhaltlichen Stärken und Schwächen noch einmal aufzeigt werden, aber durch die manuelle Auswertungen aller Detailantworten jeden Schülers, wäre dieses Vorgehen sehr aufwendig gewesen. Außerdem sollten die Schüler lernen, schon anhand der Rückmeldungen eigene Rückschlüsse auf ihre Schwächen ziehen und im Optimalfall versuchen gerade diese Wissenslücken in Eigenverantwortung zu schließen. Diesen letzten Schritt kann man den Lernern auch durch das Verteilen eines Beurteilungsbogens nicht abnehmen.

5.3.3.5 Feedback an den Lehrenden zur Optimierung der Lernprozesse

Neben dem Feedback in Bezug auf den Lösungsprozess von Aufgaben (Gliederungspunkt 5.3.3.3) sehe ich hier die größte Wahrscheinlichkeit, Effekte in Hinblick auf den Lernzuwachs der Schüler zu erzielen.

Bisher „gelingt die Diagnose von Schülerkompetenz nur unzureichend“ (Harlen & Winter 2004; Shavelson & Ruiz-Primo 1999) und durch computergestützte Testverfahren könnte nicht nur der Lernzuwachs der Schüler über einen längeren Zeitraum, sogar bezogen, auf einzelne isolierte Themen, beobachtet, sondern auch die Tiefe des konzeptuellen Verständnisses besser erfasst werden. Die Forderung nach neuen Diagnosemöglichkeiten findet sich in der Literatur häufig. Tests sollen nicht mehr nur zur Leistungsbeurteilung der Schüler herangezogen werden, sondern auch dazu dienen, Schwächen zu diagnostizieren und Empfehlungen für den anschließenden Lehrprozess zu generieren. „Vor allem im naturwissenschaftlichen Bereich werden neue Diagnosemethoden nötig“ (Keogh & Naylor 2000; McConnell et al. 2006; Harlen 2007; Brown 2008; Tan & Towndrow 2009), Diagnosemethoden, die tiefergehend sind, als: Der Schüler konnte das auswendiggelernte Faktenwissen abrufen und zu Papier bringen. Denn durch die fachspezifischen Arbeitsweisen in den naturwissenschaftlichen Fächern verlangen diese auch eine andere Herangehensweise an die Themen Diagnostik, Leistungsmessung und Feedback, als zum Beispiel Fächer wie Geschichte und Deutsch.

Durch die elektronische Auswertung der formativen Leistungstests wird eine unmittelbare Rückmeldung an die Lehrkraft möglich, z.B. wo genau die Verständnisprobleme und Schwächen der einzelnen Schüler und der gesamten Klasse liegen. Möchte die Lehrkraft eine Rückmeldung zum Wissensstand der Schüler haben, kann sie in den Detailantworten der Schüler nachsehen, welche Fragen, wie beantwortet wurden und dadurch Unklarheiten und Probleme aufdecken, um diese im folgenden Unterricht aufzugreifen und zu besprechen (adaptiver Unterricht).
Auch einzelne Schüler können in ihrem Lernfortschritt beobachtet und bezüglich ihrer Defizite besser und individueller betreut werden.

Auch Helmke, Hosenfeld & Schrader (2004); Lorenz (2005); Klieme & Leutner (2006); Koeppen et al. (2008) fordern „eine Verbesserung und Stärkung der formativ-diagnostischen Kompetenz von Lehrkräften“ und Bloom (1974) erkannte, dass „Rückmeldungen als Grundlage für die Planung individueller Unterstützungsmaßnahmen genutzt werden sollen.“

Durch die Überprüfung der einzelnen Schülerantworten nach den FLTs wird die Erstellung, eines auf die individuellen Voraussetzungen der Schüler abgestimmten und damit effektiveren Unterrichts, möglich.

Es wäre sehr interessant in einer weiteren Studie zu untersuchen, wie groß Effekte auf den Wissenszuwachs durch die Adaption des Unterrichts mit Hilfe der Auswertung von formativen Leistungstests werden. Dafür müssten die teilnehmenden Lehrkräfte zwei Klassen parallel unterrichten, damit es keine Unterschiede in der Lehrerpersönlichkeit gibt und in der einen Klasse adaptiven und in der Kontrollklasse, herkömmlichen Unterricht halten.

5.4 Technische Umsetzung

5.4.1 Vor- und Nachteile elektronischer formativer Leistungstests

Wie bereits in Kapitel 4.2.2 ausführlich beschrieben, bot sich der Einsatz von herkömmlichen schriftlichen Tests bei der Durchführung der Studie nicht an, weil eine Variation des Feedbacks und vor allem ausführliche, individuelle Rückmeldungen nicht möglich gewesen wären. Dennoch soll im folgenden Abschnitt nicht nur auf die Vorteile elektronischer Leistungstests eingegangen, sondern auch die damit einhergehenden Nachteile näher beleuchtet werden:

Vorteile:

Ein zweifelsfrei riesiger Vorteil elektronischer Tests ist die Möglichkeit, sie automatisch, direkt nach der Bearbeitung der Aufgaben mit einem Knopfdruck auszuwerten. Dadurch fallen langwierige Korrekturen weg und man kann als Lehrkraft unmittelbar die Denkfehler der Schüler beim Problemlösungsprozess aufdecken und im anschließenden Unterricht sofort auf diese eingehen.
Weiterhin konnte ich eine ungleich höhere Motivation bei der Arbeit mit den Tablets auf Schülerseite beobachten, als bei der Bearbeitung der Paper-Pencil Tests während der Vorstudie. Würde man das Medium durch den Einsatz eines umfangreicheren Autorensystems noch voll ausschöpfen (durch die Einbindung von Filmmaterial oder Animationen), würde die Motivation sicherlich noch mehr steigen.
Durch den Einsatz eines anderen Autorensystems, könnte man noch mehr unterschiedliche Aufgabentypen in ansprechender Form realisieren (Drag & Drop Aufgaben, die es erlauben bestimmte Abbildungen auf dem Tablet zu verschieben und somit an die richtige Position zu ziehen; Experimente aufbauen lassen; innerhalb der Animation eines physikalischen Vorgangs einzelne Parameter verändern, z.B. die Wölbung des Vogelflügels beim Auftrieb, und dann beobachten, was passiert; …). Nutzt man bei der Bearbeitung der Tests nicht ein in sich geschlossenes System, wie es

bei der Studie der Fall war (siehe 5.4.2), sondern Computer, die sich regulär im Schulnetz befinden, kann man sogar Links zu externen Internetseiten einbinden.
Wie bereits mehrfach erwähnt, wird ein direktes und individuelles Feedback an den Schüler eigentlich nur durch die Verwendung computergestützter Tests möglich, vor allem wenn es sich um adaptive Rückmeldungen handelt.
Das diagnostische Potential der Aufgaben könnte durch den Einsatz einen adaptiven Systems voll ausgeschöpft werden – siehe hierzu auch Kapitel 5.3.3.3**.**
Die Einrichtung von Aufgabensammlungen reduziert den Aufwand bei der Erstellung elektronischer Leistungstests. Schon jetzt gibt es Aufgabenpools auf Länderebene (wie zum Beispiel das bayerische Realschulnetz), wo Items und Kurse gesammelt und anderen Lehrkräften zur Verfügung gestellt werden. Interessant wäre es, diese Vielzahl an einzelnen „Splittergruppen“, die es für jede Schulart, bzw. jedes Bundesland gibt, sinnvoll in einer Internetpräsenz zu bündeln und so einen riesigen Aufgabenpool für alle Lehrkräfte zu schaffen. Vorteilhaft wäre in diesem Zusammenhang auch eine Vereinheitlichung der verwendeten Dateiformate, so dass sie möglichst auf allen Endgeräten nutzbar sind. Das iPad hat zum Beispiel ein Problem, Flash - Dateien anzuzeigen und diese abzuspielen.

Neben den oben erwähnten Vorteilen, gibt es natürlich auch die eine oder andere Schattenseite beim Einsatz elektronischer Tests – hier die Nachteile:

Viele, vor allem ältere Lehrkräfte, haben Berührungsängste, wenn es um den Einsatz von neuen Technologien im Unterricht geht, so dass es schwierig ist, diese von den Vorteilen dieser Technologien zu überzeugen.
Diese Ängste sind nicht immer unbegründet – es ist nicht gerade aufbauend, wenn man als Lehrer weniger Routine im Umgang mit den Unterrichtmedien hat, als die meist technisch versierteren Schüler. Die Studie hat gezeigt, dass man bei spontan auftretenden Fehlern im technischen System, schnell darauf reagieren können muss, sonst sitzt die gesamte Klasse da und wartet. Um aber wiederum schnell reagieren zu können, muss man sich mit dem System gut auskennen und genau wissen, wo man den jeweiligen Fehler finden und auch wie man ihn möglichst schnell wieder beheben kann.
Hinzu kommt, dass der Entwicklungsaufwand der Testaufgaben und die Ausarbeitung von Kursen und Unterrichtsmaterial nicht unerheblich sind, vor allem, wenn man als Lehrkraft dabei auf sich

alleine gestellt ist. Für Lehrer, die nicht so technisch versiert sind, ist die Einarbeitung in die Soft- und Hardware mit großem Aufwand verbunden. Ohne ein Content Management System, wie Moodle ist die Erstellung von elektronischen Tests ohne Programmierkenntnisse schwer durchführbar.

Dazu kommen die hohen Anschaffungskosten für die Endgeräte (Tablets) oder der zeitliche Verlust beim Wechsel in den Computerraum. An vielen Schulen gibt es bereits Computerräume, die aber fast ausschließlich von den Informatik- und Mathematiklehrern genutzt werden. Die Bedienung der dort befindlichen Smartboards und der Computerarbeitsplätze ist nicht immer trivial und die Arbeit im Rechnerraum ist oft von mehr Unruhe begleitet, als der Unterricht in der gewohnten Umgebung des Klassenzimmers. Tablets können eine Lösung dieser Probleme sein, doch bleibt hier die Finanzierungsfrage. Es gibt bisher wenige iPad Klassen, die unterschiedlich strukturiert sein können: Manchmal schaffen die Eltern der Schüler die iPads an (über eine monatliche Kaufrate), manchmal besorgt auch die Schule einige wenige iPads pro Klasse, die dann für Gruppenarbeiten genutzt werden können.

Zusätzlich dazu bleibt die Wartung und der Service der Geräte zu klären: Oft werden eine oder mehrere Lehrkräfte dazu auserkoren, sich um diese Problematik, sowie die Einbindung in das Schulnetz und die damit einhergehenden Sicherheitsfragen zu kümmern.

Damit kommen wir zu einer weiteren Problematik: Die technische Infrastruktur an den Schulen: An vielen Schulen gibt es noch nicht einmal ein Schulnetz, geschweige denn ein stabiles WLAN, so dass ein flächendeckender Einsatz elektronischer Medien im Unterricht noch längere Zeit utopisch sein wird.

Ein weiteres Problem, diesmal speziell die elektronischen Tests betreffend ist, dass eigentlich nur geschlossene, in Einzelfällen noch halboffene Fragetypen möglich sind, jedenfalls, wenn man die Auswertung der Aufgaben automatisch durchführen lassen möchte. Der Verzicht auf offene Fragestellungen ist bei formativer Leistungsmessung allerdings soundso ratsam, denn die Korrektur sollte sich in Grenzen halten, um möglichst zeitnah auf die Ergebnisse der Tests eingehen zu können. Dies setzt hohe Anforderungen an das Aufgabendesign und die Formulierung der Testfragen, vor allem, wenn man konzeptuelles Wissen abprüfen möchte.

5.4.2 Technische Umsetzung des Projekts

5.4.2.1 Software

Vor noch nicht allzu langer Zeit, musste man über gute Programmierkenntnisse verfügen, wollte man selbst eine Lernplattform entwerfen oder elektronische Leistungstests erstellen. Heute gibt es eine Vielzahl an Autorensystemen und Content Management Systemen (CMS), mit denen eine umfassende Einführung von elektronischem Unterrichtsmaterial, Kursen und Tests ohne großes Programmierwissen möglich ist. In den folgenden Abschnitten soll etwas genauer auf die unterschiedlichen Tools eingegangen werden, die den Lehrenden heute zur Verfügung stehen.

5.4.2.1.1 Das Autorensystem zur Erstellung der Testfragen

Autorensysteme und Content Management Systeme erlauben es auch informationstechnischen Laien, elektronisches Unterrichtsmaterial und Onlinekurse ohne Programmierkenntnisse zu erstellen. Dabei sind einige, weniger komplexe Systeme (wie Moodle) auch leicht verständlich aufgebaut und bedürfen nur einer geringen Einarbeitungszeit. Insgesamt gilt: Je umfassender das Autorensystem, desto mehr Möglichkeiten hat man, optisch ansprechende, individuelle und abwechslungsreiche Module und Materialien zu erstellen, aber desto höher ist auch das geforderte technische Verständnis auf Seiten der Lehrkraft. So bietet Moodle eine kleinere Auswahl an Aufgabentypen bei der Erstellung von elektronischen Tests und weniger Möglichkeiten, externe Medien einzubinden, als zum Beispiel Adobe Captivate. Hinzu kommt, dass man Captivate auch im Verbund in einer ganzen eLearning Suite kaufen kann, die eine Vielzahl von Programmen beinhaltet, mit denen man nicht nur die Lerninhalte, sondern zusätzlich eine Vielzahl anderer elektronischer Materialien erstellen kann - es handelt sich um ein umfassendes Paket zur digitalen Bildbearbeitung, zur Erstellung von Animationen, Filmen, Skizzen, Sounds und Webinhalten… Den größten Freiheitsgrad beim Design seiner Lernumgebung und –inhalte, hat man natürlich, wenn man alles selbst programmiert (zum Beispiel in HTML 5). Es gilt: Je höher die technische Versiertheit der Lehrkraft, desto höher die Freiheitsgrade bei der Erstellung der Materialien.
Im folgenden Abschnitt möchte ich kurz auf die Vor- und Nachteile von Moodle eingehen und erläutern, warum ich mich letztendlich für dessen Einsatz entschieden habe.

Die Vorteile von Moodle sind:

Moodle kostet nichts – es handelt sich um Freeware (Open Source), die man sich im Internet herunterladen und für seine Zwecke einsetzen kann. Dies ist ein klarer Vorteil gegenüber kommerzieller Autorensysteme. Und dadurch, dass ständig weiter daran programmiert wird, gibt es häufig neue Versionen und Updates, die man kostenfrei nutzen kann.
Weiterhin ist Moodle plattformunabhängig, das heißt, das alle Inhalte, die man damit entwickelt, sowohl auf Windows- Rechnern, als auch unter IOS- (Mac) oder Android – Systemen genutzt werden können. Dies ist ein klares Plus, denn auf Macs laufen zum Beispiel Flash – Animationen, die man häufig im Internet finden kann, nicht.
Außerdem kann man sich problemlos mit etwas Zeit relativ in dieses Tool einarbeiten und es wird bereits von vielen Hochschulen und auch Schulen genutzt, um Onlinekurse anzubieten.
Neben Kursen (also darbietendem Lerninhalten), können auch andere Dinge, wie zum Beispiel elektronische Tests, Umfragen, Blogs, etc. erstellt werden.
Die wichtigsten Aufgabentypen (Lückentext, Multiple- und Single- Choice und Zuordnungsaufgaben sind vorhanden.
Doch der wohl größte Vorteil ist, dass Moodle nicht nur als CMS zur Erstellung von Unterrichtsmaterial dient, sondern auch als Lernplattform herangezogen werden kann: Man kann alle Materialien, die man entwickelt hat, zu Kursen zusammenfassen und ausgewählten Lernenden zur Verfügung stellen. Dabei speichert Moodle in einer Datenbank alle Ergebnisse der Schüler ab. Würde man die Inhalte mit Adobe Captivate erstellen, müsste man sich noch in ein zweites Tool einarbeiten, das als Lernplattform und Datenbank dient.

Leider lassen sich auch einige Nachteile nicht verleugnen:

In Moodle gibt es eigentlich gar keine Möglichkeit, das Design des Kurses, oder der Testfragen zu beeinflussen – es ist einfach vorgegeben und man muss es so nehmen, wie es ist… Für meinen Geschmack ist dies wenig individuell und durch seine Eintönigkeit nicht unbedingt umwerfend, aber es lenkt auf der anderen Seite auch nicht vom Wesentlichen (den Inhalten und den Testfragen) ab.
Weiter steht nur eine begrenzte Auswahl an Aufgabentypen zur Verfügung. Schön wäre es gewesen, noch etwas mehr Abwechslung in die Fragestellungen zu bringen. Bei Captivate besteht zum Beispiel auch die Möglichkeit, Drag & Drop oder Hotspot Aufgaben zu generieren, die sich bei be-

stimmten Themen anbieten würden, dies sieht Moodle leider nicht vor. Zwar soll Moodle auch extern erstellte Testfragen importieren können (wie zum Beispiel ein Kreuzworträtsel), leider ist mir der Import aber nicht gelungen.
Auch das Einbinden von Animationen (was gerade im naturwissenschaftlichen Bereich sehr interessant wäre) funktionierte leider nicht.
Zusätzlich besteht zwar die Möglichkeit, alle Fragen aus der Fragensammlung zu exportieren, aber beim Importieren der Fragen in eine andere Moodle Plattform, wurden fast alle Bilder nicht mit importiert. Dieses Verhalten hat sich auch bei der Durchführung von mehreren Versuchen nicht geändert.
Die Tests selbst ließen sich in Moodle leider nicht kopieren, man musste jeden Test manuell zusammenstellen, obwohl die Funktion eigentlich vorhanden ist. Dadurch, dass Pre- und Posttest von den Aufgaben gleich gewesen wären, hätte man sich hier Arbeit sparen können.

Ein weiteres Manko: Die in Moodle integrierte Version des Fragebogens (Aktion: Umfragen), fasst zwar die Umfrageergebnisse aller Schüler in einer Statistik zusammen, aber die einzelnen Ergebnisse kann man nicht automatisch (zum Beispiel nach Excel) exportieren. Dies bedeutet für die Auswertung der Ergebnisse viel Arbeit, weil man die Einträge jedes einzelnen Schülers anklicken und per Copy & Paste in Excel kopieren muss.

Nachdem ein kurzer Überblick über die Vor- und Nachteile von Moodle gegeben wurde, soll im Folgenden kurz erläutert werden, warum Moodle für die Durchführung des Feldexperimentes herangezogen wurde:

Der wahrscheinlich ausschlaggebende Punkt war, dass Moodle nicht nur als Autorensystem herangezogen werden kann, sondern auch als Lernplattform und Datenbank dient. So beschränkte ich die Einarbeitung auf ein Tool. Auch sollte es durch die Nutzung eines ganzheitlichen Systems keine Probleme mit der Kompatibilität zwischen den Inhalten und der Ergebnissicherung geben. Ein zweiter Aspekt, der für den Einsatz von Moodle sprach war, dass es bereits sehr häufig von Bildungseinrichtungen genutzt wird und kostenfrei zur Verfügung steht. Weiterhin ist die Plattformunabhängigkeit kein zu vernachlässigender Punkt, da zum Zeitpunkt des Testdesigns noch nicht feststand, welche Endgeräte bei der Studie zum Einsatz kommen würden.

Im Folgenden soll wenigstens kurz ein kleiner Einblick in den Aufbau dieses Content Management Systems gegeben werden (folgender Abschnitt soll aber kein Tutorial sein und erhebt auch keinen Anspruch auf Vollständigkeit – er soll lediglich einen Eindruck vermitteln, wie die Benutzeroberfläche gestaltet ist und welche Funktionen zur Verfügung stehen):

Hat man sich erfolgreich in Moodle eingeloggt, kommt man auf den Startbildschirm, von dem aus man alle weiteren gewünschten Aktivitäten durchführen kann. Bevor man allerdings irgendetwas ändern kann (Kurse und Tests anlegen, Nutzer hoch laden etc.), muss der Bearbeitungsmodus aktiviert werden. Wichtig auch: Am besten legt man zwei Login Namen für sich selbst an, mit unterschiedlichen Berechtigungen: Einen einfachen User – Login zum Testen der angelegten Kurse, Lerninhalte und Prüfungen und einen Administrator – Login, mit dem man seine Kurse anlegen und bearbeiten kann.

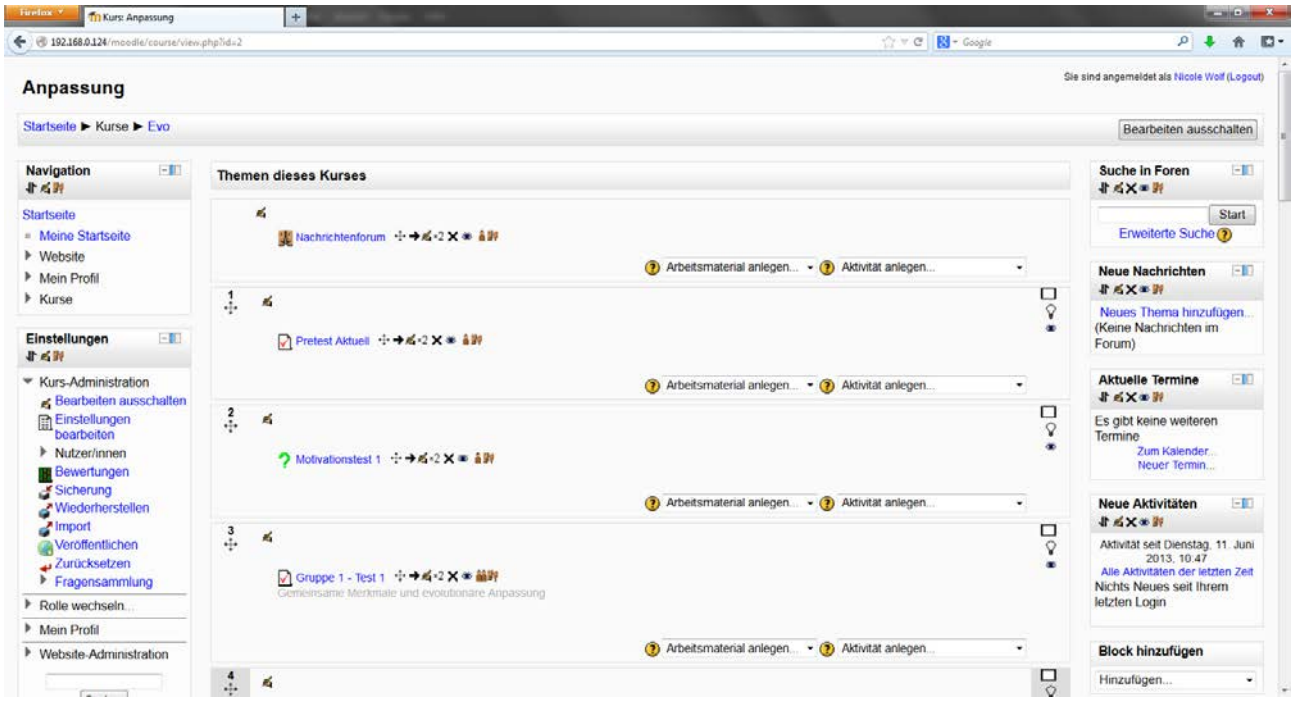


Abbildung 23: Eingabemaske in Moodle - Erstellung von Kursen und Tests

Auf dem Screenshot kann man sehen, dass der Kurs für die Studie Evo (für Evolution) genannt wurde (siehe Navigation oben) und welche Tests und Fragebögen dieser Kurs beinhaltet (siehe Mittelblock unter „Themen dieses Kurses“).

Die Navigationsleiste rechts zeigt alle Aktivitäten, die mit der Kommunikation mit und zwischen den Kursteilnehmern zu tun haben (Forum, Blog, Nachrichten, etc.). Diese Möglichkeiten wurden nicht genutzt, da außerhalb der Testzeitpunkte keine Korrespondenz mit den Schülern stattfand.

Links wird angezeigt, in welchem Unterpunkt man sich gerade befindet (unter dem Punkt „Navigation“) und welche Möglichkeiten man hat, um die Kurse und die Website zu administrieren und konfigurieren.

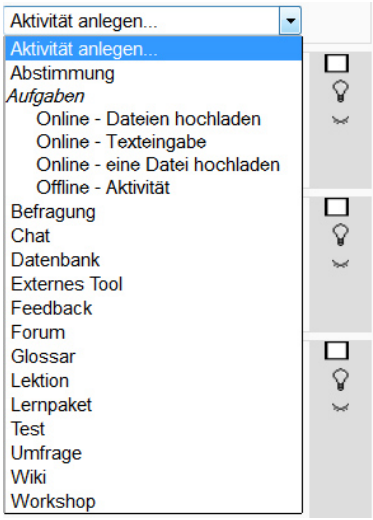


Abbildung 24: Moodle - Aktivität anlegen

In Moodle kann man unterschiedlichste Aktivitäten oder Arbeitsmaterialien anlegen. Für die Studie waren nur die Punkte „Test“ und „Umfrage“ von Bedeutung. Die Unterrichtseinheit selbst und alle damit zusammenhängenden Materialien wurden von den Lehrkräften auf herkömmliche Art und Weise zur Verfügung gestellt.

Die Einstellungen (Wie oft soll der Test wiederholbar sein? Sollen die Fragen der Aufgaben immer wieder neu gemischt werden? etc.) der angelegten Aktivitäten können im Nachhinein immer wieder geändert und angepasst werden.

Die Zusammenstellung der Tests erfolgt aus einer vorher angelegten Fragensammlung. Von dort fügt man die gewünschten Aufgaben hinzu, sortiert sie, oder löscht sie wieder. Auch die Reihenfolge der Fragen kann man bestimmen und ob für jede Aufgabe eine neue Seite geöffnet werden soll, oder nicht.

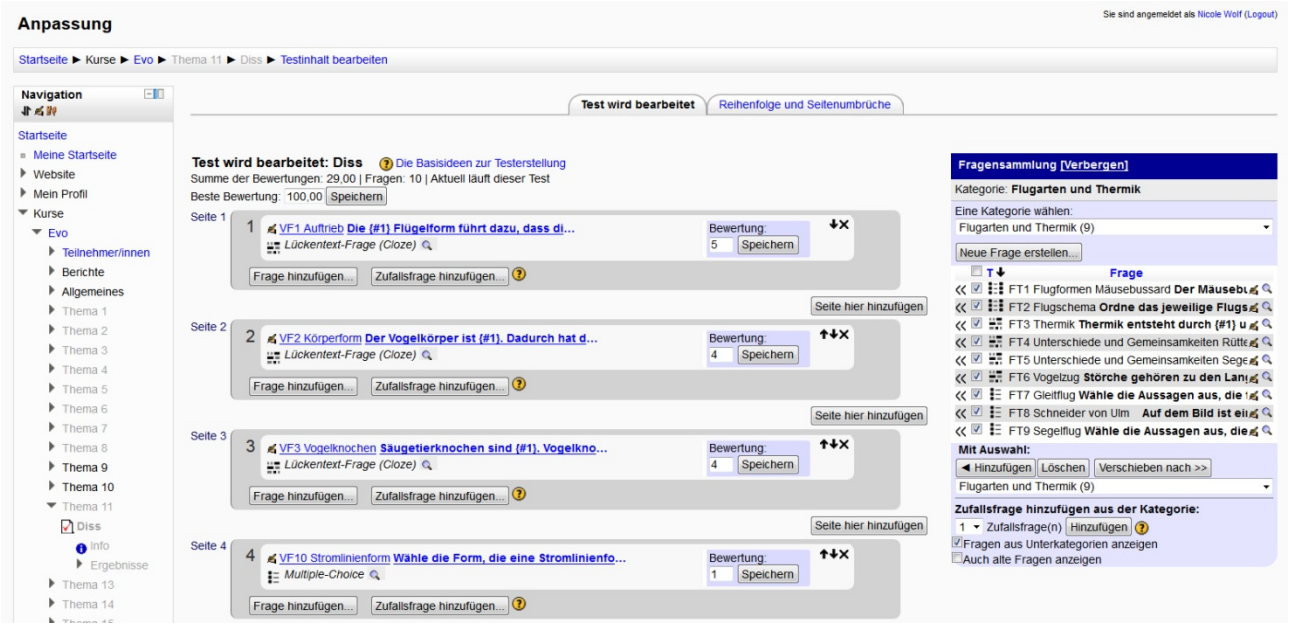


Abbildung 25: Zusammenstellung der Tests - Fragenauswahl

Bei der Erstellung der einzelnen Aufgaben kann man genau festlegen, wie viele Punkte es auf die richtige Beantwortung der jeweiligen Teilantwort geben soll und ob es für Falschantworten Punktabzug gibt. Letzteres ist anzuraten, denn sonst können die Schüler einfach alles ankreuzen und bekommen dadurch immer die volle Punktzahl…

Komfortabel und umfassend sind auch die Möglichkeiten, wenn es um das Feedback (siehe Abbildung 16) geht: Man kann die gewünschten Rückmeldungen zu jedem Testzeitpunkt zu- oder abschalten.
Zusätzlich dazu besteht die Möglichkeit ein Gesamtfeedback zu formulieren, um dem Schüler nach Abschluss des Tests noch eine Einschätzung seiner Leistung mit auf den Weg zu geben. Die Bewertungsgrenzen kann man dabei selbst festlegen, d.h. man kann vorgeben, bei wie viel Prozent der erreichbaren Punkte, welche Rückmeldung erscheinen soll.

Beispiele für Testfragen, kann man im Kapitel 5.3 ansehen. Dort werden auch die unterschiedlichen Aufgabentypen, deren Vor- und Nachteile und deren Nutzung näher beschrieben.

5.4.2.1.2 Lernplattform zur Auswertung der Schülerleistungen

Wie bereits angemerkt, ist Moodle gleich Beides in einem: Autorensystem zur Erstellung von eLearning Inhalten und Lernplattform zur Speicherung und Aufarbeitung von Schülerergebnissen. Auch die Verwaltung von Kursen und deren Teilnehmer ist mit diesem Tool möglich.

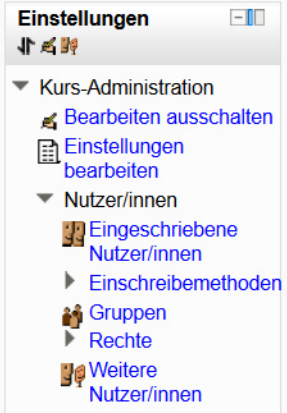


Abbildung 26: Nutzerverwaltung in Moodle Navigation

Das Anlegen der Kursteilnehmer ist sehr komfortabel, weil man mit Hilfe von Excel, gleich ganze Listen von Nutzern hoch laden kann und nicht jeden Schüler einzeln anlegen muss. Diese Datei sollte alle wichtigen Informationen enthalten, wie Benutzernamen und Passwort des Teilnehmers, Schule, Klasse, etc.

Alle Kursteilnehmer kann man sich in einer Liste anzeigen lassen, der man verschiedene Informationen zu den einzelnen Schülern entnehmen kann (siehe unten), wie zum Beispiel deren Kontaktdaten oder auch wann sie sich das letzte Mal mit den Lerninhalten, bzw. Tests auseinandergesetzt haben.

In der Nutzeransicht kann man den Teilnehmern verschiedene Rechte zuweisen, oder neue Lerner anlegen und löschen.

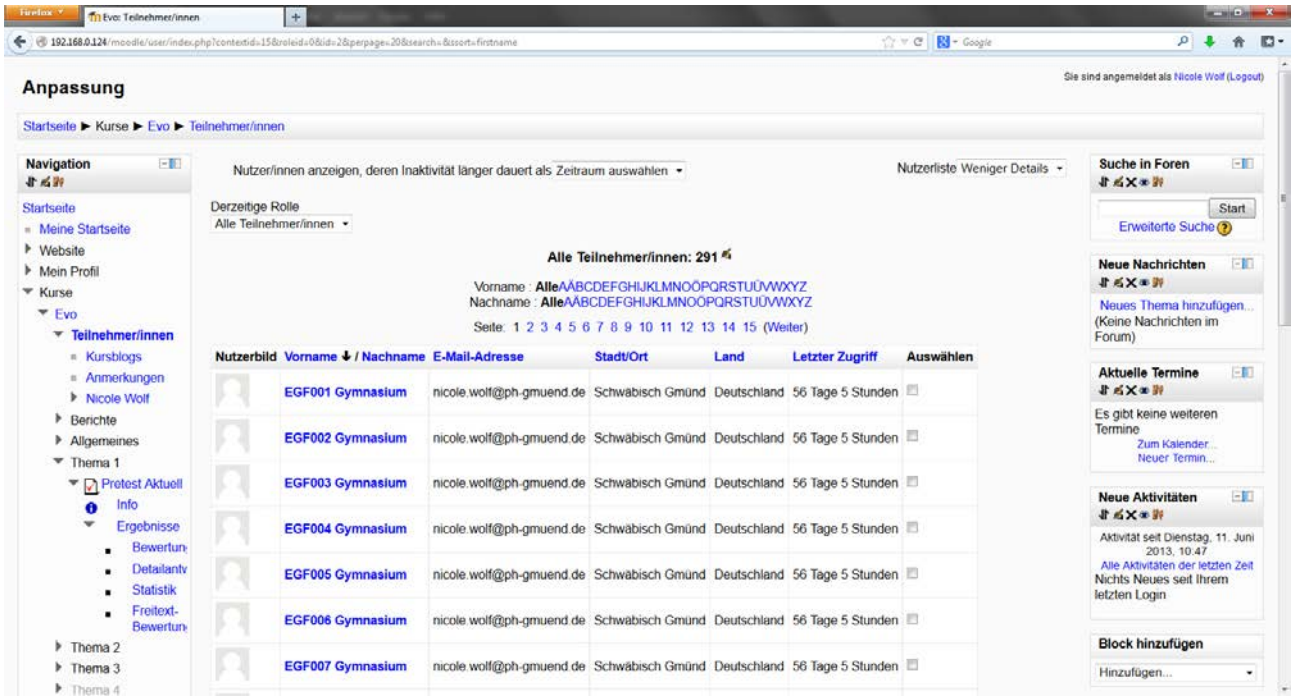


Abbildung 27: Verwaltung der Kursteilnehmer in Moodle

Bei der Auswertung der Schülerergebnisse kann man auf mehrere Optionen zurückgreifen, je nachdem, wie viele Informationen man über die einzelnen Schülerantworten haben möchte. So gibt es zum Beispiel die Möglichkeit, sich lediglich anzeigen zu lassen, welche Fragen die Schüler, richtig, teilweise richtig, oder ganz falsch beantwortet und wie viele Punkte sie auf die jeweiligen Aufgaben erhalten haben.

Vorname : **Alle**AÄBCDEFGHIJKLMNOÖPQRSTUÜVWXYZ
Nachname : **Alle**AÄBCDEFGHIJKLMNOÖPQRSTUÜVWXYZ

Seite: 1 2 3 4 5 6 7 8 9 10 (Weiter)

Tabellendaten laden als Textdatei mit kommagetrennten Werten (CVS)

Vorname / Nachname	E-Mail-Adresse	Begonnen am	Beendet	Verbrauchte Zeit	F 1 /5,56	F 2 /2,78	F 3 /5,56	F 4 /5,56	F 5 /2,78	F 6 /5,56	F 7 /5,56	F 8 /1,39	F 9 /2,78	F 10 /2,78
EGF001 Gymnasium Versuch erneut ansehen	nicole.wolf@ph-gmuend.de	16. Januar 2013 11:05	16. Januar 2013 11:41	36 Minuten 3 Sekunden	0,00 ✗	2,78 ✓	1,39 ✓	5,56 ✓	1,39 ✓	1,39 ✓	5,56 ✓	1,39 ✓	0,00 ✗	2,78 ✓
EGF002 Gymnasium Versuch erneut ansehen	nicole.wolf@ph-gmuend.de	16. Januar 2013 11:05	16. Januar 2013 11:40	34 Minuten 59 Sekunden	4,17 ✓	1,39 ✓	1,39 ✓	2,78 ✓	1,39 ✓	2,78 ✓	0,00 ✗	0,00 ✗	0,00 ✗	2,78 ✓

Abbildung 28: Auswertung der Testergebnisse in Moodle auf Aufgabenebene

Möchte man genauer wissen, welche Teilantworten gegeben wurden, kann man sich auch alle Detailantworten in einer weiteren Ansicht ansehen:

Vorname / Nachname	E-Mail-Adresse	Bewertung/100,00000	Antwort 1	Antwort 2	Antwort 3	Antwort 4	Antwort 5	Antwort 6
WolfE010 Volksschule Versuch erneut ansehen	nicole.wolf@ph-gmuend.de	58,33333	-> Kleine Nagetiere, wie Mäuse; -> Wattwürmer; -> Früchte; -> Fische ✓	-> Kletterfuß; -> Greiffuß; -> Lauffuß; -> Schwimmfuß ✓	-> Schutz vor Kälte; -> Schutz vor Verletzungen; -> Bremsen und Steuern, Balz; -> Bilden Tragflächen für den Flug ✓	Teil 1: helfen beim Steuern und Bremsen beim Flug; Teil 2: werden bei der Balz eingesetzt, um das Vogelweibchen zu beeindrucken.; Teil 3: luftdurchlässig, asymmetrisch; Teil 4: manchmal bunt und auffällig. ✓	Teil 1: erhöhte sich bei einem Teil der Haie, die dann überlebten und sich vermehrten; Teil 2: die Notwendigkeit Beute zu fangen sie dazu zwang, schneller zu schwimmen ✓	Teil 1: Im neuen Lebensraum mit dunklen Bäumen hat sich der schlauste Schmetterling getarnt, um zu überleben.; Teil 2: Alle Schmetterlinge würden dunkel werden; Teil 3: sich alle Lebewesen anpassen, um leichter zu überleben; Teil 4: Merkmale, wie die Farbe nicht im Erbmaterial gespeichert sind und somit nicht vererbt werden können. ✗

Abbildung 29: Auswertung der Testergebnisse in Moodle auf Antwortebene

Neben diesen detaillierten Informationen, hält Moodle aber auch zusammenfassende Statistiken zu allen möglichen Gegebenheiten bereit. Zum Beispiel kann man unten sehen, wie die Schüler durchschnittlich beim jeweiligen Test abgeschnitten haben:

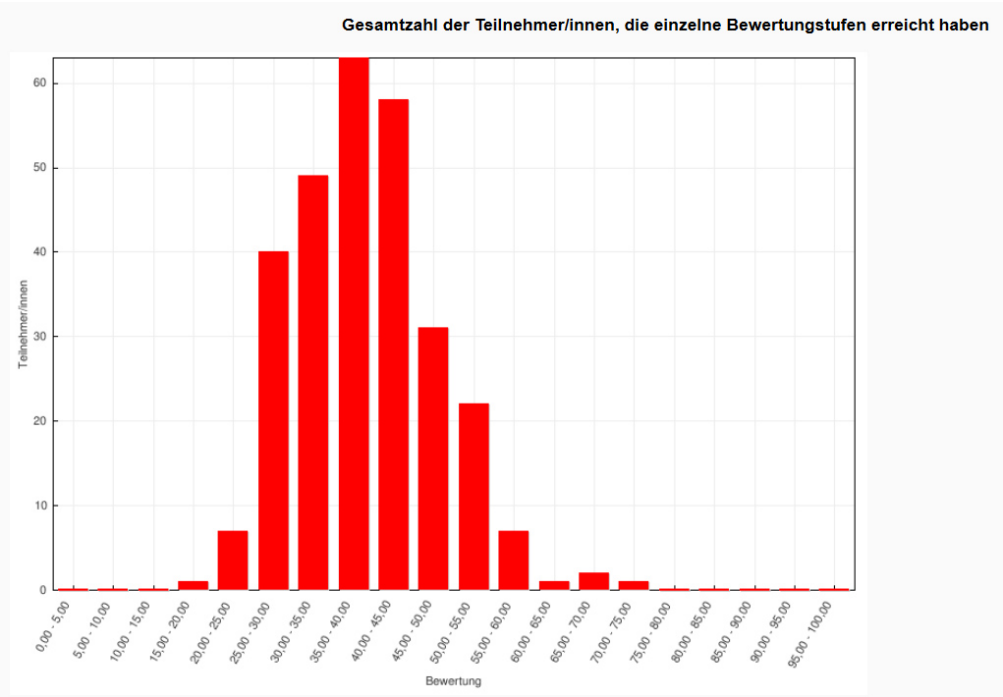


Abbildung 30: Auswertung der Tests in Moodle - Statistik

Alle Ergebnisse, Statistiken und Auswertungen, kann man aus Moodle heraus in eine Exceltabelle exportieren lassen und dort alles für den Import in SPSS vorbereiten.

Nur bei den Fragebögen zum Selbstkonzept und zur Motivation der Schüler gibt es wenige Exportmöglichkeiten: Hier kann man ausschließlich die zusammenfassende Statistiken exportieren, nicht aber die Antworten jedes einzelnen Schülers – diese muss man manuell aus den Fragebögen herauskopieren, um sie in einer Exceldatei zu sammeln.

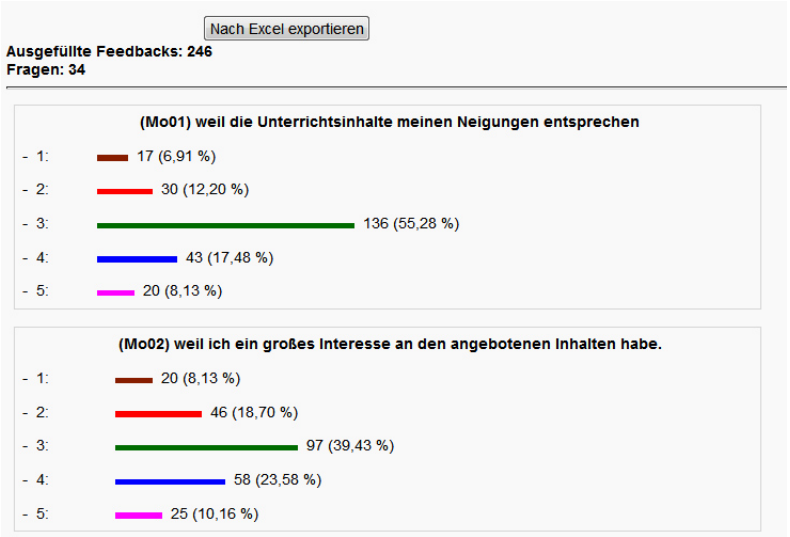


Abbildung 31: Auswertung der Motivationstests in Moodle

Beim Moodleserver wurde eine Installation für Windows 2008 Server eingesetzt, damit keine weitere Einarbeitung in Linux notwendig wurde.

Vorsicht: Das auf der Moodle - Homepage angebotene Standardpaket für Windows mit xampp[1] ist nicht einsatzfähig! Bei einer größeren Anzahl von Nutzern stürzt der Server ab. Stattdessen wurde wamp [2] verwendet.

Dies musste ich beim ersten Testlauf des technischen Systems in der iPad Klasse herausfinden: Schon der Zugriff von zwei Nutzern gleichzeitig, lies den Server zusammenbrechen.

5.4.2.2 Hardwarekomponenten

Da nicht alle Schulen über die nötige technische Infrastruktur verfügen, um eine Studie durchzuführen, die so viele technische Voraussetzungen verlangt wie dieses Projekt, wurde versucht, ein System zu erschaffen, das völlig unabhängig von den technischen Gegebenheiten der Schulen, also komplett autark funktioniert. Wie die Zusammenstellung der Hardware genau aussah, kann man in den folgenden Abschnitten erfahren.

5.4.2.2.1 Server und Accesspoint

Um ein autarkes System zu erhalten, musste ein Laptop als Server installieren, der sowohl die Tests innerhalb der Lernumgebung Moodle, als auch die Datenbank im Hintergrund verwalten konnte, um alle Schülerergebnisse speichern zu können. Um den Lernern den Zugriff auf die Tests zu ermöglichen, wurde natürlich auch die Bereitstellung eines eigenen WLANs nötig. Dieses wurde durch einen Accesspoint (Konfiguration – bei Interesse anfordern) erzeugt, der wiederrum durch ein LAN Kabel mit dem Laptop / Server verbunden war. Um nicht auf die Rechnerräume der Schule angewiesen zu sein, wurde für die Testzeitpunkte ein eigener Klassensatz Tablets angeschafft, die sich über das vom Accesspoint erzeugte WLAN mit dem Laptop verbinden konnten. Die Abbildung (siehe unten) soll verdeutlichen, wie die einzelnen Komponenten angeordnet sind und zueinander in Verbindung stehen:

[1] http://www.apachefriends.org/de/xampp.html
[2] http://www.wampserver.com/en/

Abbildung 32: Technische Realisierung der Evaluation

Der Server:

Als Server dient ein Notebook der über ausreichend Leistung und Hauptspeicher verfügt, um Moodle und alle sonstigen Dienste (DHCP) stabil und ohne größere Probleme am Laufen zu erhalten. Die Nutzung des eigenen Notebooks hat den Vorteil, dass man unabhängig vom Schulnetz und den technischen Gegebenheiten der Schulen ist. Man benötigt kein zur Verfügung gestelltes WLAN und muss das Sicherheitssystem der Schule nicht unterwandern, wenn schulfremde Geräte (die Tablets) versuchen, auf das Schulnetz zuzugreifen. Dies würde viele Änderungen in den schulinternen Sicherheitseinstellungen nach sich ziehen, was der Akzeptanz der Studie bei den Schulen und dortigen Lehrkräften nicht zuträglich gewesen wäre. Durch den Einsatz eines völlig autarken Systems, wurden in den jeweiligen Schulen lediglich zwei Steckdosen benötigt, um den Laptop und den Accesspoint mit Strom zu versorgen.

Ein weiterer klarer Vorteil war, dass die Schüler, wenn sie auf dem Moodleserver meines Laptops arbeiteten, auch mit meinem WLAN verbunden sein müssen, was die Nutzung des Internets ausschloss. Das heißt: Die Schüler konzentrieren sich auf die Bearbeitung der Tests. Das Surfen im Internet, das sowohl zu Ablenkungen, als auch zum Spicken hätte führen können, funktionierte nicht.

Ein nicht zu unterschätzender Nachteil hingegen ist, dass der Laptop keine Firewall haben durfte, was bedeutet, dass er immer ungeschützt war. Dies war nötig, damit die Endgeräte, mit denen die Schüler arbeiteten, auf den Laptop zugreifen und die bei der Testbearbeitung anfallenden Daten abgespeichert werden konnten.

Der Accesspoint (das WLAN):

Das WLAN wurde von einem Accessspoint erzeugt, der mit Hilfe eines LAN Kabels (Cross – Over Kabels) mit dem Server verbunden war. Dabei sollte man darauf achten, keine „Homegeräte“ oder Geräte von Discountern einzusetzen. Das Heimnetzwerk ist nicht dafür ausgelegt, dass 10 und mehr Geräte gleichzeitig auf den Accesspoint zugreifen. Deshalb wird der Einsatz von professionellen „standalone“ Geräten empfohlen. Bei der Studie wurde ein Cisco Access Point (CS-AIR-AP1142N-E-K9) eingesetzt.

5.4.2.2.2 Endgeräte

Zur Bearbeitung der Moodle Tests kann man theoretisch alle möglichen Endgeräte verwenden: Denkbar wären **herkömmliche PCs**, wie sie in den Rechnerräumen der Schulen vorhanden sind. Diese können dann auf einen, im World Wide Web liegenden Moodleserver zugreifen und dort die Tests bearbeiten. Es gibt bereits Zusammenschlüsse von Schulen oder ganzen Bundesländern, die im Internet in Moodle erstellte elektronische Kurse und Unterrichtsmaterial zur Verfügung stellen. Zum Beispiel kann man beim Bayerischen Realschulnetz einen Zugang beantragen und dort eigene Kurse hochladen und Kollegen zugänglich machen, oder bereits vorhandene Materialien nutzen. Schön wäre es, wenn man die entstehenden „Splittergruppen“ einzelner Schularten sammeln und zusammenfassen könnte, so dass nach Fächern und Themen geordnet, jedem Lehrer für seine Arbeit an der Schule umfangreiches Unterrichtsmaterial und eine Vielzahl an Testfragen zur Verfügung stünde, aus denen er nur noch auszuwählen bräuchte. So würde die Erstellung formativer Tests besonders effektiv und einfach von Statten gehen.

Klarer Vorteil des Einsatzes von herkömmlichen PCs: Die Rechner, und die nötige technische Infrastruktur, sind an vielen Schulen bereits vorhanden und die Monitore der Computer bieten ein viel größeres Display an, als Tablet PCs oder gar Smartphones, auf die im Folgenden genauer eingegangen werden soll. Ein Nachteil besteht darin, dass die Desktop Computer nicht mobil sind und dadurch nicht einfach irgendwo hin mitgenommen werden können. Weiter können sie nicht dazu verwendet werden, bestimmte Arbeitsaufträge zu erledigen (zum Beispiel einen englischen Dialog zwischen zwei Schülern zu filmen und diesen dann zu präsentieren).

Gibt es ein schulweites, stabiles WLAN, ist auch der Einsatz von **Tablet PCs** möglich. „Tablet PCs sind mobile Endgeräte, die über einen berührungsempfindlichen Bildschirm (Touchscreen) per Eingabestift oder dem Finger als Mausersatz bedient werden können. Neuere Geräte erlauben außerdem Multi-Touch Gesten mit mehreren Fingern, um z.B. Bildausschnitte zu vergrößern.“ (Stolpmann & Welling 2011, S. 46) Der Einsatz dieser Geräte bringt neben der einfachen und intuitiven Bedienung auch Vorteile durch die Mobilität: Man muss nicht immer mit der gesamten Klasse in den Computerraum wechseln, nur um einen kurzen formativen Leistungstest durchzufüllen. Die Anschaffung dieser Geräte ist aber mit einem hohen Kostenfaktor verbunden. Dennoch gibt es an einigen Schulen bereits verschiedene Ansätze, den Tablet PC in den Unterricht zu integrieren. So habe ich während meines Projekts in einer Klasse gearbeitet, die 5 iPads von der Schule zur Verfügung gestellt bekommen hat, um gemeinsam Gruppenarbeiten und Recherchearbeiten durchzuführen, und um Vorträge zu erarbeiten. Dadurch fallen die Kosten auf Seiten der Schüler weg, aber natürlich hat auch nicht jeder Schüler ein Gerät, mit dem er individuell arbeiten kann. Die zweite Möglichkeit sind die sogenannten iPad Klassen. Davon gibt es in Bayern jetzt schon einige. In diesen Klassen hat jeder Schüler sein eigenes iPad, was die spontanen Einsatzmöglichkeiten der Geräte in allen Fächern natürlich enorm erweitert, aber auch hohe Kosten für die Familien der Kinder bedeutet. In einer Art Leasing Modell werden die iPads in einer monatlichen Rate gekauft und gehen irgendwann in den Besitz des Schülers über…

Da in der Studie natürlich nicht nur iPad Klassen teilgenommen haben, wurden den Schülern die Tablets, die im Rahmen des Forschungsprojekts angeschafft wurden, für die Testzeitpunkte zur Verfügung gestellt. Weiter unten werden die eingesetzten Tablets und deren Vor- und Nachteile näher erläutert. Die größten Vorteile, die die Tablets mit sich bringen, sind zweifelsfrei, ihre Mobilität (man steckt sie einfach in die Schultasche und nimmt sie überall mit hin) und ihre vielfältigen Einsatzmöglichkeiten: Ausgestattet mit Foto- und Videokamera, Internetzugang, Präsentationssoftware und diversen Apps, sind sie für jegliche Art von Unterricht denkbar.

Den größten Nachteil, den Kostenfaktor, kann man beim Einsatz von **Smartphones** umgehen: Sie sind ähnlich ausgestattet wie Tablet PCs und können sich ebenfalls mit einem WLAN verbinden. Das Gute: Sehr viele Schüler besitzen ein eigenes Smartphone und wissen damit umzugehen. Eine nicht repräsentative Umfrage ergab: Je höher die Jahrgangsstufe, desto mehr Schüler besitzen ein

Smartphone (in den Klassen 9 und 10 schon fast jeder Schüler). Dies hat damit zu tun, dass die herkömmlichen Telefone, mit denen man nur telefonieren, eine SMS verschicken und fotografieren konnte, bald immer mehr vom Markt verdrängt und Smartphones im Gegenzug immer günstiger werden. Auch bei den Testzeitpunkten haben 2 Schüler die Prüfungen mit ihrem IPhone bearbeitet, weil sie ihr iPad vergessen hatten und das war ohne Probleme möglich. Dadurch, dass viele Schüler bereits ein Smartphone besitzen, war ihnen der Umgang mit den Tablets von vornherein vertraut. Fast keine Kinder brauchten Hilfe oder Anleitung, wenn es um die Bedienung der Geräte ging. Der Nachteil der Smartphones liegt natürlich beim sehr kleinen Display (Aufenanger &Schlieszeit 2013).

Eine Möglichkeit, den Vorteil eines größeren Displays, mit dem Vorteil der Mobilität zu verbinden, stellt der **Laptop** dar. Hier gibt es ja sogar schon verschiedene Mischformen aus Laptop und Tablet (mit abnehmbarer Tastatur), die noch kleiner und mobiler sind und trotzdem alle Funktionalitäten eines Laptops aufweisen. Das Feldexperiment wurde auch in einer Laptop – Klasse durchgeführt, die aber die Leistungstests der Einfachheit halber (einloggen, WLAN Verbindung, etc.) auch mit Hilfe der Tablets ausgefüllt haben. Die Laptops sind in der Anschaffung aber meist noch teurer als einfache Tablets.

In der Studie wurden insgesamt 29 Tablet PCs eingesetzt. Davon waren 23 von Lenovo A1 (von IBM) und 6 von Samsung Galaxy. Außerdem hat eine Klasse mit den eigenen iPads gearbeitet. Im Folgenden sollen die einzelnen Tablets und deren Vor- und Nachteile genauer aufgeführt werden. Auf eine detaillierte Beschreibung der technischen Ausstattung wird allerdings verzichtet – nur auf die, für das Projekt wichtigen Punkte soll näher eingegangen werden:

Tabelle 20: Vergleich der in der Studie verwendeten Tablets

	Lenovo A1 (IBM)	Samsung Galaxy Tab 2	iPad (Apple)
Bild			
Betriebssystem	Android Bei Android gab es keine Probleme bei der Ar-		Mac OS Auch Mac OS lief einwandfrei

	Lenovo A1 (IBM)	Samsung Galaxy Tab 2	iPad (Apple)
	beit mit Moodle.		mit Moodle. Bei Apple kann es nur zu Schwierigkeiten kommen, wenn man Macromedia Flash anzeigen lassen will. Dies lässt Apple aus Sicherheitsgründen nicht zu (Flash wurde aber im Projekt nicht benötigt).
Displaygröße	7 Zoll + klein und handlich – wenn man knapp 30 Tablets transportieren muss, ist die Größe durchaus wichtig - Durch das kleine Display werden nicht alle Auswahloptionen der Zuordnungsaufgaben angezeigt und die Schüler müssen jedes Mal durch die Auswahl scrollen – darauf muss man die Lerner extra hinweisen. - Trotz der Einstellung in Moodle, dass die Lernplattform auf mobilen Endgeräten laufen muss, wird ab und zu Text abgeschnitten (bei den Texten der Kontrollgruppen) und die Antworten bei den Lückentextaufgaben werden teilweise im Auswahlmenü nicht vollständig angezeigt (erst, wenn man die Antwort auswählt, wird sie komplett in die Lücke übernommen). Dies macht die Arbeit mit den kleinen Tablets etwas unübersichtlich.		9.7 Zoll + Alles Wichtige wird angezeigt, so dass die Schüler ungestört arbeiten können und nicht Gefahr laufen, irgend -etwas zu übersehen. - Müsste man einen Klassensatz iPads ständig mit sich herumtragen, wäre dies sehr anstrengend (dadurch, dass die iPads in den Klassen vorhanden waren, gab es in dieser Hinsicht aber keine Probleme).
Touchscreen	Reagiert manchmal etwas träge und ungenau.	Reagiert besser als das Lenovo, aber noch nicht so fein, wie das iPad.	Sehr genaue und schnelle Reaktion. Auch der Bildaufbau ist am iPad am schnellsten.
Benutzerfreundlichkeit	Bei beiden Tablets gibt es bei der Bedienung im Großen und Ganzen nichts zu meckern. Allerdings wurden bei der Studie auch nur 2 Dinge benötigt: Display entsperren und Webbrowser starten (über die Weltkugel). Bei einer ausführlicheren Nutzung, gäbe es an der Bedienoberfläche und den Icons schon einige Dinge, die ich persönlich verbessern würde...		Die Bedienoberfläche ist klar strukturiert, die Icons aussagekräftig (man weiß sofort, was sich dahinter verbirgt) und die Bedienung durch einfache Gesten sehr schnell und intuitiv.
WLAN Verbindung	Dadurch, dass die Tablets immer nur für den Schulversuch verwendet wurden, konnte man von vornherein in den Einstellungen festhalten, dass sich die Tablets automatisch mit dem benötigten WLAN (formal) verbinden sollen. Auch die Einrichtung des Moodle – LogIn Screens als Startseite im Internet war möglich. So mussten die Schüler, um mit der Arbeit beginnen zu können, nur auf die Weltkugel klicken.		Da die iPads vorwiegend im normalen Unterricht genutzt wurden und dadurch nicht von vornherein mit dem richtigen WLAN verbunden waren, mussten dies die Schüler manuell machen, auch das Aufrufen der Login Seite von Moodle musste von den Kindern übernommen werden, indem sie im Browser, die

	Lenovo A1 (IBM)	Samsung Galaxy Tab 2	iPad (Apple)
			richtige Webadresse eingaben (setzt man beim Login ein Lesezeichen, entfällt dieser Schritt beim zweiten Mal). Da die Schüler ständig mit den iPads arbeiten, war dies aber kein großes Problem. Um die Seite aufrufen zu können, mussten, falls noch nicht geschehen, die Cookies aktiviert werden.
WLAN Aufbau	Funktionierte Automatisch. Es dauerte immer etwas, bis die Verbindung zum WLAN aufgebaut war.	Das WLAN musste bei jedem einzelnen Gerät von immer extra manuell getrennt und wieder verbunden werden. Zum Glück waren nur 6 Samsung Geräte im Einsatz… Es dauerte immer etwas, bis die Verbindung zum WLAN aufgebaut war.	Wie bereits oben beschrieben, mussten sich die Schüler selbst mit „formal“ verbinden. Es dauerte immer etwas, bis die Verbindung zum WLAN aufgebaut war.
Akkulaufzeit	Die kürzeste Betriebszeit im Vergleich zu den anderen beiden Geräten. Dabei gab es große Unterschiede zwischen den Akkulaufzeiten einzelner Geräte. Für den Einsatz in 6 Schulstunden, hat der Akku aber ausgereicht. Danach waren die Geräte aber ziemlich am Limit.	Gute Akkulaufzeit. Die Tests haben die Akkus auch nach 6 Stunden nicht ans Limit gebracht.	Laut Hersteller 10 Stunden Betriebszeit.
Laden	Alle Geräte wurden vor jedem Einsatz komplett geladen. Dies nahm einige Zeit in Anspruch, weil es nicht möglich ist alle 29 Geräte auf einmal ans Stromnetz zu Hause hängen. Die Lenovos haben eine Ladeanzeige, die man von außen gut sehen kann und die von Orange auf Grün schaltet, wenn der Akku voll ist. Bei Samsung muss man immer nachsehen, wie weit die Geräte sind indem man kurz auf den Einschaltknopf drückt – nur dann wird der Ladestatus angezeigt.		Dadurch, dass jeder Schüler sein eigenes iPad hatte, waren diese auch selbst für das Aufladen verantwortlich. Gut bei Apple: Sollte man als Lehrkraft einen ganzen Klassensatz verwalten, gibt es einen Koffer mit Ladefunktion: Dort sind die iPads gut geschützt, können transportiert und auf einmal geladen werden, was den Arbeitsaufwand deutlich senkt.

Der größte Vorteil, den die Arbeit mit dem iPad mit sich bringt, ist das iOS – Betriebssystem, welches die Apple Produkte zu einer Art „Ökosystem“ macht: „Alles ist aufeinander abgestimmt und nichts kann ohne Zustimmung von Apple hinzugefügt werden“ (Aufenanger & Schlieszeit 2013, S. 7). Dadurch hat man die Sicherheit, dass keine Viren, Trojaner oder Ähnliches in den Anwendungen und Programmen enthalten sind.

Fazit: Sollte man ganze Klassen mit Tablets ausstatten (Tablet = Eigentum des Schülers) und soll sehr viel im Unterricht mit den Tablets gearbeitet werden, wäre es sinnvoll, ein Tablet mit größerem Display anzuschaffen (ca. 10 Zoll). Welche Marke und welches Betriebssystem man wählt, ist wohl Geschmacks- und Preissache. Allerdings sollte sich die Klasse auf ein Betriebssystem und einen Gerätetyp einigen, sonst hat man als Lehrkraft nur Chaos, wenn es darum geht irgendwelche Änderungen in den Einstellungen vorzunehmen. Die Lehrkraft sollte sich ja auch mit den eingesetzten Geräten auskennen und das wird sehr schwierig, wenn die Schüler mit unterschiedlichen Typen arbeiten. Von der Bedienung und der Nutzeroberfläche her, sollte das Tablet ansprechend und intuitiv gestaltet sein und auch der Touchscreen sollte einwandfrei funktionieren. Angenehm ist auch, wenn der Seitenaufbau und Speicherprozesse nicht zu viel Zeit in Anspruch nehmen. Bei der Auswahl des richtigen Tablets ist es wichtig, sich bereits vorher genau zu überlegen, was damit alles gemacht werden soll (Dialoge Filmen, Bildbearbeitung, Erstellung von Präsentationen und elektronischen Büchern/Heften). Diese Vorüberlegungen werden interessant, wenn es darum geht, die richtige Speichergröße und eine vernünftige Kameraauflösung zu wählen. Grundsätzlich waren für die Studie alle drei Geräte mehr oder weniger gut geeignet – die einzelnen Vor- und Nachteile kann man der Tabelle (siehe oben) entnehmen.
Um die Tablets hinreichend zu schützen, ist anzuraten für jedes Tablet eine Schutzhülle aus Leder zu besorgen, in denen die Geräte während der Arbeit auch bleiben können. Der Umgang mit den Tablets von Schülerseite aus war sehr lobenswert: Es wurde nie Unfug damit getrieben und alle Kinder gingen vorsichtig damit um. In der iPad Klasse war die Arbeit mit den Tablets natürlich am unkompliziertesten, weil die Schüler dort von vornherein geübter mit den Tablets agiert haben und die Schlepperei der Technik wegfiel: Ein Koffer mit den 29 Tablet PCs ist ziemlich schwer und lässt sich nicht unbedingt gut durch die Treppenaufgänge der Schulen transportieren.

Die angeschafften Tablets mussten vor ihrem ersten Einsatz konfiguriert werden.

Folgende Einstellungen wurden vorgenommen:

- Uhrzeit und Datum
- WLAN: Automatische Verbindung mit Formal
- Screensaver: Verlängerung des Zeitintervalls, so dass das Display nicht bei jeder Überlegung dunkel wird, oder sich ganz abschaltet
- Startseite im Internet: Login – Seite der Moodle Homepage
- Sprache bei der Tastatur: Deutsch
- Ton auf lautlos, damit es im Klassenzimmer nicht klingt, wie in der Kommandozentrale eines Ufos

Vor jedem weiteren Einsatz wurden die Tablets geladen, gereinigt und überprüft, ob der Ton auch wirklich ausgeschaltet war.

6 Evaluationsergebnisse

6.1 Einleitung

Die für die Hypothesenprüfung notwendigen Variablen wurden mit standardisierten bzw. in Vorstudien erprobten Testverfahren in den Treatmentgruppen (Experimentalgruppen und Kontrollgruppe) getestet und überarbeitet. Die genaue Beschreibung der Messinstrumente, befindet sich unter Gliederungspunkt 5.3.2. In den folgenden Abschnitten wird die Stichprobe genauer beschrieben. Auch wird besonderes Augenmerk auf die Auswertung der Ergebnisse und somit auch die Überprüfung der Hypothesen gelegt. Außerdem soll noch einmal auf die biologiedidaktische Frage eingegangen werden, inwieweit sich die Fehlvorstellungen der Schüler zum Konzept der evolutionären Anpassung bereits in einer niedrigen Jahrgangsstufe durch gezielte Rückmeldungen beheben lassen und wie gut sich elektronische Leistungstests entwickeln, auswerten und in den Unterricht integrieren lassen. Auch der Umgang der Schüler mit den Tablet PCs soll näher beschrieben werden.

6.2 Ergebnisse des Feldexperiments

6.2.1 Allgemeine Informationen über die Stichprobe

Insgesamt haben *N* = 261 Schüler der 6. (*n* = 83 Schüler) und 7. Jahrgangsstufe (*n* = 178 Schüler) an der Studie teilgenommen. Das Durchschnittsalter der Schüler lag bei 13.34 Jahren (Mittelwert) mit einer Standardabweichung von sd = .82. Sie besuchten 10 verschiedene Klassen aus sechs unterschiedlichen Schulen. Dabei waren alle Schularten Bayerns vertreten, also Volks-, Mittel- und Realschulen und ein Gymnasium. Die Schüler verteilten sich auf die soeben genannten Schularten wie folgt: 29 Gymnasiasten (angestrebter Abschluss: Abitur), 141 Realschüler (Mittlere Reife), 91 Schüler (Qualifizierender Hauptschulabschluss) aus Mittelschulen. Durch die Durchmischung der unterschiedlichen Bildungslevels, konnte die ökologische Validität der Studie gewährleistet werden.

Hinsichtlich des Geschlechts ergab sich folgende Aufteilung:

Geschlecht:		
	Absoluter Wert	Relativer Wert
männlich:	123	43.31 %
weiblich:	161	56.69 %

Auf die Frage, ob ihre Eltern in Deutschland geboren wurden und welche Sprache zu Hause gesprochen würde, antworteten die Schüler folgendermaßen:

Wurde dein Vater in Deutschland geboren?		
	Absoluter Wert	Relativer Wert
Ja:	211	74.30 %
Nein:	73	25.70 %

Wurde deine Mutter in Deutschland geboren?		
	Absoluter Wert	Relativer Wert
Ja:	213	75.00 %
Nein:	71	25.00 %

Welche Sprache wird bei dir zu Hause gesprochen?		
	Absoluter Wert	Relativer Wert
nur deutsch:	187	65.85 %
deutsch + andere Sprache:	89	31.34 %
nur eine andere Sprache:	8	2.82 %

Obwohl ca. 1/3 der Schüler aus Migrantenfamilien kamen, gab es nur einen Lerner in einer der Mittelschulen, der kein Deutsch verstand und daher auch an der Studie nicht teilnehmen konnte, alle anderen Schüler waren der Sprache mächtig und haben diese gut verstanden und gesprochen. Dennoch kann man nicht ausschließen, dass ein paar Schüler größere Verständnisprobleme bei der Beantwortung der Fragen hatten als andere, obwohl diese sich, trotz Hilfsangebot, nicht gemeldet haben: Nachfragen sprachlicher Art gab es eigentlich nicht – nur zu biologischen Fachbegriffen, wie der „Population“.

6.2.2 Vorbereitungen für die Ergebnisauswertung

Überprüfung der Vorbedingungen:

Bei folgenden ***Variablen*** wurde vor Beginn der Berechnungen überprüft, ob sie einen Einfluss auf das Treatment haben:

- Positives Selbstkonzept – SKTA01 bis SKTA03 (Helmke 1992)
- Negatives Selbstkonzept – SKTA04 bis SKTA08 (Helmke 1992)
- Intrinsische Motivation (Interesse) – 1Mo01 bis 1Mo03 (Wild et al. 1995)
- Intrinsische Motivation (Kompetenzorientiert) – 1Mo04 bis 1Mo06 (Wild et al. 1995)
- Instruktionsbezogene Angst (Trait) – INA03, INA06, INA08, INA12

- Instruktionsbezogenes Interesse (Trait) - INA01, INA02, INA04, INA11
- Instruktionsbezogenes Wohlbefinden (Trait) - INA05, INA07, INA09, INA10 (Gläser–Zikuda et al. 2005)

Unter Kapitel 5.3.2.5 kann man einsehen, welche Items für die Erhebung dieser Persönlichkeitsmerkmale verwendet wurden. Diese wurden von den Schülern im Selbstreport durch das Ausfüllen des Motivationstests 1 beantwortet.
Die Überprüfung der Vorbedingungen zeigte, dass es keine signifikanten Unterschiede zwischen den Gruppen hinsichtlich der oben genannten Merkmale gab, außer bei der Schulleistung. Daher wurde dieses Merkmal als Kovariate ins Modell mit aufgenommen.

*Das Berechnungsmodell enthält folgende weiteren **Kovariaten**:*

- Schulleistung: Erhebung retrospektiv (letzte Zeugnisnote) durch einen Selbstreport der Schüler in den Fächern Deutsch, Mathematik und Biologie/PCB/Natur und Technik – je nach Schulart.

Die Noten der Fächer Deutsch, Mathematik und Biologie/PCB/Natur und Technik wurden einer Hauptkomponentenanalyse unterzogen. Nach dem Kriterium des Eigenwertes > 1 wurde ein einziger Faktor extrahiert, der 51.58% der Varianz erklärte (Tabelle 21).

Tabelle 21: Hauptkomponentenanalyse der Schulnoten

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
	Gesamt	*% der Varianz*	*Kumulierte %*	*Gesamt*	*% der Varianz*	*Kumulierte %*
1	1.547	51.580	51.580	1.547	51.580	51.580
2	.780	26.009	77.589			
3	.672	22.411	100.000			

Extraktionsmethode: Hauptkomponentenanalyse.

Alle drei Schulnoten bilden einen Faktor (Schulleistung). Die Ladungen auf den Faktor bzw. die Komponente sind als gut zu bezeichnen, da sie über 0.4 liegen (Bühner 2011).

Tabelle 22: Komponentenmatrix der Schulnoten

	Komponente
	1
Mathematiknote aus dem vergangenen Schuljahr	.686
Biologie/PCB-Note aus dem vergangenen Schuljahr	.760
Deutschnote aus dem vergangenen Schuljahr	.706

Für die weitere Verwendung in den statistischen Analysen wurden die gespeicherten Faktorenwerte aus der Regression verwendet.

- Pretest/ Vorwissen: Als Vergleichsgröße. Genau wie später der Post- und der Behaltenstest, wurde der Pretest in diejenigen Items unterteilt, die das Fakten- bzw. Konzeptwissen abfragen. Diese Unterteilung wurde in der statistischen Auswertung durchgeführt. Das Vorwissen aller Schülergruppen wurde mit Hilfe eines elektronischen Pretests vor Beginn der Unterrichtseinheit erhoben.
- Geschlecht: Evaluation innerhalb des Motivationstests 1 durch Selbstreport.
- AbschlussID: Festlegung durch die Versuchsleiterin (Unterteilt in Abitur, Mittlere Reife und Qualifizierter Hauptschulabschluss).

Zusätzlich zu den eben aufgeführten Kovariaten befindet sich das als ***Faktor*** im Modell:

Die Treatmentgruppe „ausführliches Feedback“ wurde in zwei Subgruppen unterteilt - je nach Nutzung des Feedbacks. Dazu wurden die beiden Items

FB09GRUPPE1 Ich habe das Feedback ausführlich gelesen.
FB10GRUPPE1 Mir war nur wichtig, ob ich richtig, oder falsch geantwortet habe – die genaue Rückmeldung war mir egal.

einer Faktorenanalyse unterzogen (Hauptkomponenten) und die Residuen gespeichert (Faktorenwerte). Positive Werte wurden als die „Nutzung von Feedback“ interpretiert, negative Werte als „keine Nutzung“.

Als externe Validierung wurde die Feedbacknutzung mit Gewissenhaftigkeit aus dem Big Five korreliert. Der Wert von r = .295 (p = .008), zeigt eine mittlere Korrelation an, die belegt, dass die gewissenhafteren Schüler das Feedback eher annehmen und auch nutzen. Somit kann von einer validen Angabe der Schülerinnen zur Feedbacknutzung ausgegangen werden.

Vergleichbarkeit der Treatmentgruppen

Bezüglich der vier Treatmentgruppen zeigten sich, außer bei den Schulnoten, keinerlei Unterschiede in den Variablen (siehe Tabelle 23). Alle Treatmentgruppen wiesen vergleichbare Ausgangsbedingungen auf, sowohl das „positive und negative Selbstkonzept“ betreffend, als auch die „Intrinsische Motivation“ und die Variablen des Big Five (Offenheit, Extraversion, Gewissenhaftigkeit, Verträglichkeit und Neurotizismus). Auch in den Variablen „Instruktionsbezogene(s)_Interesse / Angst und Wohlbefinden“, sowie in den Ergebnissen des Pretests, waren keine Unterschiede auszumachen.

Tabelle 23: Vergleichbarkeit der Treatmentgruppen

		df	Mittel der Quadrate	F	Signifikanz
Schulleistung	Zwischen den Gruppen	3	3.748	3.842	.010
	Innerhalb der Gruppen	243	.975		
	Gesamt	246			
Intrinsische Motivation Interesse (T1Mo_IntrinsicMot_Interest)	Zwischen den Gruppen	3	.685	1.015	.387
	Innerhalb der Gruppen	243	.675		
	Gesamt	246			
Intrinsische Motivation Kompetenz (T1Mo_IntrinsicMot_Competence)	Zwischen den Gruppen	3	.522	.539	.656
	Innerhalb der Gruppen	243	.968		
	Gesamt	246			
Positives Selbstkonzept (SKTA_PositivesSK)	Zwischen den Gruppen	3	1.258	1.756	.156
	Innerhalb der Gruppen	243	.716		
	Gesamt	246			
Negatives Selbstkonzept (SKTA_NegativesSK)	Zwischen den Gruppen	3	.427	.366	.778
	Innerhalb der Gruppen	243	1.167		
	Gesamt	246			
Extraversion	Zwischen den Gruppen	3	1.255	1.765	.154
	Innerhalb der Gruppen	239	.711		
	Gesamt	242			

Verträglichkeit (Agreeableness)	Zwischen den Gruppen	3	1.315	2.032	.110
	Innerhalb der Gruppen	239	.647		
	Gesamt	242			
Gewissenhaftigkeit (Conscientiousness)	Zwischen den Gruppen	3	1.734	2.211	.087
	Innerhalb der Gruppen	239	.784		
	Gesamt	242			
Neurotizismus (Neuroticism)	Zwischen den Gruppen	3	.784	.891	.446
	Innerhalb der Gruppen	239	.879		
	Gesamt	242			
Offenheit (Openness)	Zwischen den Gruppen	3	.382	.446	.720
	Innerhalb der Gruppen	239	.857		
	Gesamt	242			
Instruktionsbezogenes Interesse (MOT1INA_Interesse)	Zwischen den Gruppen	3	.395	.515	.672
	Innerhalb der Gruppen	243	.766		
	Gesamt	246			
Instruktionsbezogenes Angst (MOT1INA_Angst)	Zwischen den Gruppen	3	.701	.741	.528
	Innerhalb der Gruppen	243	.945		
	Gesamt	246			
Instruktionsbezogenes Wohlbefinden (MOT1INA_Wohlbefinden)	Zwischen den Gruppen	3	.926	1.069	.363
	Innerhalb der Gruppen	243	.867		
	Gesamt	246			
Ergebnisse Pretest (Z-Wert: SMEAN(PunkGesPre_72))	Zwischen den Gruppen	3	.302	.311	.818
	Innerhalb der Gruppen	248	.972		
	Gesamt	251			

Als ***abhängige Variablen*** bei der Berechnung dienen der Posttest und der Behaltenstest, deren Items, in Fakten- und Konzeptwissen unterteilt und separat berücksichtigt wurden. Um Unterschiede in der Leistung zwischen den einzelnen Schulklassen anzugleichen, wurden die Leistungswerte für Posttest und Behaltenstest jeweils klassenweise z-standardisiert. Dadurch wird die Varianz, die durch den Faktor „Klasse“ als Gruppierungsvariable vorhanden ist, nivelliert. Aufgrund der geringen Stichprobengröße (Anzahl der Klassen: N = 10) können hierarchische Modelle bzw. Mehrebenenanalysen nicht durchgeführt werden (Ditton 1998). Fehlende Daten bei den Post- und Behaltenstests wurden imputiert und im konkreten Fall durch den Mittelwert ersetzt (vergleiche Lüdtke 2007).

Überprüfungen der Bedingungen zum letzten Testzeitpunkt:

Zeitgleich mit dem Posttest wurde der zweite Motivationstest durchgeführt (die ausführliche Beschreibung des Tests befindet sich im Kapitel 5.3.2.5). Alle Schüler beantworteten die Fragen in einem elektronischen Formular im Selbstreport.

Als Signifikanzniveau wurde, wie in den Sozialwissenschaften üblich, die Grenze von .05 gewählt (Bühner 2011).

6.2.3 Reliabilität der Tests

Die Reliabilitäsanalyse wird über Cronbachs α (Alpha) berechnet. „Cronbachs α ist eine nach Lee Cronbach benannte Maßzahl für die interne Konsistenz einer Skala und bezeichnet das Ausmaß, in dem die Aufgaben bzw. Fragen einer Skala miteinander in Beziehung stehen. Es wird angewendet, um die interne Konsistenz eines psychometrischen Instruments zu schätzen“ (Bühner 2011). Dabei sind Werte über 0.7 als „akzeptabel“, über 0.8 als „gut und über 0.9 als „sehr gut“ zu interpretieren.

Tabelle 24: Reliabilität der Tests

	Anzahl der Items	Reliabilität
Pretest gesamt	21	0.488
Pretest Faktenwissen	12	0.444
Pretest Konzeptwissen	9	0.302
Posttest gesamt	21	0.816
Posttest Faktenwissen	12	0.754
Posttest Konzeptwissen	9	0.639
Behaltenstest gesamt	16	0.873
Behaltenstest Faktenwissen	9	0.785
Behaltenstest Konzeptwissen	7	0.745
State Interesse 1	3	0.813
State Interesse 2	3	0.833
State Interesse 3	3	0.846
Wohlbefinden 1	3	0.819
Wohlbefinden 2	3	0.828
Wohlbefinden 3	3	0.835
Langeweile 1	3	0.690
Langeweile 2	3	0.776
Langeweile 3	3	0.793

	Anzahl der Items	Reliabilität
Intrinsische Motivation Interesse (Mot 1)	3	0.747
Intrinsische Motivation Kompetenz (Mot 1)	3	0.829
Intrinsische Motivation Interesse (Mot 2)	3	0.819
Intrinsische Motivation Kompetenz (Mot 2)	3	0.850
Positives Selbstkonzept	3	0.674
Negatives Selbstkonzept	5	0.906
Instruktionsbezogenes Interesse	4	0.812
Instruktionsbezogene Angst	4	0.808
Instruktionsbezogenes Wohlbefinden	4	0.862
FLT 1	16	0.801
FLT 2	14	0.765
Items zur Unterrichtseinheit	6	0.858
Items zur Tablet Nutzung	8	0.794

Die Reliabilität des Pretests ist mit 0.488 nicht besonders gut. Alle anderen Skalen weisen eine akzeptable bis sehr gute Reliabilität auf (siehe Tabelle 24).

Der Big Five wurde keiner Reliabilitätsanalyse unterzogen, da jedes Konstrukt nur aus zwei Items besteht. Dennoch handelt es sich beim Big Five um eine reliable und valide Skala, wie verschiedene Studien belegen (Gosling et al. 2003; Rammstedt & John 2007).

6.2.4 Hypothesenprüfung

Im ersten allgemeinen linearen Modell (multivariat) soll untersucht werden, inwieweit sich die These bewahrheitet, dass Schüler, die während der formativen Leistungstests ein ausführliches Feedback erhalten und dieses auch nutzen (Treatment 1a), beim Post- und Behaltenstest bei der Reproduktion von **Faktenwissen** besser abschneiden, als Schüler der anderen Gruppen. Hierzu wurden bei der Berechnung des Modells mit Posttest und Behaltenstest (Faktenwissen) als abhängiger Variable und Schulleistung, sowie Pretest als Kovariaten, und Geschlecht, Schultyp und Treatment als feste Faktoren, zuerst die Ergebnisse für die multivariaten Tests ausgegeben. Dabei zeigten sich in der multivariaten Analyse ein hochsignifikanter Einfluss des Pretests und ein signifikanter Einfluss der Schulleistung auf das Gesamtset der abhängigen Variablen Post- und Behaltenstest. (siehe Tabelle 25).

Tabelle 25: Erstes allgemeines lineares Modell (mulitvariat)

Effekt		Wert	F	Sig.	Partielles Eta-Quadrat
Konstanter Term	Pillai-Spur	.117	14.387[a]	.000	.117
Pretest Faktenwissen	Pillai-Spur	.132	16.561[a]	.000	.132
FAC1_1_Schulleistung	Pillai-Spur	.039	4.436[a]	.013	.039
AbschlussID	Pillai-Spur	.023	1.246	.291	.011
Geschlecht	Pillai-Spur	.009	1.040[a]	.355	.009
Treatment	Pillai-Spur	.019	.690	.658	.009
AbschlussID * Geschlecht	Pillai-Spur	.025	1.367	.245	.012
AbschlussID * Treatment	Pillai-Spur	.019	.347	.980	.009
Geschlecht * Treatment	Pillai-Spur	.017	.633	.704	.009
AbschlussID * Geschlecht * Treatment	Pillai-Spur	.047	.872	.576	.023

In den nachfolgenden univariaten Tests wurden die Einflüsse der unabhängigen Variablen (Treatment) nun separat jeweils auf den Posttest und den Behaltenstest berechnet. Es zeigte sich keinerlei Einfluss der Treatmentgruppen auf das Faktenwissen in beiden Tests. Sowohl Schulleistung als auch Vorwissen hatten einen signifikanten Einfluss auf die beiden univariaten Modelle.

Tabelle 26: Univariate Tests zum ersten allgemeinen linearen Modell

Quelle	Abhängige Variable	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	Z-Wert: SMEAN(PosttestFaktenWissen)	25	1.897	2.166	.002	.198
	Z-Wert: SMEAN(BehaltenFaktenWissen)	25	1.631	1.911	.007	.179
Konstanter Term	Z-Wert: SMEAN(PosttestFaktenWissen)	1	21.639	24.708	.000	.101
	Z-Wert: SMEAN(BehaltenFaktenWissen)	1	10.706	12.545	.000	.054
Pretest Faktenwissen	Z-Wert: SMEAN(PosttestFaktenWissen)	1	24.040	27.449	.000	.111
	Z-Wert: SMEAN(BehaltenFaktenWissen)	1	13.474	15.787	.000	.067
FAC1_1_Schulleistung	Z-Wert: SMEAN(PosttestFaktenWissen)	1	4.463	5.096	.025	.023
	Z-Wert: SMEAN(BehaltenFaktenWissen)	1	5.668	6.641	.011	.029
AbschlussID	Z-Wert: SMEAN(PosttestFaktenWissen)	2	1.096	1.252	.288	.011
	Z-Wert: SMEAN(BehaltenFaktenWissen)	2	1.044	1.223	.296	.011
Geschlecht	Z-Wert: SMEAN(PosttestFaktenWissen)	1	.151	.173	.678	.001
	Z-Wert: SMEAN(BehaltenFaktenWissen)	1	1.182	1.385	.241	.006

Treatment	Z-Wert: SMEAN(PosttestFaktenWissen)	3	.328	.374	.772	.005
	Z-Wert: SMEAN(BehaltenFaktenWissen)	3	.493	.578	.630	.008
AbschlussID * Geschlecht	Z-Wert: SMEAN(PosttestFaktenWissen)	2	1.301	1.485	.229	.013
	Z-Wert: SMEAN(BehaltenFaktenWissen)	2	1.492	1.748	.177	.016
AbschlussID *	Z-Wert: SMEAN(PosttestFaktenWissen)	6	.432	.493	.813	.013
	Z-Wert: SMEAN(BehaltenFaktenWissen)	6	.172	.201	.976	.005
Geschlecht *	Z-Wert: SMEAN(PosttestFaktenWissen)	3	.602	.687	.561	.009
	Z-Wert: SMEAN(BehaltenFaktenWissen)	3	.213	.250	.861	.003
AbschlussID * Geschlecht * Treatment	Z-Wert: SMEAN(PosttestFaktenWissen)	6	.719	.821	.554	.022
	Z-Wert: SMEAN(BehaltenFaktenWissen)	6	.569	.666	.677	.018
Fehler	Z-Wert: SMEAN(PosttestFaktenWissen)	219	.876			
	Z-Wert: SMEAN(BehaltenFaktenWissen)	219	.853			
Gesamt	Z-Wert: SMEAN(PosttestFaktenWissen)	245				
	Z-Wert: SMEAN(BehaltenFaktenWissen)	245				
Korrigierte Gesamtvariation	Z-Wert: SMEAN(PosttestFaktenWissen)	244				
	Z-Wert: SMEAN(BehaltenFaktenWissen)	244				

Im zweiten allgemeinen linearen Modell (multivariat) sollte untersucht werden, inwieweit die These, dass Schüler, die während der formativen Leistungstests ein ausführliches Feedback erhalten und dieses auch nutzen (Treatment 1a), beim Post- und Behaltenstest bei der Reproduktion von **Konzeptwissen** besser abschneiden, als Schüler der anderen Gruppen, zutrifft.

Hierzu wurden auch hier mit Posttest und Behaltenstest (Konzeptwissen) als abhängiger Variable und Schulleistung, sowie Pretest als Kovariaten, und Geschlecht, Schultyp und Treatment als feste Faktoren, zunächst die Ergebnisse für die multivariaten Tests ausgegeben. Dieses Modell zeigte in der multivariaten Analyse einen signifikanten Einfluss des Pretests (Pillai's Spur = 0.049, F = 5.595, p = 0.004, eta^2 = 0.049) und einen signifikanten Einfluss der Schulleistung (Pillai's Spur = 0.028, F = 3.158, p = 0.044, eta^2 = 0.028). Außerdem hatten die Bedingungen des „Treatments" einen signifikanten Einfluss auf das Erlernen von konzeptuellem Wissen (Pillai's Spur = 0.061, F = 2.297, p = 0.034, eta^2 = 0.031).

Tabelle 27: Zweites allgemeines lineares Modell (multivariat)

Effekt		Wert	F	Sig.	Partielles Eta-Quadrat
Konstanter Term	Pillai-Spur	.044	5.064[a]	.007	.044
Pretest Konzeptwissen	Pillai-Spur	.049	5.595[a]	.004	.049
FAC1_1_Schulleistung	Pillai-Spur	.028	3.158[a]	.044	.028
AbschlussID	Pillai-Spur	.011	.599	.663	.005
Geschlecht	Pillai-Spur	.011	1.221[a]	.297	.011
Treatment	Pillai-Spur	.061	2.297	.034	.031
AbschlussID * Geschlecht	Pillai-Spur	.015	.805	.523	.007
AbschlussID *Treatment	Pillai-Spur	.064	1.203	.278	.032
Geschlecht *Treatment	Pillai-Spur	.022	.815	.559	.011
AbschlussID * Geschlecht * Treatment	Pillai-Spur	.054	1.019	.430	.027

In den nachfolgenden univariaten Analysen war der Effekt des Treatments beim Behaltenstest stärker als beim Posttest, da die Signifikanz beim Posttest den festgelegten Wert von .05 nicht ganz erreicht. Dennoch kann man angesichts des Wertes von .053 von einer Tendenz zur Signifikanz sprechen. Beim Konzeptwissen zeigte sich ein signifikanter Einfluss des Pretest auf den Posttest (p = .001), beim Behaltenstest allerdings nicht. Dafür hatte die Schulleistung einen signifikanten Einfluss auf den Behaltenstest (p = .013). Beides belegt, dass das Vorwissen einen Einfluss auf das Behalten hat.

Tabelle 28: Univariate Tests zum zweiten allgemeinen linearen Modell

Quelle	Abhängige Variable	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	Z-Wert: SMEAN(PostKonzeptWissen)	1.286	1.388	.111	.137
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	1.368	1.482	.072	.145
Konstanter Term	Z-Wert: SMEAN(PostKonzeptWissen)	9.325	10.067	.002	.044
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	2.021	2.190	.140	.010
Pretest Konzeptwissen	Z-Wert: SMEAN(PostKonzeptWissen)	10.123	10.929	.001	.048
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	2.803	3.037	.083	.014
FAC1_1_Schulleistung	Z-Wert: SMEAN(PostKonzeptWissen)	.599	.646	.422	.003
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	5.839	6.327	.013	.028
AbschlussID	Z-Wert: SMEAN(PostKonzeptWissen)	.571	.616	.541	.006
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	.212	.230	.795	.002
Geschlecht	Z-Wert: SMEAN(PostKonzeptWissen)	1.579	1.705	.193	.008
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	.095	.103	.749	.000

Treatment	Z-Wert: SMEAN(PostKonzeptWissen)	2.411	2.603	.053	.034
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	2.838	3.076	.029	.040
AbschlussID * Geschlecht	Z-Wert: SMEAN(PostKonzeptWissen)	1.455	1.571	.210	.014
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	.213	.231	.794	.002
AbschlussID * Treatment	Z-Wert: SMEAN(PostKonzeptWissen)	1.314	1.419	.208	.037
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	.895	.970	.446	.026
Geschlecht * Treatment	Z-Wert: SMEAN(PostKonzeptWissen)	.765	.826	.481	.011
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	.981	1.063	.366	.014
AbschlussID * Geschlecht * Treatment	Z-Wert: SMEAN(PostKonzeptWissen)	.349	.377	.893	.010
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	1.369	1.484	.185	.039
Fehler	Z-Wert: SMEAN(PostKonzeptWissen)	.926			
	Z-Wert: SMEAN(BehaltenKonzeptWissen)	.923			

Aufgrund der Signifikanz wurden die geschätzten Randmittel für die vier Treatmentgruppen in Bezug auf das Konzeptwissen im Post- und Behaltenstest ausgegeben (Tabelle 29) und in den Abbildungen 33 und 34 dargestellt.

Tabelle 29: Geschätzte Randmittel der vier Treatmentgruppen

Abhängige Variable	Versuchsbedingung 2 items für Gruppe 1a	Mittelwert	Standardfehler
Posttest Konzeptwissen (Z-Wert: SMEAN(PostKonzeptWissen))	FLT mit ausführlicher Rückmeldung. nicht genutzt	-.487[a]	.227
	FLT mit ausführlicher Rückmeldung. genutzt	-.073[a]	.218
	FLT mit dichotomer Rückmeldung (richtig/falsch)	.334[a]	.191
	Texte lesen	-.127[a]	.157
Behaltenstest Konzeptwissen (Z-Wert: SMEAN(BehaltenKonzeptWissen)	FLT mit ausführlicher Rückmeldung. nicht genutzt	-.415[a]	.227
	FLT mit ausführlicher Rückmeldung. genutzt	.314[a]	.217
	FLT mit dichotomer Rückmeldung (richtig/falsch)	.291[a]	.191
	Texte lesen	-.222[a]	.157

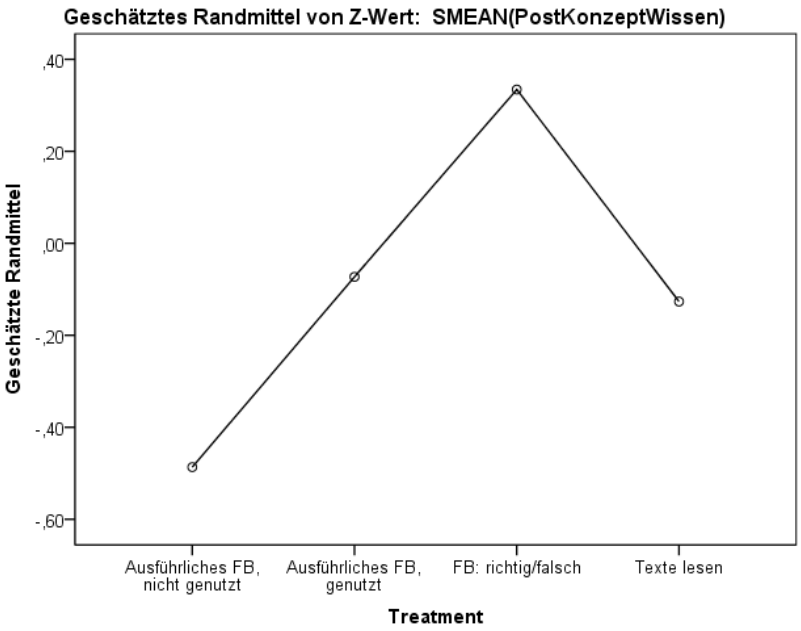


Abbildung 33: Geschätzte Randmittel (Posttest - Konzeptwissen)

Im ***Posttest*** bestehen signifikante Unterschiede zwischen, der Teatmentgruppe 1b, die das Feedback im formativen Leistungstest nicht nutzt, und der Treatmentgruppe 2, die lediglich die Rückmeldung erhält, ob sie richtig oder falsch geantwortet haben (p = .007), sowie zwischen der Treatmentgruppe 2 (dichotomes Feedback) und der Treatmentgruppe 3, die Texte liest und somit als Kontrollgruppe fungiert (p = .066). Keine Unterschiede bestehen zwischen den anderen Treatmentgruppen.

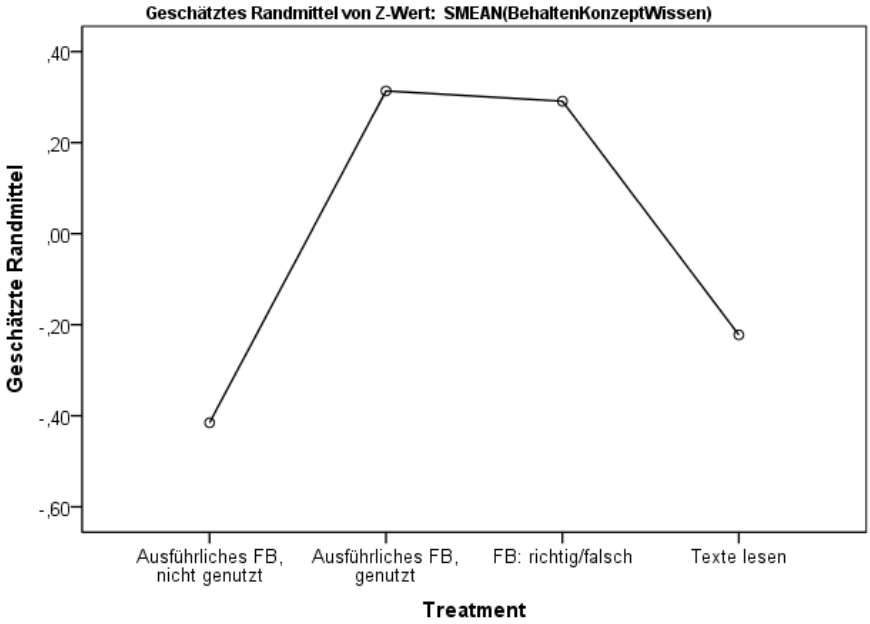


Abbildung 34: Geschätzte Randmittel (Behaltenstest - Konzeptwissen)

Im ***Behaltenstest*** ergeben sich beim Konzeptwissen signifikante Unterschiede zwischen Treatmentgruppe 2 (Dichotom) und Treatmentgruppe 3 „Texte lesen“ (p = .041) und Treatmentgruppe 2 und Treatmentgruppe 1b „keine Feedbacknutzung“ (p = .019). Die Treatmentgruppe 1a, die die ausführlichen Rückmeldungen auch genutzt haben, schnitten ebenfalls besser ab, als die Texte lesende Kontrollgruppe (p = .048) und signifikant besser, als die Schüler des gleichen Treatments (1b), die das Feedback nicht nutzten (p = .022). Keine Unterschiede hingegen gab es beim Lernzuwachs des Treatments 2 und des Treatments 1a (p = .937), sowie zwischen den Schülern der Treatmentgruppe 1b und den Lernern, die sich mit den Texten auseinandersetzten (Treatment 3) (p = .482).

6.2.5 Motivationale Auswirkungen

Bei der Auswertung der Daten zeigte sich, dass die Veränderung in der Motivation der Schüler vom ersten Messzeitpunkt bis zum Posttest nicht signifikant war:

Tabelle 30: Intrinsische Motivation der Schüler

		Mittelwert	N	SD	T	P
Paaren 1	Intrinsische Motivation Interesse (Mot 1)	3.1129	242	.836	-.521	.603
	Intrinsische Motivation Interesse (Mot 2)	3.1446	242	.896		
Paaren 2	Intrinsische Motivation Kompetenz (Mot 1)	3.3581	242	.996	.996	.320
	Intrinsische Motivation Kompetenz (Mot 2)	3.2893	242	1.003		

Die situativen Emotionen:

Tabelle 31: Situative Emotionen

ONEWAY ANOVA

		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
State_Interesse	Zwischen den Gruppen	2.347	2	1.173	1.495	.226
	Innerhalb der Gruppen	200.134	255	.785		
	Gesamt	202.481	257			
State_Wohlbefinden	Zwischen den Gruppen	.414	2	.207	.292	.747
	Innerhalb der Gruppen	181.096	255	.710		
	Gesamt	181.510	257			
State_Langeweile	Zwischen den Gruppen	.299	2	.149	.226	.798
	Innerhalb der Gruppen	168.483	255	.661		
	Gesamt	168.782	257			

Für die Analyse wurde alle drei Messzeitpunkte gemittelt und zu einem Wert zusammengefasst. Bei der Auswertung der Werte zeigen sich keine signifikanten Unterschiede zwischen den Treatments bezüglich der situativen Emotionen Langeweile, Wohlbefinden und Interesse während der gesamten Unterrichtseinheit. Dies sollen folgende Grafiken noch einmal verdeutlichen:

Situative Emotion „Interesse“:

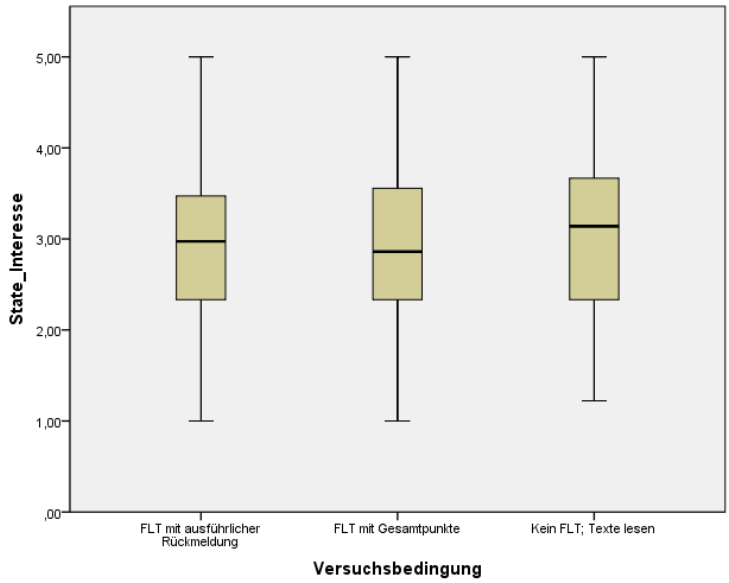


Abbildung 35: Situative Emotion "Interesse"

Situative Emotion „Wohlbefinden“:

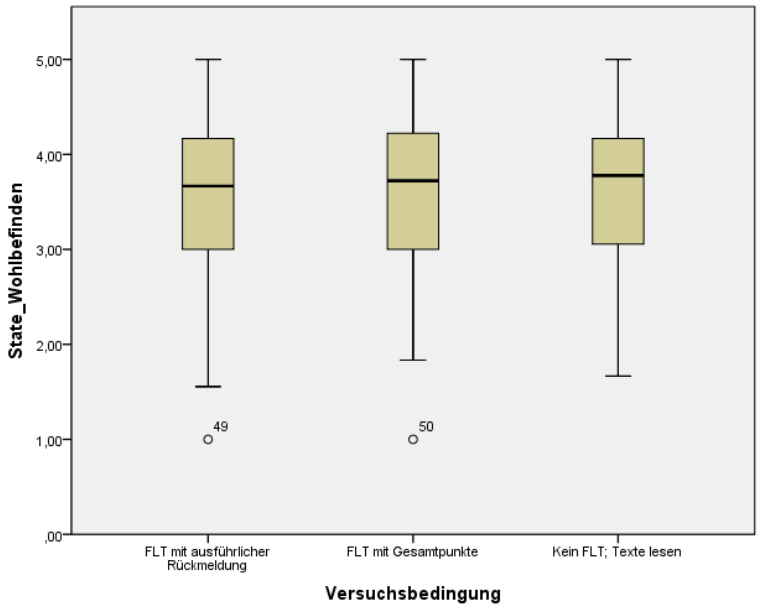


Abbildung 36: Situative Emotion "Wohlbefinden"

Situative Emotion „Langeweile“:

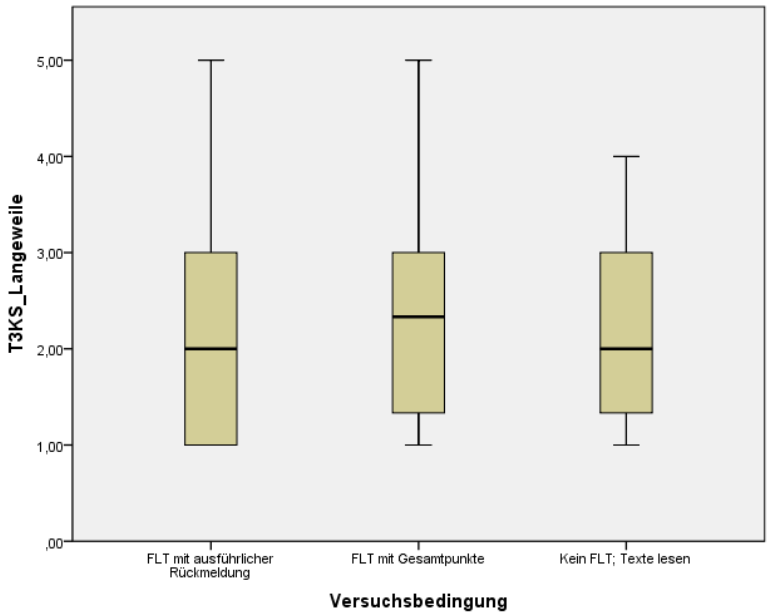


Abbildung 37: Situative Emotion "Langeweile"

Korrelationen

In der unten angeführten Tabelle, kann man die Korrelationen zwischen den *situativen Emotionen* „Interesse, Wohlbefinden und Langweile“, sowie der „Bewertung der Unterrichtssequenz“ mit dem *Wissenszuwachs* genauer betrachten:

Tabelle 32: Korrelationen

		Posttest Konzept Wissen	Posttest Fakten Wissen	Behaltenstest Konzept Wissen	Behaltenstest Fakten Wissen
State_Interesse	Korrelation nach Pearson	.013	.034	.086	.062
	Signifikanz (2-seitig)	.839	.601	.189	.344
	N	244	244	237	238
State_Wohlbefinden	Korrelation nach Pearson	.175**	.178**	.162*	.140*
	Signifikanz (2-seitig)	.006	.005	.013	.031
	N	244	244	237	238
State_Langeweile	Korrelation nach Pearson	-.113	-.086	-.144*	-.132*
	Signifikanz (2-seitig)	.078	.181	.027	.042
	N	244	244	237	238
Bewertung der UE	Korrelation nach Pearson	.033	.079	.119	.111
	Signifikanz (2-seitig)	.608	.213	.065	.084
	N	247	247	240	241

*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.
**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Es zeigte sich, dass das Wohlbefinden signifikant positiv mit den nachfolgenden Wissenstests korrelierte. Langeweile hingegen korrelierte im Behaltenstest negativ mit den Leistungen.

6.2.6 Änderung der Schülervorstellungen zum Konzept der evolutionären Anpassung

Im folgenden Kapitel sollen alle Fragen zur evolutionären Anpassung aus dem Pre- und Posttest noch einmal genauer dargestellt und herausgearbeitet werden, welche Fehlvorstellungen der Schüler zu diesem Konzept besser, welche schlechter behoben werden konnten.
Dafür wurden alle Fragen zu diesem Themenbereich mit all ihren Antwortmöglichkeiten aufgelistet und festgehalten, wie sich die Schülerantworten auf diese Antwortmöglichkeiten beim Vorwissenstest verteilt haben, und wie sich die Anzahl der falschen und richtigen Antworten beim Posttest verändert hat.

Um genau auswerten zu können, welche Teilkonzepte (genetische Variabilität, Natürliche Selektion, etc.) nach der Unterrichtseinheit und den formativen Leistungstests besser verstanden und durchdrungen wurden, wurde am Ende dieses Kapitels eine Tabelle erstellt, die alle richtigen und falschen Antworten zu diesen Teilkonzepten zeigt und dadurch direkt vergleichbar macht. Dabei können manche Teilantworten auch mehreren dieser Fehlvorstellungen zugeordnet sein.

Doch zunächst sollen die Teilkonzepte der Untersuchung vorgestellt werden:

Tabelle 33: Teilkonzepte innerhalb des Konstruktes "Evolutionäre Anpassung"

Thema	Verwendetes Kürzel
Genetische Variabilität	**GENVAR**
Die zufällige Mutation von Genen ist für die Evolution verantwortlich, man kann als Lebewesen nicht gezielt auf eine bessere Anpassung hinarbeiten (z.B. durch Training). Von außen zugefügte Merkmale können nicht vererbt werden. Anpassung ist niemals zielgerichtet oder bedarf einer Absicht.	**MUT/ZIEL**
Die Lebewesen, die das vorteilhafte Merkmal besitzen, setzen sich durch und überleben (bessere Anpassung an den Lebensraum) = natürliche Selektion	**NATSEL**
Evolutionäre Anpassung dauert sehr lange, also mehrere Generationen.	**DAUER**
Evolutionäre Anpassung endet nie – ändern sich äußere Bedingungen in einem Lebensraum, werden sich die Lebewesen, mit den für diesen Lebensraum vorteilhafteren Merkmalen, durchsetzen und überleben.	**ENDE**
Anpassung betrifft nicht nur ein Lebewesen aus einer Population, weil für den Lebensraum positive Merkmale an die nächsten Generationen weitergegeben werden.	**POP**

In der folgenden Auflistung der *Fragen zum Themenbereich „evolutionäre Anpassung“*, wurde jede Antwort einer dieser Teilkonzepte zugeordnet (siehe eingefügtes Kürzel). Alle Items sind selbst konstruiert oder in Anlehnung an unterschiedliche Studien entstanden (Angaben direkt unter der jeweiligen Frage) und für die Zwecke der Studie abgewandelt worden. Die genaue Skalendokumentation befindet sich im Dokument „Fragensammlung“ (bei Interesse anfordern). Die Items wurden in einer Vorstudie getestet und innerhalb des Pre-/Posttest (Konzeptwissen) auf ihre Reliabilität hin überprüft (siehe Gliederungspunkt 6.3.2.1).

Fragentext	**Antwortalternativen**	**Richtig**	**Häufigkeit der Schüler-antworten**	
			Pretest	**Posttest**
In einer Population von Schmetterlingen sind fast alle Schmetterlinge hell, nur ein Schmetterling hat eine dunkle Farbe. Was hat ihn wohl dunkel werden lassen?	Innerhalb einer Population kann es gar keine Schmetterlinge mit unterschiedlichen Farben geben. **GENVAR**		27	13
	Eine zufällige Veränderung im Erbmaterial. **MUT/ZIEL**	x	60	132
	Im neuen Lebensraum mit dunklen Bäumen hat sich der schlauste Schmetterling getarnt, um zu überleben. **MUT/ZIEL**		168	101
Wenn das dunkle Merkmal die Chance zu überleben erhöhen würde, was würde dann mit der Population nach einer längeren Zeitspanne wohl passieren? **GENVAR**	Alle Schmetterlinge würden dunkel werden,		58	33
	Die Anzahl der dunklen Schmetterlinge würde sich erhöhen, es würde aber dennoch hellere Schmetterlinge geben,	x	175	183
	Es würde sich nichts ändern,		20	31
weil	sich alle Lebewesen anpassen, um leichter zu überleben **MUT/ZIEL**		124	99
	vorwiegend die Lebewesen mit dem dunklen Merkmal überleben und dann vermehren **NATSEL**	x	85	113
	sich die geringe Anzahl dunkler Schmetterlinge nicht durchsetzen könnte **NATSEL**		43	34
und	Merkmale, wie die Farbe nicht im Erbmaterial gespeichert sind und somit nicht vererbt werden können. **MUT/ZIEL**		99	63
	durch Variationen im Erbmaterial nie die ganze Population gleich aussehen wird. **GENVAR**	x	145	178

In Anlehnung an Murray (1996)

Fragentext	Antwortalternativen	Richtig	Häufigkeit der Schüler-antworten	
			Pretest	Posttest
Haie können Geschwindigkeiten von bis zu 30 Knoten schwimmen, Ihre Vorfahren waren da noch viel langsamer. Die Fähigkeit schnell zu schwimmen	entwickelte sich bei ALLEN Haien nach ein paar Generationen, **GENVAR**		116	80
	erhöhte sich bei einem Teil der Haie, die dann überlebten und sich vermehrten, **NATSEL**	x	140	167
weil **MUT/ZIEL**	die Haie immer schneller wurden, je mehr sie ihre Muskeln verwendeten und diese trainierten.		49	26
	es zuvor eine zufällige Veränderung im Erbmaterial einzelner Haie gab.	x	45	121
	die Notwendigkeit Beute zu fangen sie dazu zwang, schneller zu schwimmen.		161	99

In Anlehnung an Murray (1996)

Fragentext	Antwortalternativen	Richtig	Häufigkeit der Schüler-antworten	
			Pretest	Posttest
Kreuze den Punkt an, der die Aussage wahr werden lässt! Evolutionäre Anpassung ...	… braucht lange Zeit. Merkt ein Tier, dass es ein bestimmtes Merkmal gibt, das sich vorteilhaft auf sein Leben auswirken würde, dann wird dieses Merkmal nach und nach an die nächsten Generationen weiter gegeben. **MUT/ZIEL** und **DAUER**		116	78
	…findet sofort statt. Durch gezieltes Training kann ein Tier ein vorteilhaftes Merkmal fördern (schnelleres Laufen), und dieses gleich an die nächste Generation weitergeben. **MUT/ZIEL** und **DAUER**		25	11
	… braucht lange Zeit und passiert durch die natürliche Auslese: Ist ein vererbbares Merkmal vorteilhaft für eine Art, überleben die Tiere, die dieses Merkmal besitzen und geben es an weitere Generationen weiter. **NATSEL** und **DAUER**	x	83	138
	… findet sofort statt. Ist es für das Überleben einer Art nötig, dann wird durch Mutation die Erbinformation geändert und gleich an die nächste Generation weiter gegeben. **MUT/ZIEL** und **DAUER**		32	20

Fragentext	Antwortalternativen	Richtig	Häufigkeit der Schüler-antworten	
			Pretest	Posttest
Ende des 19. Jahrhunderts der Zoologe August Weismann folgendes Experiment durch: Er schnitt Mäusen die Schwänze komplett ab, um festzustellen, welche Auswirkung dies auf die Nachkommen haben würde. Bei der ersten Generation von Nachkommen, waren die Schwänze der Mäuse **MUT/ZIEL**	ein wenig kürzer,		97	23
	unverändert,	x	140	217
	komplett verschwunden,		19	6
weil **MUT/ZIEL**	der Anpassungsprozess mehrere Generationen in Anspruch nimmt.		83	56
	ein von außen zugefügtes Merkmal nicht vererbt werden kann.	x	115	179
	Einwirkungen von außen sofort im Erbmaterial gespeichert und an die Nachkommen vererbt werden.		56	12
Schneidet man bei 20 Generationen immer wieder den Schwanz ab, dann sind die Schwänze der 21. Generation **MUT/ZIEL**	ein wenig kürzer,		92	47
	unverändert,	x	74	171
	komplett verschwunden,		87	29
weil **MUT/ZIEL**	Anpassungen sehr lange dauern und sich daher auch erst nach mehreren Generationen durchsetzen.		160	89
	ein von außen zugefügtes Merkmal nicht in den Genen gespeichert und vererbt wird.	x	90	155

In Anlehnung an Johannsen & Krüger (2005)

Fragentext	Antwortalternativen	Richtig	Häufigkeit der Schüler-antworten	
			Pretest	Posttest
Welche Aussagen über die Evolution sind wahr? Kreuze an!	Die Bedingungen in einem spezifischen Lebensraum haben Einfluss auf die Überlebenschance bestimmter Merkmale – besser angepasste Merkmale setzen sich durch, weil vor allem die Lebewesen überleben, die dieses vorteilhafte Merkmal besitzen. **NATSEL**	x	119	140
	Alle Lebewesen einer Population passen sich gleichermaßen an - das heißt, am Ende des Anpassungsvorgangs sehen alle Tiere gleich aus und haben die gleichen Eigenschaften. **GENVAR**		64	31
	Evolution passiert durch zufällige Veränderungen im Erbmaterial. **MUT/ZIEL**	x	85	156
	Anpassung betrifft nur einzelne Lebewesen. **POP**		44	42
	Durch Training lassen sich bestimmte Eigenschaften (zum Beispiel: Schnelles laufen) stärken und diese werden dann an die nächste Generation weitergegeben. **MUT/ZIEL**		83	36
	Ist ein Lebewesen erst mal optimal angepasst, endet der Anpassungsvorgang, denn es gibt nichts mehr zu verbessern. **ENDE**		67	38
	Die Artenvielfalt auf unserem Planeten entstand durch die Variationen und Neukombinationen im Erbmaterial bei der Fortpflanzung. **GENVAR**	x	91	130
	Anpassung findet statt, weil sie für das Überleben einer Art notwendig ist. **MUT/ZIEL**		141	140
	Anpassungsvorgänge benötigen viele Generationen. **DAUER**	x	142	152
	Lebewesen produzieren mehr Nachkommen, als in einem Lebensraum überleben können (Futterangebot, etc.). Dadurch entsteht eine Konkurrenzsituation und nur die besser angepassten Individuen überleben. **NATSEL**	x	91	107

Fragentext	Antwortalternativen	Richtig	Häufigkeit der Schüler-antworten	
			Pretest	Posttest
Wahr oder falsch? Ist ein Tier erst einmal optimal an seinen Lebensraum angepasst, endet die Evolution! **ENDE**	Das ist richtig, denn sind Lebewesen erst einmal optimal an ihren Lebensraum angepasst, können die Lebensbedingungen nicht mehr besser werden und somit findet auch keine weitere Anpassung statt.		42	39
	Das ist falsch, denn Lebensbedingungen und Umwelt können sich immer ändern, somit muss auch immer wieder eine neuerliche Anpassung stattfinden.	x	111	157
	Das ist richtig, Anpassungen verbrauchen von den Lebewesen so viel Energie, dass sie nur im äußersten Notfall stattfinden, deshalb finden sie auch nur sehr selten statt und sind sie einmal abgeschlossen, bleibt es auch dabei.		36	12
	Das ist falsch, Lebewesen streben immer danach, besser zu werden. Wenn Löwen zum Beispiel merken, dass Sie durch Training schneller laufen können und dann ihre Beute leichter fangen können, werden Sie dies tun und ihre antrainierte Muskelmasse an ihre Nachkommen weiter vererben, damit diese einen Vorteil gegenüber den anderen Lebewesen ihrer Art haben und leichter überleben. **ENDE** und **MUT/ZIEL**		67	39

Fragentext	Antwortalternativen	Richtig	Häufigkeit der Schüler-antworten	
			Pretest	Posttest
Die Giraffen auf dem Bild sind hinsichtlich ihrer Halslänge optimal an ihren Lebensraum angepasst - wie wird die Halslänge nach einigen weiteren Generationen aussehen? (Das Bild zeigt drei Giraffen, deren Hals so lange wie der Baum ist, von dem sie fressen.) **GENVAR**	Die Hälse aller Giraffen werden immer länger und gleichen sich an. (Das Bild zeigt drei Giraffen mit gleicher Halslänge, höher als der Baum)		31	35
	Es wird auch in den nächsten Generationen Giraffen mit unterschiedlich langen Hälsen geben. Die Halslänge wird sich dabei um das Optimum herum bewegen. (Bild mit drei Giraffen unterschiedlicher Halslänge, alle ungefähr so hoch wie der Baum)	x	97	130
	Die Hälse der Giraffen sind alle bereits optimal an ihren Lebensraum angepasst und müssen sich nicht weiter verändern. Alle Giraffen der Population werden weiterhin Hälse in der optimalen Länge haben. (Das Bild zeigt drei Giraffen mit gleich langen Hälsen in genauer Baumhöhe).		128	82

In Anlehnung an Johannsen & Krüger (2005) und Kattmann et al. (2009)

Zusammenfassung der Schülerantworten:

Tabelle 34: Schülerantworten zu den konzeptuellen Fragen zur "Evolutionären Anpassung"

Fehlvorstellung / Teilkonzept	Pretest (256 Schüler)						Posttest (247 Schüler)					
	Max. mögl. falsche Antw. pro Schüler und Aufgabe	Max. mögl. falsche Antw. bezo-gen auf alle Schüler*	gewählte falsche Antw.	Max mögl. richtige Antw. pro Aufgabe	Max. mögl. richtige Antw. Bezo-gen auf alle Schüler*	Richtige angekreuzt	Max. mögl. falsche Antw. pro Schüler und Aufgabe	Max. mögl. falsche Antw. bezo-gen auf alle Schüler*	gewählte falsche Antw.	Max mögl. richtige Antw. pro Aufgabe	Max. mögl. richtige Antw. Bezo-gen auf alle Schüler*	Richtige angekreuzt
Genetische Variabili-tät	8	2048	**444**	4	1024	**508**	8	1976	**305**	4	988	**621**
Zufällige Genmuta-tion	18	4608	**1592**	7	1792	**609**	18	4446	**935**	7	1729	**1131**
Natürliche Selektion	1	256	**43**	5	1280	**518**	1	247	**34**	5	1235	**665**
Dauer von Anpas-sungs-vorgängen	3	768	**173**	2	512	**225**	3	741	**109**	2	494	**290**
Evolution endet bei optimaler Anpas-sung	4	1024	**212**	1	256	**111**	4	988	**128**	1	247	**157**
Anpassung betrifft nur einzelne Indivi-duen	1	256	**44**	0	0	-	1	247	**42**	0	0	-

* max. mögliche falsche/richtige Antworten pro Schüler und Aufgabe multipliziert mit der Anzahl der beim jeweiligen Test teilgenommenen Schüler

Ergebnisse:

Tabelle 35: Ergebnisse zu den Schülerantworten im Bereich "Evolutionäre Anpassung"

Fehlvorstellung / Teilkonzept	Pretest		Posttest		delta falsch **	delta richtig**
	falsch pre*	richtig pre*	falsch post*	richtig post*		
Genetische Variabilität	21.7%	50%	15.4%	63%	-6.2%	13%
Zufällige Genmutation	35%	34%	21%	65%	-14%	31%
Natürliche Selektion	17%	40%	14%	54%	-3%	13%
Dauer von Anpassungsvorgängen	23%	44%	15%	59%	-8%	15%
Evolution endet bei optimaler Anpassung	21%	43%	13%	64%	-8%	20%
Anpassung betrifft nur einzelne Individuen	17%	-	17%	-	0%	-

* relative Anzahl der falsch bzw. richtig beantworteten Fragen/Aufgaben (Datenbasis siehe Tabelle 34)

** relative Änderung zwischen Pre- & Posttest bezogen auf die „falschen/nicht korrekten“ bzw. „richtigen/korrekten“ Antworten Ein Vergleich der absoluten Zahlen ist aufgrund der unterschiedlichen Anzahl der Schüler bei Pre-&Posttest nicht möglich!

6.2.7 Der Umgang mit den Tablets , den elektronischen Tests und dem Feedback

Im zweiten Motivationstest wurden die Schüler nach ihren Erfahrungen beim Umgang mit den Tablet PCs, möglichen Schwierigkeiten beim Ausfüllen der elektronischen Leistungstests und der Attraktivität der Lerninhalte befragt.

Zunächst wurde eine *Faktorenanalyse der selbst konstruierten Items* zu diesen Themen durchgeführt:

Diese zeigt zwei Faktoren mit einem Eigenwert > 1 (siehe Tabelle 36), was auch durch die Analyse des Screeplots unterstützt wird (siehe Abbildung 38).

Tabelle 36: Faktorenanalyse - "Umgang mit den Tablets" und "Feedbacknutzung"

Komponente	Anfängliche Eigenwerte		
	Gesamt	% der Varianz	Kumulierte %
1	3.320	41.500	41.500
2	1.312	16.404	57.904
3	.895	11.189	69.094
4	.679	8.483	77.577
5	.559	6.988	84.565
6	.490	6.124	90.688
7	.446	5.580	96.269
8	.299	3.731	100.000

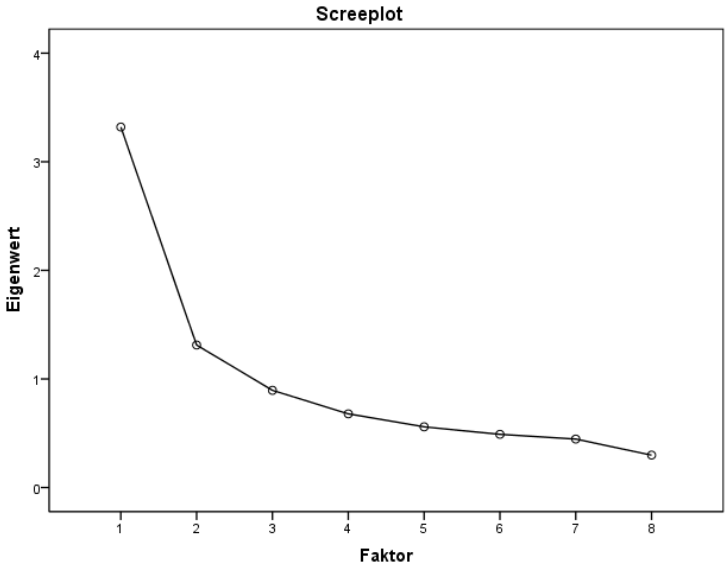


Abbildung 38: Screeplot zur Faktorenanalyse (Tab. 32)

Die Komponentenmatrix zeigt zwei klare Komponenten:

1) Den technischen Umgang und die Bereitschaft mit den Tablets zu arbeiten und
2) Die Kompetenz, Wissenslücken zu identifizieren und diese zu schließen.

Tabelle 37: Komponentenmatrix zu - "Umgang mit den Tablets" und "Feedbacknutzung"

Item	Komponente	
	1	2
Der Umgang mit den Tablet - PCs hat mir Spass gemacht.	**.819**	-.192
Ich hätte die Tests lieber auf herkömmlichem Weg mit Papier und Stift bearbeitet.	**-.698**	.403
Ich bin mit der Bearbeitung der Tests gut klar gekommen.	**.693**	-.031
Ich hatte Schwierigkeiten. bei der Arbeit mit den Tablets.	**-.691**	.405
Die Vielzahl der Tests hat mich genervt.	**-.657**	.142
Durch die Wiederholung des Lerninhaltes war ich besser auf die Probe vorbereitet.	**.540**	.466
Ich habe versucht. diese Wissenslücken zu schließen.	.487	**.638**
Ich wusste nach den Tests. wo meine Wissenslücken sind.	.493	**.551**

Extraktionsmethode: Hauptkomponentenanalyse.

Der Umgang mit den Tablets (Tablet handling):

Tabelle 38: Korrelationen mit dem Tablethandling

		Tablet handling
Posttest Faktenwissen (Z-Wert: SMEAN(PosttestFaktenWissen))	Korrelation nach Pearson	.123*
	Signifikanz (2-seitig)	.047
	N	261
Behaltenstest Faktenwissen (Z-Wert: SMEAN(BehaltenFaktenWissen))	Korrelation nach Pearson	.143*
	Signifikanz (2-seitig)	.021
	N	261
Posttest Konzeptwissen (Z-Wert: SMEAN(PostKonzeptWissen))	Korrelation nach Pearson	.222**
	Signifikanz (2-seitig)	.000
	N	261
Behaltenstest Konzeptwissen (Z-Wert: SMEAN(BehaltenKonzeptWissen))	Korrelation nach Pearson	.166**
	Signifikanz (2-seitig)	.007
	N	261

Es stellte sich heraus, dass das Tablethandling mit dem Wissen korreliert, d.h. wer gut mit den Tablets umgehen konnte zeigte auch eine höhere Leistung im Wissenstest. Dies gilt allerdings auch für den Behaltenstest, der handschriftlich und nicht computergestützt durchgeführt wurde.

Im folgenden Abschnitt soll noch einmal genauer auf die, im Motivationstest 2 erhobenen 12 Einzelitems eingegangen werden, in denen die Schüler die Arbeit mit den Tablets bewerten, und ankreuzen, wie sie mit den elektronischen Tests klar gekommen sind. Auch, wie sie das Feedback genutzt haben soll näher erläutert werden. Dabei wurden die Items FB01 – FB08 von allen Schülern beantwortet, die Fragen zum ausführlichen Feedback (Item FB09 – FB12) hingegen, waren den Lernern der Gruppe 1 vorbehalten.

Die nachfolgenden Tabellen zeigen die Ergebnisse der Befragung. Dabei steht die 1 für „Trifft gar nicht auf mich zu“ und die 5 für „Trifft völlig auf mich zu“.

(FB01) Ich bin mit der Bearbeitung der Tests gut klar gekommen.		
	Absoluter Wert	Relativer Wert
1:	11	4.47 %
2:	23	9.35 %
3:	69	28.05 %
4:	77	31.30 %
5:	66	26.83 %

(FB04) Ich hatte Schwierigkeiten, bei der Arbeit mit den Tablets.		
	Absoluter Wert	Relativer Wert
1:	127	51.63 %
2:	51	20.73 %
3:	32	13.01 %
4:	21	8.54 %
5:	15	6.10 %

In der heutigen Zeit kommen unsere Kinder immer früher mit technischen Geräten in Berührung und wachsen mit diesen auf. Ob Smartphones, Tablets oder Computer – die meisten Schüler der Sekundarstufe können damit umgehen. Deshalb ist es auch nicht weiter verwunderlich, dass die Lerner bei der Bearbeitung der elektronischen Tests keine großen Schwierigkeiten hatten. Auch meine Beobachtungen bezüglich des Umgangs mit den Tablets zeigten, dass die Schüler versiert und intuitiv mit den Geräten umgehen können. Die 14 % der Schüler, die bei der Frage FB01 eine 1 oder 2 angekreuzt haben, waren wahrscheinlich eher mit dem Trennen und Verbinden des WLANs oder dem Einloggen überfordert. Probleme, sie sich relativ schnell lösen ließen und im Laufe der Studie immer weniger auftauchten. Vor allem die Schüler der iPad Klassen benötigten so gut wie keine Anleitung und lösten kleinere technische Probleme (Zulassen der Cookies bei der Arbeit mit der Projekt - Webseite) teilweise sogar eigenständig.

(FB02) Der Umgang mit den Tablet - PCs hat mir Spaß gemacht.		
	Absoluter Wert	Relativer Wert
1:	20	8.13 %
2:	14	5.69 %
3:	24	9.76 %
4:	33	13.41 %
5:	155	63.01 %

Die Motivation beim Umgang mit den Tablets war während der gesamten Studie hoch, auch zum Ende hin, arbeiteten die Schüler noch gerne mit den Geräten, selbst wenn sie darauf nur die vorgegebenen Tests bearbeiten durften. Schade fanden sie allerdings, dass es ihnen nicht erlaubt war, die Tablet PCs zum Spielen oder Surfen zu nutzen. Die hohe Motivation gründet wahrscheinlich darin, dass es den Schülern normalerweise nicht erlaubt ist (außer in den iPad Klassen, wo sie einen festen Platz in den Stunden einnehmen) während des Unterrichts mit technischen Geräten zu arbeiten. Sie kommen gerade einmal im Informatikunterricht oder bei einem motivierten Lehrer im Mathematikunterricht in den Computerraum und dort wird in der Regel auch nicht mit Tablets, sondern mit normalen PCs gearbeitet. Smartphones sind bereits weit verbreitet in der Schüler-

schaft und in den höheren Jahrgangsstufen fast klassendeckend vorhanden. Tablets hingegen sind noch nicht in allen Haushalten zu finden und gelten eher als Luxus. Der Reiz des Neuen ist wohl auch ein Faktor, warum den Schülern die Arbeit mit den kleinen PCs Freude gemacht hat.

(FB03) Ich hätte die Tests lieber auf herkömmlichem Weg mit Papier und Stift bearbeitet.		
	Absoluter Wert	Relativer Wert
1:	160	65.04 %
2:	31	12.60 %
3:	25	10.16 %
4:	7	2.85 %
5:	23	9.35 %

Über 65 % der Lerner fanden, dass sie Tests lieber häufiger auf elektronischem Weg schreiben würden. Dies hat wohl mehrere Gründe: Die möglichen Antworten zu den Testfragen sind bereits vorformuliert, so dass man sich nicht selbst die Mühe machen muss, Fakten und Zusammenhänge genau darzulegen und zu beschreiben. Es reicht, die richtige Antwort auszuwählen und diese zu markieren. Dies ist nicht nur bequem, sondern geht auch schneller. Bild und Videomaterial illustrieren die Fragen und wirken motivierend. Ein nicht zu unterschätzender Faktor ist, das die Schüler gleich nach der Beantwortung der Fragen, eine Rückmeldung erhalten, ob sie falsch oder richtig geantwortet haben – dabei ist die Frage, welche Note sie wohl erreicht hätten wichtiger, als die Frage, wie die richtige Antwort gelautet hätte.

(FB05) Ich wusste nach den Tests, wo meine Wissenslücken sind.		
	Absoluter Wert	Relativer Wert
1:	17	6.91 %
2:	25	10.16 %
3:	84	34.15 %
4:	71	28.86 %
5:	49	19.92 %

(FB06) Ich habe versucht, diese Wissenslücken zu schließen.		
	Absoluter Wert	Relativer Wert
1:	23	9.35 %
2:	27	10.98 %
3:	81	32.93 %
4:	70	28.46 %
5:	45	18.29 %

Ca. 83 % (Ziffern 1-3) der Schüler geben an, dass sie nach den Leistungstests einigermaßen gut wussten, wo ihre Wissenslücken lagen und immerhin ca. 79 % der Lerner haben nach eigenen Angaben auch versucht, diese Wissenslücken zu schließen. Dies wäre, wenn es wirklich zutrifft, ein

schönes Ergebnis, würde es doch ein großes Interesse an der Thematik der Unterrichtseinheit vermuten lassen.

Ob die intrinsische Motivation der Schüler aber wirklich so hoch war, ist zweifelhaft, denn schaut man zum Beispiel die Ergebnisse der Fragen FB09 und FB10 (die immerhin von der Treatmentgruppe 1 beantwortet werden musste), wo es um die Feedbacknutzung geht, genauer an, fragt man sich schon, warum die Schüler nicht gleich die vorliegenden Informationen nutzten, um ihre falschen oder lückenhaften Vorstellungen zu berichtigen…

(FB07) Durch die Wiederholung des Lerninhaltes war ich besser auf die Probe vorbereitet.		
	Absoluter Wert	Relativer Wert
1:	14	5.69 %
2:	31	12.60 %
3:	75	30.49 %
4:	77	31.30 %
5:	49	19.92 %

Zu diesem Thema wurden die Schüler während der Studie einzeln befragt und viele fanden, dass sie furch die formativen Leistungstests und die damit einhergehenden Übungsmöglichkeiten, besser auf die Probe am Ende der Unterrichtseinheit vorbereitet waren. Es bleibt zu vermuten, dass die ca. 18 % der Schüler, die gar keinen oder nur wenig Mehrwert in den Tests sehen, das Angebot auch nicht wirklich genutzt haben. Den meisten Schülern gab es mehr Sicherheit bei der Beantwortung der Fragen, wenn sie sich häufiger konzentriert mit der Thematik auseinandersetzen mussten. Auch die Lehrer waren ganz zufrieden mit den Ergebnissen des Posttests. Der Notendurchschnitt der einzelnen Klassen war meist besser, als bei den normalen Schulaufgaben.

(FB08) Die Vielzahl der Tests hat mich genervt.		
	Absoluter Wert	Relativer Wert
1:	98	39.84 %
2:	57	23.17 %
3:	56	22.76 %
4:	16	6.50 %
5:	19	7.72 %

Erfreulich war, dass keiner der Schüler genervt reagiert oder gar gestreikt hat, wenn „schon wieder“ ein Test anstand. Die meisten Schüler haben durchgängig gut mitgearbeitet und die Tests auch ernst genommen. Dieser Eindruck spiegelt sich auch in den Befragungsergebnissen der Frage FB08 wieder. Lediglich der Behaltenstest, der 6 – 8 Wochen nach dem Posttest stattfand wurde nicht unbedingt mit Begeisterung aufgenommen und teilweise nur ungern bearbeitet. Dies hat wohl

zwei Gründe: Erstens war es ein Paper – Pencil Test mit halboffenen Fragen, wo die Schüler viel schreiben mussten und zweites ging es bei dieser Prüfung auch um nichts mehr (keine Note, etc.).

Die folgenden Fragen mussten nur von den Schülern der ersten Treatmentgruppe beantwortet werden, weil auch nur sie das ausführliche Feedback zur Verfügung hatten. Diese beiden Items wurden auch dafür herangezogen, die Treatmentgruppe 1 (ausführliches Feedback) in die beiden Subgruppen 1a (Nutzung der Rückmeldungen) und 1b (keine Nutzung des Feedbacks) zu unterteilen:

FB09: Ich habe das Feedback ausführlich gelesen.		
	Absoluter Wert	Relativer Wert
1:	8	10 %
2:	9	11 %
3:	24	30 %
4:	17	22 %
5:	21	27 %

FB10: Mir war nur wichtig, ob ich richtig, oder falsch geantwortet habe – die genaue Rückmeldung war mir egal.		
	Absoluter Wert	Relativer Wert
1:	17	22 %
2:	20	25 %
3:	22	28 %
4:	11	14 %
5:	9	11 %

Die Zahlen zeigen, dass ca. die Hälfte der Schüler (FB09 Ziffern 4 und 5 = 49 %) das Feedback gewissenhaft gelesen hat und mit dessen Hilfe versucht hat, zu erfahren, wo die Fehler bei der Beantwortung der Fragen lagen. 30 % der Schüler haben die Rückmeldetexte wenigstens teilweise genutzt, um Fehler zu isolieren und ihre Wissenslücken zu schließen. 21 % geben an, die Rückmeldung eher nicht beachtet zu haben.

In FB10 wird der Umgang der Schüler mit dem Feedback noch etwas konkretisiert: Wurde nur die dichotome Rückmeldung richtig/falsch genutzt, oder wirklich das ausführliche Feedback berücksichtigt? Hier zeigt sich, dass sich die Ergebnisse aus Item FB09 bestätigen.

FB11: Ich habe die weiterführenden Informationen gut.		
	Absoluter Wert	Relativer Wert
1:	3	4 %
2:	10	13 %
3:	26	33 %
4:	24	31 %

5:	15	19 %

FB12: Das Feedback hat mir sehr geholfen.		
	Absoluter Wert	Relativer Wert
1:	6	8 %
2:	15	19 %
3:	26	33 %
4:	20	25 %
5:	12	15 %

Gut die Hälfte der Schüler (FB11 Ziffern 4 und 5) hat die Rückmeldungen gut bis sehr gut bewerteten und nur 17 % der Lerner konnten weniger Nutzen daraus ziehen. Immerhin noch 40 % der Schüler (FB12 Ziffer 4 und 5) hat das gut bis sehr gut bewertete Feedback im weiteren Verlauf der Unterrichtseinheit und Testphase auch gut bis sehr gut weitergeholfen.

6.2.8 Die Bewertung der Unterrichtseinheit

Genau wie die Tablet- und Feedbacknutzung, durften die Schüler auch die Lerninhalte der Unterrichtseinheit bewerten. Dafür wurden 6 Items selbst konstruiert und im folgenden Abschnitt einer Faktorenanalyse unterzogen:

Faktorenanalyse der selbst konstruierten Items zur Unterrichtseinheit:

Die Faktorenanalyse zeigt einen Faktor mit einem Eigenwert > 1 (siehe Tabelle 39), dies wird durch die Analyse des Screeplots unterstützt (siehe Abbildung 39).

Tabelle 39: Faktorenanalyse "Bewertung der Unterrichtseinheit"

Erklärte Gesamtvarianz			
Kompo-nente	Anfängliche Eigenwerte		
	Gesamt	% der Vari-anz	Kumulierte %
1	3.626	60.438	60.438
2	.859	14.313	74.752
3	.535	8.917	83.669
4	.431	7.180	90.849
5	.305	5.075	95.924
6	.245	4.076	100.000
Extraktionsmethode: Hauptkomponentenanalyse.			

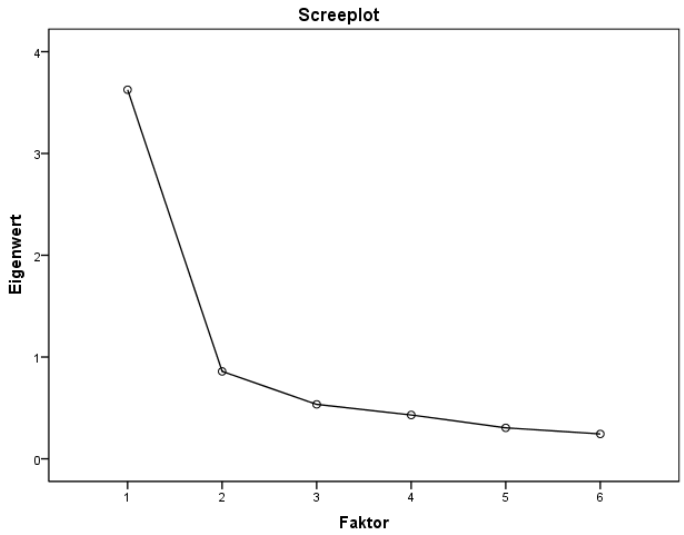


Abbildung 39: Screeplot zur Faktorenanalyse "Bewertung der Unterrichtseinheit"

Die Komponentenmatrix zeigt eine klare Komponente:

Tabelle 40: Komponentenmatrix zu "Bewertung der Unterrichtseinheit"

	Komponente
	1
Ich habe mich gerne mit diesem Thema beschäftigt.	.879
Die Inhalte fand ich spannend.	.853
Ich habe gerne etwas Neues zu diesem Thema erfahren.	.843
Das Thema interessiert mich nicht.	-.781
Ich fand die Inhalte langweilig.	-.719
Auch zu Hause habe ich das Eine oder Andere zu diesem Thema nachgelesen.	.538

Vergleicht man alle drei Treatmentgruppen, hinsichtlich der Bewertung der Unterrichtseinheit, der Tabletnutzung und des Schließens der Wissenslücken, stellt man fest, dass es keine Unterschiede innerhalb des Treatments gab.

Tabelle 41: Unterschiede in den Treatmentgruppen (UE, Tablets, Feedbacknutzung)

ONEWAY ANOVA

		Quadrat-summe	df	Mittel der Quadrate	F	Signi-fikanz
Bewertung der UE	Zwischen den Gruppen	.832	2	.416	.438	.646
	Innerhalb der Gruppen	245.168	258	.950		
	Gesamt	246.000	260			
Tablet handling	Zwischen den Gruppen	1.194	2	.597	.629	.534
	Innerhalb der Gruppen	244.806	258	.949		
	Gesamt	246.000	260			
Wissenslückenschließer	Zwischen den Gruppen	.277	2	.139	.146	.865
	Innerhalb der Gruppen	245.723	258	.952		
	Gesamt	246.000	260			

6.2.9 Teilnehmende Beobachtung

Während der Testzeitpunkte hatte ich, dadurch, dass ich bei den Schülern öfters technische Hilfe leisten musste, bzw. beim Abgaben und Öffnen der Tests unterstützen musste, leider nicht so viel Zeit, die Lerner bei der Arbeit mit den Tablets und den elektronischen Leistungstests zu beobachten. Doch immer, wenn sich die Gelegenheit bot, wurde diese auch wahrgenommen und folgende Dinge beobachtet:

- Die Schüler kommen mit dem Aufgabenformat und den Tablets größtenteils gut bis sehr gut zurecht. Sie sind es gewohnt mit elektronischen Medien zu arbeiten und besitzen meistens auch ein Smartphone, das ähnlich funktioniert.
- Das ausführliche Feedback wurde wenig gelesen. Schüler wollten vor allem die erreichte Punktzahl wissen und welche Note dies in einem bewerteten Test gewesen wäre.
- Die Kontrollgruppe war häufig unzufrieden damit, dass sie „nur“ Texte lesen, die anderen Schüler aber die Tests bearbeiten durften. Einige waren sehr schnell mit dem Lesen der Texte fertig (manche Schüler klickten einfach nur auf „weiter“ und gaben nach nur 5 Minuten ab).
- Es gab große Unterschiede bei den Schularten, was die Auseinandersetzung mit den Rückmeldetexten anging: Am Gymnasium beschäftigten sich die Schüler länger und konzentrierter mit dem angebotenen, ausführlichen Feedback. Auch die Treatmentgruppe 3 las die Texte gewissenhafter durch. Die intrinsische Motivation der Gymnasiasten war, wie zu erwarten höher, als die vieler Schüler aus den anderen Schularten.
- Die Arbeit mit den iPads, war aufgrund ihrer Größe einfacher, als die Nutzung der anderen Tablets mit 7-Zoll Display: Dort wurden die Auswahlmenüs nicht vollständig angezeigt – dadurch mussten die Schüler scrollen, um alle Antwortmöglichkeiten sehen zu können. Und trotz der Einstellung in Moodle, dass das Endgerät ein Tablet PC mit kleinem Display ist, wurden teilweise Texte abgeschnitten, so dass ein paar Aufgaben manuell geändert und dadurch angepasst werden mussten.

Ein Schüler (Gesamtpunktzahl: 26 von 46 Punkten beim FLT 1), der das Feedback bei keiner Aufgabe gelesen hatte und bei dem die Zeit zwischen "prüfen" und dem Beginn der nächsten Aufgabe im Wesentlichen mit dem Laden der neuen Aufgabe verging, wurde nach der Bearbeitung seines Tests kurz befragt:

Warum hast du das Feedback nicht gelesen?
S: Das kommt nicht in der Schulaufgabe (benoteter Test) dran.
Wie findest du es, Aufgaben auf dem Tablet zu bearbeiten?
S: Der Test ist OK, weil man nichts schreiben muss.
Sind die Fragen verständlich?
S: Ja.

Der Schüler möchte wissen, welche Note er mit dieser Punktzahl bekommen würde…

6.3 Diskussion der Ergebnisse

Der wichtigste Aspekt der Studie war es, herauszufinden, ob Feedback bei der Durchdringung von Faktenwissen und konzeptuellen Wissens im naturwissenschaftlichen Unterricht helfen kann und welche Art von Feedback sich dabei als besonders geeignet erweist. Auch die Frage, welche Rückmeldeformate sich bei elektronischen Leistungstests anbieten, war im Focus des Interesses.
In den folgenden Abschnitten sollen die in Kapitel 6.2 dargestellten Ergebnisse der Studie diskutiert werden. Dabei wird vor allem das methodische Vorgehen und die Wahl der Instrumente kritisch hinterfragt und der Frage nachgegangen ob es Sinn macht, elektronische formative Leistungstests einzusetzen, oder ob traditionelle handschriftliche Tests doch effektiver sind. Außerdem sollen die aufgestellten Hypothesen (siehe Gliederungspunkt 3) überprüft und näher erläutert werden.

6.3.1 Methodisches Vorgehen

6.3.1.1 Die Unterrichtssequenz

Aufgrund der bereits beschriebenen Forschungsdefizite hinsichtlich der formativen Leistungsmessung in den naturwissenschaftlichen Fächern, wurde die Studie im Fach Biologie durchgeführt. Dieses Vorgehen bietet sich auch deshalb an, da es in der Biologie weitreichende Konzepte gibt, die sich über mehrere Jahrgangsstufen erstrecken und viele fachliche Teilaspekte miteinander vernetzen. So streift das Konstrukt der evolutionären Anpassung Fachbegriffe und weiterführender Themen, wie das der Genetik (Genmutation, genetische Variabilität, Artenvielfalt), der Entwicklungsgeschichte der Lebewesen, der Anpassung an bestimmte Lebensräume, der Bionik, der ökologischen Nische, usw. Aufgrund dieser großen Vernetzung, war es möglich, in der Unterrichtssequenz nicht nur Faktenwissen zu vermitteln, sondern auch konzeptuelles Wissen und die damit einhergehenden Zusammenhänge näher zu beleuchten. Eine besondere Herausforderung bestand darin, die Fehlvorstellungen der Schüler, die sich im Laufe ihres Lebens durch Erfahrungen und falsche Schlussfolgerungen gebildet haben, aufzudecken und diese zu berichtigen (siehe hierzu Kapitel 2.3.1.1). Man kann also verschiedene Wissensebenen (Fakten- und Konzeptwissen) vermitteln und untersuchen, bei welchen Wissensarten, welche Rückmeldungen am meisten Sinn machen.

Leider bietet sich das Fach Biologie nicht an, adaptive elektronische Tests einzusetzen, weil sich die Aufgaben nicht auf Teilprobleme reduzieren lassen, die man dann gezielt durch Übungsaufgaben beheben kann.
Mathematische Probleme hingegen lassen sich besser in Teilschritte unterteilen, die man, jedes für sich genommen üben kann (genaueres über die adaptiven Verfahren, kann man unter 5.3.3.3 nachlesen).

Da während der Studie in den einzelnen Klassen nur eine begrenzte Anzahl an Stunden zur Verfügung stand, mussten hinsichtlich der Unterrichtseinheit, ein paar Abstriche gemacht werden. Man hätte die Lerninhalte an einigen Stellen vertiefen und sich zum Beispiel noch genauer mit flugunfähigen Vögeln / flugfähigen Wirbeltieren auseinandersetzen können. Auch das Thema „Bionik“ ist äußerst spannend und bietet sich für das trainieren biologischer Arbeitsweisen, wie dem Durchführen von genauen Beobachtungen und eigenen Forschungen durch die Schüler an, aber das Einbinden weiterer Themen hätte den Rahmen der Studie gesprengt. Darum sind die Inhalte letztendlich gut gewählt und stimmig aufeinander abgestimmt und gerade so weit gefasst, dass die wichtigsten Inhalte des übergeordneten Themas „Anpassung der Vögel an das Fliegen“ behandelt werden konnten.

Die Unterrichtssequenz selbst begann an einem Punkt, an dem die Schüler ganz unterschiedliches Vorwissen zu diesem Themenbereich mitbrachten, denn es gab vorher keinen vorbereitenden Unterricht, um alle auf den gleichen Wissensstand zu bringen.
Die komplette Unterrichtseinheit (mit Stundenvorschlägen, Übungsaufgaben, Experimentierheften, Arbeitsblättern und sämtlichen Materialien) wurde von der Versuchsleiterin entworfen und allen Lehrkräften zur Verfügung gestellt. So bekamen alle Lerner den gleichen Unterricht und die gleichen Materialien. Lediglich die Methodik des Unterrichts (Schülerexperiment, Lehrerexperiment oder Frontalunterricht) und die Anpassung der Stunden (z.B. Wahl der richtigen Übungsaufgaben) für die jeweilige Klasse, war den Lehrern überlassen. So wurde sichergestellt, dass alle Schüler nach Beendigung der Studie, den gleichen Wissensstand aufweisen müssten. Zunächst stand zur Diskussion, den Lehrern lediglich Vorgaben hinsichtlich der Stundenanzahl und der Lernziele zu machen, aber durch die strengen Vorgaben, wurde eine einheitliche Durchführung der Erhebung in allen Klassen erst möglich. Auch wurde eine unterschiedliche Schwerpunktsetzung der einzelnen Lehrkräfte vermieden. Ein weiterer, wichtiger Vorteil der genauen Festlegung war im organisatori-

schen Bereich verankert: Alle Tests und Kurzskalen waren an genauen Zeitpunkten verortet und es konnte gut geplant werden, wann der nächste Erhebungszeitpunkt anstehen musste.

Im zweiten Motivationstest (Kapitel 5.3.2.5) konnten die Schüler die Lerninhalte im Selbstreport bewerten. Dabei konnten sie ankreuzen, ob sie diese langweilig fanden und wie gerne sie sich mit dem Thema beschäftigt haben, usw.
Vergleicht man alle drei Treatmentgruppen, hinsichtlich der Bewertung der Unterrichtseinheit, stellt man fest, dass es keine Unterschiede in den Gruppen gab. Dies lässt sich dadurch erklären, dass die Unterrichtssequenz an sich für die Gesamtstichprobe gleich war.
Sieht man sich die einzelnen Ergebnisse der Befragung an, kann man zusammenfassend festhalten, dass das Thema mittelmäßig interessant für die Schüler war. Der Großteil der Lerner hat sich bei der Frage, ob sie die Inhalte spannend / langweilig fanden für eine mittlere Einschätzung (Kreuz bei 3) entschieden. Knapp 40 % haben sich immerhin in der Schule ganz gern mit dem Thema auseinandergesetzt und fanden es gut, Neues über die Lerninhalte zu erfahren, aber die Frage, ob sie sich zu Hause weiter mit dem Thema beschäftigten, also selbst Einzelheiten nachgelesen hätten, beantworteten nur ca. 22 % mit „ja“.
Dies ist wahrscheinlich bei den meisten schulischen Themen so. Viele Schüler finden die Themen mäßig interessant, einige können sich für das eine oder andere Thema begeistert und nur wenig bringen genug Interesse und intrinsische Motivation mit, um sich auch selbständig zu Hause weiter in die Inhalte zu vertiefen.

Wünschenswert wäre es, über die naturwissenschaftlichen Lerninhalte hinaus, auch Inhalte aus anderen Fächergruppen für vergleichbare Studien heranzuziehen.

6.3.1.2 Das Treatment

Eine der Stärken dieser Studie bestand darin, dass die Klassen während des gesamten Zeitraums und Unterrichts im Klassenverbund zusammen waren und somit immer die gleichen äußeren Bedingungen (Raum, Zeit, Störungen, etc.) vorfanden. Auch die Unterteilung in die drei Treatmentgruppen während der formativen Leistungstests fand innerhalb der Klasse statt. Neben den äußeren Bedingungen, war auch die Lehrkraft konstant, dadurch konnte es nicht zu zusätzlichen Varianzen in den Untersuchungsergebnissen kommen (Randler 2012). Auch während der Testzeitpunkte wurde versucht möglichst homogene Bedingungen unter den drei Gruppen zu schaffen: Alle Schüler hatten gleich lange Zeit, sich mit dem Test / den Texten zu beschäftigen und alle haben sich mit

den gleichen Lerninhalten auseinandergesetzt. So war es möglich, vergleichbare Ergebnisse zu erhalten.

Folgendes Treatment wurde bei der Studie verwendet:

Treatmentgruppe 1 – formativer Leistungstest mit ausführlichem Feedback.
Bei der Auswertung der Ergebnisse wurde die Treatmentgruppe 1 in zwei Subgruppen (1a: Feedback genutzt und 1b: Feedback nicht genutzt) unterteilt.
Diese Aufteilung wurde nötig, um eine Aussage darüber treffen zu können, ob ein ausführliches Feedback einen größeren Lernzuwachs nach sich zieht, als eine dichotome Rückmeldung (Treatment 2), durch die die Schüler nur erfahren, ob sie richtig oder falsch geantwortet haben. Denn diejenigen Schüler, die die ausführlichen Rückmeldetexte zwar zur Verfügung hatten, diese aber nicht genutzt haben, kamen ja auch nicht dazu, die eventuellen positiven Einflüsse, die ein ausführliches Feedback auf den Lernzuwachs haben könnte, für sich zu nutzen.

Treatmentgruppe 2 – formativer Leistungstest mit dichotomen Feedback (richtig / falsch).
Zu Beginn der Studie stand die Überlegung im Raum, neben der Rückmeldung, ob die Frage richtig oder falsch beantwortet wurde, den Schüler auch eine Übersicht zur Verfügung zu stellen, der sie entnehmen können, welche Antwort die richtige gewesen wäre. Dies wäre allerdings sehr nah am Untersuchungsdesign der ersten Treatmentgruppe gewesen, so dass eine eindeutige Trennung für die Untersuchung nicht mehr möglich gewesen wäre, denn letztendlich hätten sich die beiden Gruppen nur noch im Grad der Ausführlichkeit der Rückmeldetexte unterschieden. Nach der Überprüfung der Hypothesen, stellt sich allerdings die Frage, ob gerade diese Art der Rückmeldung, die effektivste sein könnte, aber dies wird näher in Kapitel 6.3.3 (Hypothesenprüfung) diskutiert.

Treatmentgruppe 3 – Texte lesen.
Hier soll noch einmal angemerkt werden, dass beim Design der dritten Treatmentgruppe sehr harte Kontrollbedingung gewählt wurden, denn die Kontrollgruppe bekam keine fachfremden Texte zu lesen oder gar Unterricht in einem anderen Fach, sondern inhaltlich die gleichen Informationen zur Verfügung gestellt, wie die Treatmentgruppe 1. Die Texte gingen sogar noch ausführlicher auf die bereits erarbeiten Lerninhalte der vorhergehenden Unterrichtsstunden ein, indem sie alle Themen noch einmal detailliert wiederholten und darstellten. Das strenge Kontrolldesign zeigt sich nicht

nur in inhaltlicher, sondern auch in zeitlicher und motivationaler Hinsicht, denn auch die Schüler der dritten Gruppe durften die Tablets nutzen und mussten nicht mit Texten aus Büchern oder ihrem Schulheft arbeiten und hatten genauso viel Zeit, wie alle anderen.
Es wäre denkbar gewesen, noch eine weitere Kontrollgruppe „0“ zu haben, die

1) nur den Pre- und Posttest vor und nach der Unterrichtseinheit schreibt (ohne die beiden formativen Leistungstests), oder sogar
2) nur den Pre- und nach einigen Wochen den Postest schreibt, ohne überhaupt Unterricht zur Thematik zu erhalten.

Doch erstens war die Stichprobe nicht groß genug, um weitere Treatmentgruppen zu bilden, und zweitens hätte man im ersten Fall eher untersucht, ob formative Leistungsmessung an sich etwas bringt und im zweiten Fall, wie gut die Unterrichtseinheit war. Beides sollte nicht genauer betrachtet werden und somit wurde die Entscheidung getroffen, keine weiteren Gruppen zu bilden.

Die Überprüfung, inwieweit sich der Wissenszuwachs bei elektronischen und traditionellen, handschriftlichen formativen Tests unterscheidet, war ebenfalls nicht im Focus der Studie und wurde deshalb nicht berücksichtigt.

Hätte man genügend Lehrkräfte und Schüler zur Verfügung gehabt, wäre es sehr interessant gewesen, zu untersuchen, wie stark sich eine Rückmeldung an die Lehrkraft und die daraus resultierende Adaption des nachfolgenden Unterrichts auf die Leistungen der Lerner im Post- und Behaltenstest auswirkt:

Wenn eine Lehrkraft gleich 2 Klassen parallel führt, könnte man in einer der beiden Klassen formative Leistungstests schreiben lassen und, genau wie in dieser Studie eine elektronische Rückmeldung an die Schüler geben. In der zweiten Klasse könnte die Lehrkraft genau ansehen, welche Fragen der formativen Leistungstests die Schüler richtig oder falsch hatten (sogar auf Teilantwortebene möglich) und im anschließenden Unterricht auf die Denkfehler und Probleme der Schüler gezielt eingehen. Wahrscheinlich wären hier die Effekte auf die Leistung am größten. Eine Untersuchung dieser Feedbackvariante in einem nachfolgenden Projekt wäre wünschenswert.

6.3.1.3 Das Feedback

Im Allgemeinen lässt sich sagen, dass Feedback an sich einen positiven Effekt auf das Lernen und Behalten von konzeptuellen Lerninhalten hat, denn die Treatmentgruppen, die eine Rückmeldung

erhalten haben, schnitten im Posttest besser ab, als die Kontrollgruppe, die kein Feedback zur Verfügung hatte. Dieses Ergebnis deckt sich mit früheren Studien zur formativen Leistungsmessung (Hattie & Timperley 2007; Black & Wiliam 2009), den Feedbackstudien (Bangert-Drowns et al. 1991) und den Feedbacktheorien (Kluger & DeNisi 1996). Dabei blieb der Effekt des Feedbacks moderat (in einem mittleren Bereich), was ebenfalls Parallelen zu bisherigen Forschungsergebnissen aufweist. Vor allem bei komplexeren Lerninhalten (Kluger & DeNisi 1996; Kingston & Nash 2011), wie dem Erlernen naturwissenschaftlicher Konzepte und Zusammenhänge zeigten sich bisher niedrige bis mittlere Effekte. Bei dieser Evaluation kann man diese Effekte in einer randomisierten Feldstudie beobachten, einer Studie, die direkt in den naturwissenschaftlichen Unterricht eingebettet und nicht unter Laborbedingungen durchgeführt wurde.

Diskussion des Feedbackdesigns

In der folgenden Tabelle werden noch einmal alle Feedbackmöglichkeiten hinsichtlich der Darreichungsform, Ausführlichkeit, Verfügbarkeit, etc. dargestellt.

Grün hervorgehoben sind diejenigen Optionen, die in der Studie verwendet wurden. Die Optionen, die hinsichtlich eines Leistungszuwachses aus Sicht der Versuchsleiterin sehr erfolgversprechend scheinen und zukünftig genauer untersucht werden sollten, sind blau dargestellt.

Feedback	**Einsatzmöglichkeiten**			
Bei	Falschen Antworten		Richtigen Antworten	Richtigen und falschen Antworten
Von	Anderen Schülern (auch einer Schülergruppe)		Der Lehrkraft (zusätzlich)	Dem Computer
An	Den Lernenden		Den Lehrenden	Beide
In Bezug auf	Die Aufgabe	Den Lösungsweg (Kompetenzförderung) *Hier nicht möglich, da adaptives System notwendig.*	Die Person (Leistung, Arbeitshaltung, Lernfortschritt)	Den Unterricht (Verlauf und weitere Lernschritte)

Feedback	**Einsatzmöglichkeiten**			
Hinsichtlich der Ausführlichkeit	Richtig / Falsch	Richtig / Falsch und hervorheben der richtigen Antwortmöglichkeiten	Detaillierte Rückmeldung zu richtigen und falschen Antworten	Detaillierte Antwort mit der Möglichkeit zusätzliche Informationen einzuholen. (PC)
In Hinsicht der Verfügbarkeit	Vollständig vorhanden		Hinweis, wo man richtige Antwort finden kann	Gar kein Hinweis: Schüler sollen eigenverantwortlich nachlesen (Steigerung: Kompetenzerwerb)?
Hinsichtlich des Zeitpunktes	Direkt nach der Aufgabe		Direkt nach dem Test	Erst in den darauffolgenden Stunden
Sonstiges	Positive, negative oder neutrale Formulierungen?	Schriftlich / mündlich / beides	Individuell, auf den jeweiligen Fehler eingehend oder allgemeine Formulierungen?	In der Gruppe (zusätzliches Besprechen der Ergebnisse mit der gesamten Klasse) – für den Einzelnen

Im Nachhinein betrachtet wäre es besser gewesen, bereits von der ersten Vorstudie an, mit elektronischen Tests zu arbeiten, um sich verstärkt auf die Inhalte der Rückmeldung konzentrieren zu können, anstatt auf deren Art der Darreichung. Es ging viel Zeit verloren, zu versuchen, bei einem traditionellen Test in Schriftform, jedem Schüler, direkt nach Bearbeitung der Tests, individuelles Feedback zu geben. Dies ist so gut wie unmöglich und geht auch nur auf sehr unübersichtliche und komplizierte Art und Weise mit Hilfe von Kontrollfolien, etc. Die Möglichkeiten eines individuellen Feedbacks bestehen bei Tests in herkömmlicher Schriftform eher darin, dass die Lehrkraft, alle Tests korrigiert und auswertet und im nachfolgenden Unterricht auf die Schüler zugeht, die besonders große Probleme beim Verständnis der erarbeiteten Lerninhalte zeigten.

Dann kann man aber wirklich nur sehr kurze formative Leistungserhebungen durchführen, die aus wenigen, kurzen Fragen bestehen und dadurch schnell auswertbar sind. Auf die Vor- und Nachteile von elektronischen Tests (vgl. 5.4.1) wird aber zu einem späteren Zeitpunkt (unter 6.3.1.4) noch einmal eingegangen.

Wie bei der Diskussion über das Treatment bereits angemerkt, wäre eine dichotom gestaltete Rückmeldung (richtig/falsch) in Kombination mit einer kurzen Darstellung, welche Antworten richtig gewesen wären, sowohl für die Schüler, als auch für die Lehrkraft am effektivsten. (siehe Feedbackinterventionstheorie nach Kluger & DeNisi (1996, 1998)).

Für die Schüler deshalb, weil sie keine ausführlichen Texte mehr lesen müssen (viele tun dies prinzipiell nicht), sondern auf einen Blick sehen, was sie falsch gemacht haben und was richtig gewesen wäre. Die Erfahrungen aus der Studie zeigen, dass die Lerner vor allem wissen wollen, wie viele Punkte sie erzielt haben und welche Note dies bedeuten würde. Hinsichtlich der Inhalte ist nur interessant, was falsch war und was richtig wäre, nicht aber, warum es falsch war. Eine nochmalige ausführliche Auseinandersetzung mit dem Lerninhalt ist häufig auf Schülerseite nicht gewünscht. Für die Lehrer ist ein klarer Vorteil, dass die Erstellung der Tests sehr viel schneller geht, wenn man kein ausführliches Feedback formulieren muss. Wie und wann das Feedback (richtig / falsch und richtige Antwort) angezeigt werden soll, kann man per Mausklick einstellen und bedeutet demnach nur einen geringen Aufwand. Besser ist es, wenn die Lehrkraft die gewonnene Zeit dafür einsetzt, Aufgaben noch einmal mit allen Schülern im Unterrichtsgespräch zu klären, die häufig nicht korrekt von den Lernern beantwortet wurden.

Insgesamt war das Design der Rückmeldungen in Ordnung, nur einzelne Formulierungen hätten noch einfacher sein können.

Leider konnten einige Dinge hinsichtlich des Feedbacks, die bei der Untersuchung wünschenswert gewesen wären, durch die Einschränkungen von Moodle nicht berücksichtigt werden:

- Adaptives Feedback könnte sehr nützlich sein, geht aber mit einem großen Programmieraufwand einher und ist nur möglich, wenn die Lehrkraft informationstechnisch sehr versiert ist. Einsetzbar sind adaptive Rückmeldungen auch nur, wenn eine bestimmte Aufgabenstruktur vorhanden ist. So können Aufgaben in Mathematik in kleine Teilschritte unterteilt werden, die ein lösungsprozessbezogenes Feedback erlauben. Damit kann man gezielter

auf individuelle Stärken und Schwächen eingehen und Strategien zur Leistungsverbesserung anhand von Beispielaufgaben zur Verfügung stellen.

- Eine ausführliche Begleitung des individuellen Lernfortschritts jedes einzelnen Schülers wäre eventuell eine weitere Art der Motivation gewesen. In der Vorstudie konnten die Schüler in ein Diagramm eintragen, wie viele Punkte sie bei welchem Wissensgebiet bei den einzelnen Wissenstests hatten und ihren Lernfortschritt beobachten. Dies haben die Kinder gerne gemacht, denn neben der Beobachtung der eigenen Fortschritte, wurde auch ein direkter Vergleich mit den Mitschülern möglich. Leider konnte man sich in Moodle keine individuelle Lernverlaufskurve anzeigen lassen. Erschwert wurde ein direkter Vergleich der Testzeitpunkte auch, weil nicht alle Tests die gleichen Fragen beinhalteten.

Optimal wäre, wenn die Schüler anhand dieser Lernverläufe zusätzlich lernen würden, zu erkennen, welche Defizite sie haben und wenn sie im Anschluss selbst versuchen würden, diese aufgedeckten Wissenslücken zu schließen. Lerntagebücher zu führen und mit deren Hilfe eine Veränderung im Arbeitshaltung und Lernverhalten der Schüler zu erzielen braucht aber auch einen längeren Zeitraum (mehrere Monate oder gar Jahre) und kann nicht Teil einer kurzen Studie sein. Dennoch ist die Hinführung der Schüler zu einem selbstregulierten Lernverhalten das wichtigste Ziel formativer Leistungsmessung.

6.3.1.4 Die Technik

Aufgrund des Untersuchungsdesigns der Studie war es mehr oder weniger zwingend erforderlich, elektronische formative Leistungstests einzusetzen (siehe hierzu auch: 4.2.2), denn nur so wurde eine Variation des Feedbacks möglich. Welche Lernplattform und welches Autorensystem man für die Entwicklung der Tests und die Speicherung von Schülerdaten verwendet, bedurfte allerdings größeren Überlegungen und Abwägungen. Näheres über die verschiedenen Möglichkeiten, kann man in Kapitel 5.4 nachlesen.

Letztendlich war die Wahl von Moodle als Autorensystem und Lernplattform wohl die einfachste und effektivste Lösung. Natürlich wäre es schön gewesen, auch andere Aufgabenformate, wie Drag and Drop einzusetzen, aber die Entwicklungen im elektronischen Bereich gehen so schnell voran, dass es wahrscheinlich auch nicht lange dauern wird, bis derartige Formate in das bestehende System aufgenommen oder deren Einbindung zumindest möglich sein wird. Es ist einfach ein komfor-

tabler Rundumservice, der von der Lehrkraft kein allzu großes technisches Know How erfordert und ein schnelles Erstellen unterrichtlicher Inhalte und Tests erlaubt. Auch für die Studie hat sich die Wahl als sehr gut herausgestellt, denn abgesehen von ein paar Dingen, wie zum Beispiel der wenig komfortablen Auswertemöglichkeiten der Motivationstests, war die Durchführung und das Ergebnis sehr zufriedenstellend. Das System hat gut funktioniert, es gab keine größeren technischen Aussetzer und alle Daten wurden zuverlässig erfasst. Nur einmal wollte Moodle bei einer Kennung nicht funktionieren – leider war es nicht möglich herauszufinden, warum. Nachdem dieser Fehler aufgetreten war, wurden immer Ersatzteilnehmer angelegt, um schnell reagieren zu können, wenn der Fehler noch einmal auftreten sollte. Derartige Routinen entwickelt man erst wenn man längere Zeit mit einem System gearbeitet hat. Würden Lehrkräfte Moodle über einen längeren Zeitraum im Unterricht einsetzen, wären sie wahrscheinlich nach und nach auf alle Eventualitäten eingerichtet. Der größte Vorteil allerdings war der, dass das System autark funktionierte und dadurch keine Verbindung zu den Schulnetzen nötig war. Dies hätte den organisatorischen Aufwand bei Weitem überschritten und hätte wohl zum Scheitern der Evaluation geführt.

Moodle wird bereits in zahlreichen Bildungseinrichtungen verwendet und nach und nach immer weiter ausgebaut. Es ist nur eine Frage der Zeit, bis alle Bausteine, die man sich wünscht zur Verfügung stehen. Legt man allerdings großen Wert auf ein individuelles Design und der Testfragen und eventuell auch adaptive Rückmeldungen, wird man wohl um eigene Programmierungen oder die Nutzung kommerzieller Systeme nicht herumkommen.
Über die *Vor- und Nachteile elektronischer Tests* wurde bereits ausführlich berichtet. Hier noch einmal eine kurze tabellarische Zusammenfassung der wichtigsten Punkte:

Tabelle 42: Vor- und Nachteile elektronischer Leistungstests

Vorteile	Nachteile
Automatische Auswertung der Tests	Hoher Entwicklungsaufwand bei der ersten Ausarbeitung und Entwicklungen von elektronischen Tests und Unterrichtseinheiten. Dafür ist später eine schnelle Änderungen und Anpassungen möglich.
Hohe Motivation der Schüler	Nicht jeder Aufgabentyp bietet sich für elektronische Leistungstests an. Halboffene und offene Fragen sind schwer auswertbar, zeigen aber besonders gut, inwieweit Konzepte verstanden und durchdrungen wurden.

Testfragen können in ansprechender Form und als interaktive Aufgaben entwickelt werden. (Drag and Drop, Hot Spot, Experimente planen und ablaufen lassen, etc.)	Hohe Anschaffungskosten für Endgeräte (Tablets) oder Wechsel in den Computerraum
Verwendung unterschiedlicher Medien möglich: Der Einsatz von Animationen und Filmen erlaubt eine bessere Veranschaulichung der Lerninhalte.	Die Infrastruktur an den Schulen ist teilweise schlecht, oft auch gar nicht vorhanden → Fehlende Computerräume, kein WLAN, etc.
Ein direktes und individuelles Feedback an den Schüler ist eigentlich nur elektronisch möglich.	
Lerninhalte, Unterrichtssunden, Lernmaterialien und Wissenstests können in einer großen Datenbank gesammelt und allen zur Verfügung gestellt werden, so dass der Entwicklungsaufwand einzelner immer mehr abnimmt.	
Das diagnostische Potential von Aufgaben könnte voll ausgeschöpft werden.	
Sofortige Adaptierbarkeit des weitergehenden Unterrichts.	

Der Umgang mit den Tablets

Vergleicht man alle drei Treatmentgruppen, hinsichtlich der Tabletnutzung (siehe 6.2.7), stellt man fest, dass es keine Unterschiede zwischen den einzelnen Gruppen gab. Dies lässt sich unter anderem dadurch erklären, dass fast alle Schüler mit dem Umgang von Smartphones und Tablets gleichermaßen vertraut sind – es gibt keine Berührungsängste mit der Technik auf Schülerseite, denn sie wachsen damit auf. Je höher man in den Jahrgangsstufen geht, desto flächendeckender besitzen die Schüler ein eigenes Smartphone.

Bei der Auswertung der Ergebnisse zeigte sich auch, dass der Umgang mit den Tablets mit dem Wissenszuwachs korreliert, d.h. wer gut mit den Tablets umgehen konnte, zeigte auch eine höhere Leistung im Wissenstest.
Diese Korrelation könnte auftreten, weil

1) die klugen Schüler generell besser im Unterricht und in den Tests sind und demnach auch leichter mit den Tablets zurechtkommen, oder
2) die Schüler, die besser mit dem Tablets umgehen konnten sich eher auf die Inhalte konzentrierten und damit auch besser in den Tests abgeschnitten haben.

Allerdings haben die Schüler, die beim Umgang mit dem Tablet wenige Probleme hatten auch im Behaltenstest, der handschriftlich und nicht computergestützt durchgeführt wurde, höhere Ergebnisse erzielt. Dies spricht für eine höhere Wahrscheinlichkeit des ersten Erklärungsversuchs.

Die Motivation bei der Arbeit mit den Tablet PCs ist hoch, dies liegt wahrscheinlich an mehreren Dingen:
Normalerweise sind elektronische Geräte im Unterricht untersagt (außer bei den iPad Klassen). Dabei gäbe es gerade hier sehr viele Einsatzmöglichkeiten (vgl. Kapitel 2.2.2.1 und 5.4.1).
Die Tests selbst werden von den Schülern lieber auf elektronischem Weg erledigt, denn dies bedeutet weniger Arbeit – man muss keine eigenen Antworten formulieren.
Die Lerner erhalten eine sofortige Rückmeldung und auch gleich ihre Note und müssen nicht erst warten, bis die Arbeit korrigiert ist. Wünschenswert wäre, wenn die Schüler eine individuelle Lernkurve anzeigen lassen könnten, wo sie genau sehen können, in welchen Bereichen sie bereits besser geworden sind und wo es noch Schwierigkeiten gibt, die man durch gezielte Arbeit abbauen kann.
Die genauen Antworten der Schüler im zweiten Motivationstest hinsichtlich des Umgangs mit den Tablets, werden in Kapitel 6.2.7 ausführlich dargestellt und besprochen.

Die Hardware

Wie bereits erwähnt, war die Durchführung der Studie nur möglich, weil ein autarkes System mit eigenem WLAN, eigenem Server und eigener Datenbank zur Verfügung stand.

Was die Auswahl der Endgeräte betrifft, so muss man sagen, dass es sowohl vorteilhafte, als auch nachteilige Punkte gab (vgl. auch 5.4.2.2.2). Im folgenden Abschnitt möchte ich noch einmal kurz auf die wichtigsten Dinge eingehen:

Tabelle 43: Vor- und Nachteile von 7'' Tablets

Vorteile	Nachteile
Das kleine Format der 7‘‘ Tablets erlaubte es, einen Klassensatz von 30 Stück in einem Rollkoffer mit sich herum zu tragen. Größere Displays hätten mehr Gewicht und Platz bedeutet, was die Mobilität mehr eingeschränkt hätte.	Die 7‘‘ Displays der Tablets waren allerdings auch nachteilig bei der Anzeige der Inhalte: In den Drop Down Menüs wurden nicht alle Antwortmöglichkeiten angezeigt, so dass die Schüler scrollen mussten, um alle betrachten zu können. Außerdem wurden Sätze abgeschnitten, die man dann manuell in den Einstellungen überarbeiten musste. Beim größeren

	Display der I Pads gab es da keine Probleme.
Beide verwendeten Tablet PCs hatten ein gutes Preis – Leistungsverhältnis und ließen sich gut bedienen.	Die Administration der Geräte und das Laden waren sehr aufwendig, weil man immer jedes Gerät einzeln konfigurieren musste. Jede Änderung in den Einstellungen musste also 30 Mal hintereinander manuell durchgeführt werden. Auch zum Laden musste jeder PC extra angeschlossen werden. Beim I Pad gibt es einen Koffer mit Ladeeinheit, in dem man alle I Pads gemeinsam laden, konfigurieren und transportieren kann, was den administrativen Aufwand sehr reduziert.

Zusammenfassend lässt sich sagen, dass die Situation/die äußeren Bedingungen die Auswahl der Tablets bestimmen: Hat man die Möglichkeit, dass jeder Schüler der Klasse sein eigenes Tablet anschafft, es selbst transportiert und im Klassenverband nach Anweisung der Lehrkraft konfiguriert, bietet es sich an, ein größeres Display zu wählen. Da macht auch das Recherchieren und Arbeiten mit elektronischen Materialien mehr Freude. Gut ist immer, wenn die Lehrkraft das gewünschte Endgerät ausgiebig testet, denn es ist nicht motivierend mit einem schlecht reagierenden Touchscreen arbeiten zu müssen und lange Ladezeiten in Kauf zu nehmen. Preislich muss man wohl immer schauen, was man sich leisten möchte und kann. Für die Studie wurde ein guter Kompromiss gefunden, der gut funktioniert hat.

Sollte man über eine Arbeit mit elektronischen Materialien im Unterricht nachdenken, kann man sagen, dass generell viele unterschiedliche Endgeräte denkbar sind. Man kann sowohl die Computerräume der Schule nutzen (falls vorhanden), als auch Tablet PCs (Kostenfrage), oder auch die Smartphones der Schüler. Letzteres ist wohl die preislich günstigste Lösung und hat den Vorteil, dass jeder Lerner mit seinem eigenen Gerät bereits vertraut ist. Warum nicht die vorhandenen Kapazitäten nutzen, anstatt sie immer zu verbieten? Natürlich ist bei der Nutzung der Smartphones schon bei der Entwicklung der elektronischen Unterrichtsmaterialien und Tests darauf zu achten, dass diese plattformunabhängig sind und somit von jedem Fabrikat einsetzbar sind.

Machen elektronische formative Leistungstests im naturwissenschaftlichen Unterricht Sinn?

Die Ergebnisse der Studie zeigen Effekte, beim Einsatz elektronischer Leistungstests, doch die Ergebnisse der Wang Studie (2011) sind klarer und vielversprechender. Allerdings gab es bei Wang kein so eingegrenztes Treatment wie in der vorliegenden Evaluation. Hier waren die Unterschiede

zwischen Treatment- und Kontrollgruppe geringer (nur das Feedback wurde variiert) und die Schüler hatten mehr Freiheitsgrade innerhalb ihrer elektronischen Lernumgebung und beim Umgang mit den bereitgestellten elektronischen Unterrichtsmaterialien. Eine ausführlichere Beschreibung der Studie von Wang (2011) findet man im Kapitel 2.2.2.1.

Was meine Studie klar zeigt ist, dass es viel Zeit braucht, um elektronische Leistungstests und elektronisches Lernen an sich, in den Unterricht zu integrieren. Der Einsatz von Tablets ist sowohl für die Schüler, als auch für die Lehrkräfte, die technischen Neuerungen aufgeschlossen gegenüberstehen, motivierend - dieser Einsatz bedarf aber auch ein gewisses Maß an Engagement und Fachwissen. Meine Erhebung ist erst ein erster Schritt, wenn es darum geht, computerunterstützte formative Leistungsmessung zu untersuchen und diese für den naturwissenschaftlichen Unterricht attraktiv zu gestalten. Wie Hickey und Zuiker (2012) denke ich, wird es mehrere Jahre dauern, bis sich diese Art des Lernens und Prüfens an die Bedürfnisse naturwissenschaftlichen Unterrichts angepasst hat und größere Effekte auf das Lernen ausübt.

6.3.1.5 Organisatorisches / Ablauf der Evaluation

Insgesamt lässt sich sagen, dass die Studie zufriedenstellend, ohne größere Probleme abgelaufen ist. Alle teilnehmenden Lehrkräfte waren sehr kooperativ und motiviert.

Wünschenswert wären eine größere Stichprobe und noch mehr teilnehmende Gymnasien gewesen, um die Unterschiede beim Umgang mit dem Feedback in den unterschiedlichen Schularten noch genauer untersuchen zu können.

Es hat sich als positiv herausgestellt, die Testzeitpunkte (FLTs) von 3 auf 2 zu verringern und die Unterrichtseinheit dadurch ein wenig zu kürzen, denn auch so hat sich die Studie über einen sehr langen Zeitraum hingezogen (ca. 9 Wochen).

Nachdem bereits beim zweiten Messzeitpunkte ein paar Schüler ihre Logins verlegt hatten, war ich auf dieses Problem vorbereitet und habe alle Benutzerkennungen der Schüler in den folgenden Klassen gleich in eine Klassenliste eintragen lassen, um sofort herausfinden zu können, welche Kennung fehlte. Bei der ersten Klasse, wo ich für dieses Problem noch nicht sensibilisiert war, habe ich beim zweiten Tests mühsam herausfinden müssen, welche Kennung die Schüler besaßen, die ihre Daten verloren hatten und alle anderen mussten warten, bis es los gehen konnte.

6.3.2 Die Messinstrumente

Zusammenfassend lässt sich sagen, dass die Lerninhalte und Tests sehr unterschiedlich von den Kindern / Lehrkräften in Hinsicht auf den Schwierigkeitsgrad eingeschätzt wurden. Dadurch, dass die Schüler aus verschiedenen Jahrgangsstufen und unterschiedlichen Schularten kamen, waren sie mehr oder weniger mit Transferaufgaben vertraut. Auch die Migrationsdichte in den Klassen war sehr unterschiedlich ausgeprägt. Leider konnte man die Inhalte nur ansatzweise an die Gegebenheiten der einzelnen Klassen anpassen, denn sonst wäre die Studie nicht mehr überall identisch gewesen.

6.3.2.1 Reliabilität der Tests

Die Reliabilität des Pretests ist mit 0.488 nicht besonders gut, dies ist aber oft bei Wissenstests der Fall (vgl. Diaz-Morales et al. 2012). Dies liegt daran, dass man streng genommen nicht von der Reliabilität eines Tests sprechen kann, sondern immer nur von der Messgenauigkeit einer bestimmten Messung. Im Rahmen eines Testverfahrens werden häufig mehrere Messwerte ermittelt, die eine unterschiedliche Reliabilität aufweisen können (vergleiche Bühner 2011). Dies zeigt sich auch darin, dass der identische Test zu einem anderen Zeitpunkt (Posttest - nach der Unterrichtseinheit) eine Reliabilität von 0.816 aufweist. Wahrscheinlich ist die unzureichende Reliabilität des Pretests auf die Heterogenität des Vorwissens der Schüler zurückzuführen, die vor der Unterrichtseinheit stärker ausgeprägt ist, als nach der Sequenz. Die Reliabilität aller anderen Tests und Skalen kann als gut bis sehr gut eingestuft werden.

6.3.2.2 Pre- und Posttest

Den Pre- und den Posttest inhaltlich gleich zu gestalten war notwendig, um den Wissenszuwachs genau untersuchen zu können. Insgesamt waren die Tests mit 21 Fragen sehr lang. Dies war aber nicht zu umgehen, denn nur so konnte man sich einen Überblick verschaffen, was die Schüler über die Lerninhalte der gesamten Unterrichtssequenz an den verschiedenen Messzeitpunkten wussten.

Möchte man in der Schule mit formativen Leistungstests arbeiten, ist es nicht unbedingt notwendig das Vorwissen zu einer bevorstehenden Unterrichtseinheit zu erheben, außer man möchte den Lernern die Möglichkeit geben, in einer individuellen Lernkurve die eigenen Fortschritte zu verfolgen.

6.3.2.3 Formative Leistungstests

Die formativen Leistungstests bezogen sich inhaltlich nur auf die vorhergehenden Unterrichtsstunden. Man hätte diese ebenfalls identisch zum Pre- und Posttest gestalten können, aber von dieser Möglichkeit wurde aus folgenden Gründen Abstand genommen:

21 Items wären zu umfassend gewesen. Im normalen Unterricht, wo man als Lehrkraft seine eigenen Tests erstellen kann, ist es effektiver, mehrere kleine Übungstests zu schreiben. Diese kann man normalerweise am Ende der Stunde in wenigen Minuten durchführen. Vor allem, wenn die Lerner immer mit Tablets arbeiten und diese bereits gestartet sind. Dadurch haben die Schüler häufigere Übungsmöglichkeiten und arbeiten kontinuierlich an den Themen / ihren Problemen. Innerhalb der Studie sind noch häufigere, dafür kürzere Tests allerdings organisatorisch nicht möglich, denn zu jedem Testzeitpunkt muss die Versuchsleiterin anreisen, das Equipment aufbauen, etc.
Weiterer Vorteil: Konzentriert man sich auf bestimmte Lerninhalte, kann man deren Verständnis mit zusätzlichen Fragen stärker vertiefen und besser üben.
Außerdem ist es nicht sinnvoll, in den formativen Leistungstests auch immer die Lerninhalte mit abzufragen, die im weiteren Verlauf der Unterrichtseinheit erst noch kommen werden, denn dann haben die Aufgaben keinen Übungscharakter, sondern laden zum Raten ein.

Eine mehrmalige Bearbeitung der einzelnen Testfragen, wurde ausgeschlossen, denn dann hätten die Schüler so lange raten können, bis sie die richtige Antwort gefunden haben. Innerhalb der Studie macht dies keinen Sinn, denn da sollte ja der Wissenstand zum jeweiligen Testzeitpunkt erhoben werden. Im normalen Unterrichtsgeschehen kann ein anderes Vorgehen durchaus seine Berechtigung haben, denn bei der Suche nach der richtigen Antwort wird ja auch schon geübt und gelernt.

Während der Evaluation, wurden einzelne Schüler gefragt, ob sie die formativen Leistungstests als sinnvoll erachten. Dabei kam heraus, dass sie es durchaus als eine Übungsmöglichkeit und bessere Vorbereitung auf die Probearbeit am Ende der Sequenz wahrnahmen.
Durch die Tests sind vor allem die Schüler dazu „gezwungen“, sich im Unterricht noch einmal ausführlich mit den Lerninhalten auseinanderzusetzen, die es zu Hause nicht hinreichend tun würden.

Auch ist es eine Möglichkeit für die Lehrkraft, häufiger individuelle Rückmeldungen zu geben, auch wenn sie nur aus „richtig oder falsch“ besteht, die sie in den großen Klassen nicht jedem einzelnen geben könnte.

6.3.2.4 Behaltenstest

Beim Behaltenstest stellte sich die Frage, ob der Posttest einfach noch einmal wiederholt werden sollte, oder ein handschriftlicher Test mit halboffenen und offenen Fragen besser sein würde.
Letztendlich wurde die zweite Variante als vorteilhafter eingestuft, obwohl dann ein direkter Vergleich zum Posttest nicht mehr möglich war. Die Gründe für diese Entscheidung werden im Folgenden kurz aufgeführt:
Durch die Beantwortung von halboffenen Fragen, waren die Schüler gezwungen, ihr erlerntes Wissen genauer zu reflektieren und vor allem, ihr Verständnis der Inhalte selbst zu formulieren. Bei den elektronischen Tests waren alle Antwortmöglichkeiten bereits vorhanden und ausformuliert – die Lerner mussten sich nur noch für eine der Möglichkeiten entscheiden… Nur so konnte untersucht werden, inwieweit die Kinder das konzeptuelle Wissen wirklich verstanden haben und wie sicher sie mit ihrem erworbenen Wissen umgehen konnten.
Außerdem wollte man sicher gehen, dass die Schüler sich die richtigen Antworten nicht nur anhand einiger Stichworte gemerkt hatten, bzw. die richtigen Antworten durch den Übungseffekt innerhalb des gleichen Testdesigns, herausfanden.

Insgesamt war es sehr schwer, die Schüler, zu einer gewissenhaften und leistungsbereiten Bearbeitung des Behaltenstests zu motivieren. Dafür gab es wohl mehrere Gründe:

- Das Ergebnis des Tests war irrelevant für sie. Es hatte keine Bedeutung mehr für ihre Note.
- Sie hätten einen elektronischen Test bevorzugt – die Beantwortung der Fragen war ihnen zu anstrengend.
- Die Studie war gedanklich bereits beendet – eine nochmalige Auseinandersetzung mit den Lerninhalten schien nicht mehr sinnvoll.

6.3.3 Hypothesenprüfung

Bevor die Hypothese anhand der Forschungsergebnisse auf ihre Richtigkeit hin überprüft wird, soll die in Kapitel 3 aufgestellte Hypothese noch einmal aufgegriffen werden:

Schüler, die während der formativen Leistungstests ein ausführliches Feedback erhalten und dieses auch nutzen, schneiden beim Post- und Behaltenstest bei der Beantwortung von Fragen zum

a) Fakten- und

b) Konzeptwissen

signifikant besser ab, als Schüler, die kein elaboriertes Feedback erhalten, oder die angebotene Rückmeldung nicht nutzen.

Um die Evaluationsergebnisse strukturierter auswerten und erläutern zu können, wird die Hypothese im folgenden Abschnitt etwas differenzierter und auf das Treatment der Studie zugeschnitten formuliert:

Schüler, die während der formativen Leistungstests ein ausführliches Feedback erhalten und dieses auch nutzen (Treatment 1a), schneiden beim Post- und Behaltenstest bei der Beantwortung von Fragen zum

a) Fakten- und

b) Konzeptwissen signifikant besser ab, als Schüler der anderen Treatmentgruppen.

Besser als:

1) Treatmentgruppe 1b (ausführliches Feedback – nicht genutzt)
2) Treatmentgruppe 2 (dichotome Rückmeldung)
3) Treatmentgruppe 3 (Texte lesen)

Die Überprüfung der Hypothese muss also von verschiedenen Standpunkten aus durchgeführt werden. Zum einen soll eine Unterscheidung der beiden untersuchten Wissensebenen (Fakten- und Konzeptwissen) vorgenommen und zum anderen innerhalb dieser Unterteilung gezielt auf den Wissenszuwachs der einzelnen Treatmentgruppen geschaut werden. Hier steht die Frage im Vordergrund, welche Art von Feedback am effektivsten ist, wenn es um das Erlernen und Behalten von einfachen (Fakten) und komplexen (naturwissenschaftliche Konzept) Lerninhalten geht.

Die Unterteilung der Testfragen in Fakten- und Konzeptwissen wurde von mir durchgeführt:

Die Testfragen, die gezielt Zusammenhänge und Inhalte der Konzepte „Auftrieb“ und „evolutionäre Anpassung“ aufgriffen, wurden dem Konzeptwissen, diejenigen Testfragen, die nur reproduzierbare Fakten (Schnabelformen, Federarten, etc.) abfragten, dem Faktenwissen zugeordnet.

Einen genaueren Überblick über die einzelnen Testfragen und deren Zuteilung, kann man sich mit Hilfe der Datei „Fragensammlung“ verschaffen (Dokument bei Interesse bitte anfordern).
Die Zuteilung der Schüler in die jeweiligen Treatmentgruppen erfolgte nach dem Zufallsprinzip.

6.3.3.1 Faktenwissen

Bei der Auswertung der Evaluationsdaten im Bereich „Faktenwissen“ hat sich gezeigt, dass nur die Schulleistung und das Vorwissen einen Einfluss auf das Treatment haben, nicht aber die Ergebnisse der beiden Posttests.

- Hochsignifikanter Einfluss des Pretests (p = .000) und
- signifikanter Einfluss der Schulleistung (p = .013) auf die abhängigen Variablen.

Daraufhin wurden in univariaten Tests die Einflüsse der unabhängigen Variablen jeweils separat auf den Posttest und den Behaltensstest berechnet. Auch hier konnte kein Einfluss der Treatmentgruppen auf das Faktenwissen in beiden Tests festgestellt werden.

- Kein Einfluss der Treatmentgruppen auf das Faktenwissen in beiden Tests (Posttest p = .772; Behaltenstest p = .630).

Die These, dass Schüler durch das Bereitstellen ausführlicher Rückmeldungen in formativen Leistungstests im summativen Test am Ende der Unterrichtseinheit bei der Beantwortung von Fragen zum Faktenwissen besser abschneiden, als Schüler, die dieses Feedback nicht zur Verfügung haben, hat sich demnach als falsch herausgestellt.

Die Ergebnisse weisen darauf hin, dass es bei der Wiederholung von reinen Daten und Fakten nicht darauf ankommt, wie ausführlich man auf die Schülerantworten in formativen Leistungstests eingeht. Eine Erklärung hierfür könnte sein, dass es bei dieser Art von Fragen nur um die reine Reproduktion von erlerntem Wissen geht, bei dem in erster Linie nur interessant ist, ob richtig oder falsch geantwortet wurde, um eine Korrektur im Schülergedächtnis vorzunehmen. Es müssen in der Rückmeldung keine Zusammenhänge erläutert oder richtig gestellt werden, weil diese nicht Teil der gefragten Inhalte sind. Eine genaue Erläuterung, in welcher Hinsicht die Frage falsch beantwortet wurde, ist also nicht unbedingt erforderlich. Warum die Antwort falsch war, ist bekannt: Der Schüler hat sich nicht genügend mit den Lerninhalten auseinandergesetzt. Detaillierte Rück-

meldungen sind an dieser Stelle eher verwirrend und zeitraubend. Besser man hebt sich die ausführlichen Erklärungen für die konzeptuellen Fragen und Transferfragen auf, damit die Schüler dann noch genügend konzentriert und motiviert sind, um auf das bereitstehende Feedback genauer einzugehen.
Wichtiger beim Erlernen von Faktenwissen ist es, im Unterricht selbst genügend sinnvolle und anschauliche Beispiele zu bringen, die es den Lernern erleichtern, die dargebotenen Fakten zu erlernen und zu behalten. Häufiges Wiederholen ist ebenfalls zuträglich, wenn es darum geht, Inhalte dieser Wissensebene im Gedächtnis zu verankern.

Dies stimmt auch mit früheren Forschungsergebnissen überein: Unmittelbares Feedback auf der Aufgabenebene (z.B. *knowledge of results*) führt bei wenig komplexen Aufgaben (z.B. Silben memorieren) zu hohen Effektstärken (vgl. Bangert-Drowns et al. 1991). Dieser Feedbacktyp ist umso effektiver, je schneller die Rückmeldung erfolgt.

Dadurch, dass auch während der Studie der Unterricht in den Klassen normal weitergeführt wurde, also zu Beginn jeder Unterrichtsstunde die Inhalte der vorhergehenden Lerneinheiten in Abfragen wiederholt wurden, konnten die Schüler aller Treatmentgruppen gleichermaßen den Vorteil der regelmäßigen Wiederholung nutzen. Die Rückmeldung, welche Antworten korrekt oder falsch waren, bekamen also alle Lerner im Unterrichtsgespräch noch einmal zusätzlich zum Feedback in den formativen Leistungstests durch die unterrichtende Lehrkraft. Dies kann die geringen Unterschiede innerhalb des Treatments erklären.

6.3.3.2 Konzeptwissen

Betrachtet man die Auswertung der Evaluationsergebnisse im Bereich Konzeptwissen genauer, stellt man bei der Berechnung des allgemeinen linearen multivariaten Modells fest, dass neben der Schulleistung und dem Vorwissen auch die Ergebnisse der beiden Posttests einen signifikanten Einfluss auf das Treatment hatten.

- Signifikanter Einfluss des Pretests (p = .004),
- der Schulleistung (p = .044) und
- **des Treatments** (p = .034) auf die abhängigen Variablen.

Berechnet man in den nachfolgenden univariaten Analysen die Einflüsse der unabhängigen Variablen jeweils separat auf den Posttest und den Behaltensstest, zeigt sich, dass der Effekt des Treatments beim Behaltenstest stärker als beim Posttest ausfällt.

- Einfluss des Treatments beim Behaltenstest stärker (p = .029) als
- beim Posttest (p = .053).
- Signifikanter Einfluss des Vorwissens auf den Posttest (p = .001) und
- der Schulleistung auf den Behaltenstest (p = .013).

Aufgrund dieses Ergebnisses wurden die Werte für die einzelnen Treatmentgruppen, getrennt für den Post- und den Behaltenstests errechnet (geschätzte Randmittel).

*Dabei stellte sich folgendes für den **Posttest** heraus:*

Es gab keine signifikanten Unterschiede zwischen der Treatmentruppe 1a und den übrigen Treatmentgruppen!

Es bestehen lediglich signifikante Unterschiede zwischen,

- der Teatmentgruppe 1b (ausführliches Feedback – nicht genutzt) und Treatmentgruppe 2 (dichotome Rückmeldung): p = .007,
- sowie zwischen der Treatmentgruppe und der Treatmentgruppe 3, die Texte liest und somit als Kontrollgruppe fungiert (p = .066).

Die These, dass Schüler der Treatmentgruppe 1a, die ausführliche Rückmeldungen in den formativen Leistungstests erhalten und dieses auch genutzt haben, im Posttest am Ende der Unterrichtseinheit bei der Beantwortung von Fragen zum konzeptuellen Wissen signifikant besser abschneiden, als Schüler, die dieses Feedback nicht zur Verfügung haben, konnte also ebenfalls nicht bestätigt werden.

Die einzelnen Ergebnisse zeigen, dass die Schüler der Treatmentgruppe 1b, die das bereitstehende ausführliche Feedback nicht nutzte, am schlechtesten von Allen abschnitten, wenn es darum ging, konzeptuelle Lerninhalte zu verinnerlichen. Eine Erklärung hierfür könnte sein, dass diese Schüler nicht besonders gewissenhaft arbeiten und deshalb zu den eher leistungsschwachen Lernern zählen. Möchte ein Schüler seine Leistungen verbessern, wird er versuchen, die angebotene Hilfestel-

lung heranzuziehen, um dieses Ziel zu erreichen, andernfalls zeugt dies von einer geringen intrinsischen Motivation und lässt eine schlechtere Arbeitshaltung vermuten.

Die gewissenhafteren Schüler der Gruppe 1a, die das ausführliche Feedback heranziehen, um sich zu verbessern, schneiden zwar viel besser ab, als die Treatmentgruppe 1b und immerhin ein wenig besser, als die Kontrollgruppe (Treatment 3), die Texte lesen, aber dennoch schlechter, als die Gruppe (2), die lediglich erfährt, ob ihre Antwort richtig war, oder nicht. Dies ist zunächst verwunderlich, denn eigentlich liegt die Vermutung nahe, dass die häufigere Erläuterung der Zusammenhänge innerhalb der behandelten Konzepte auch gleichzeitig zu einer besseren Verinnerlichung und Durchdringung der damit zusammenhängenden Lerninhalte zur Folge hätte.

Versucht man für diese Ergebnisse Gründe dafür zu finden, könnte man folgende Punkte anführen:

- Die Treatmentgruppe 2 ist nicht durch zu viel Text von den eigentlichen Fragen und deren Antworten abgelenkt und kann sich besser auf diese konzentrieren. Später im Posttest, tun sich die Schüler dieser Gruppe vielleicht leichter, sich an bestimmte Eckdaten (Stichworte, Formulierungen) in den richtigen Antworten zu erinnern und dadurch die richtige Wahl zu treffen. Die Konzentration auf das Wesentliche hilft also, gemeinsam mit dem Übungseffekt bei der Auswahl der richtigen Antwortmöglichkeit.
- Vielleicht war ein ausführliches Feedback in manchen Fällen gar nicht nötig, das heißt die Schüler der zweiten Treatmentgruppe wussten eventuell bei manchen Antworten sofort, wo und warum sie falsch lagen, schon dadurch, dass der Fehler als solcher entlarvt und angezeigt wurde.
- Die dritte Treatmentgruppe wollte sich nicht mit dem Lesen der Texte aufhalten, sondern einfach fertig werden und sich mit anderen Dingen beschäftigen.
- Eine spätere Rückmeldung wäre eventuell effektiver, wenn es um komplexere Inhalte geht (siehe Ergebnisse der Feedbackinterventionstheorie nach Kluger & DeNisi (1996, 1998): Feedback zum Aufgabenlösungsprozess bringt die Schüler dazu, ihre eigenen Lösungswege kritisch zu hinterfragen und zu überprüfen, ihre Fehler zu reflektieren und über neue Lösungsansätze nachzudenken. Diese Art von Rückmeldung kann zu einem tieferen Verständnis der Lerninhalte und Konzepte führen und ist effektiver, wenn sie zeitverzögert erfolgt.)

Der Umstand, dass die Kontrollgruppe, die keine formativen Leistungstests bearbeitet hat, insgesamt schlechter abgeschnitten hat, als die Treatmentgruppen 2 und 1a, lässt vermuten, dass die Bearbeitung der Tests bzw. das Vorhandensein von Rückmeldungen an sich effektiver hinsichtlich des Wissenszuwachses war, als das reine Lesen der Texte. Und das, obwohl die Schüler dieser Gruppe, die meisten Informationen zu den Lerninhalten zur Verfügung hatten.

Dies könnte mehrere Gründe haben:

- Dadurch, dass die Testfragen in den formativen Leistungstests teilweise mit denen im Posttest übereingestimmt haben, konnten die Treatmentgruppen 1 und 2 die Antworten und das Testformat häufiger üben.
- Dadurch, dass die Schüler keinerlei Rückmeldung bei ihrer Auseinandersetzung mit dem Lerninhalt hatten und auch keine inhaltlichen Fragen beantworten mussten, waren sie nicht dazu angehalten, das Gelesene nochmals zu reflektieren.
- Beim Lesen von „bekannten“ Texten ist man schnell dazu geneigt, die Inhalte zu überfliegen und zu glauben, das wisse man ja schon alles. Gerade die wichtigen Details, die man eben nicht so gut verstanden hat, werden dadurch überlesen und vernachlässigt.
- Dadurch, dass die Lerner „nur“ Texte lesen durften waren sie weniger motiviert. Die Beobachtungen zeigten, dass viele Schüler der Gruppe 3, die Texte sehr schnell (innerhalb weniger Minuten) überflogen und diese dann weggelegten.

*Im **Behaltenstest** ergeben sich beim Konzeptwissen signifikante Unterschiede zwischen:*

- Treatment 1a und 3 (p = 0.048),
- Treatment 1a und 1b (p = 0.022),
- Treatment 2 und 3 (p = 0.041),
- Treatment 2 und 1b (p = 0.019).

Keine Unterschiede hingegen gab es beim Lernzuwachs des Treatments 2 und des Treatments 1a (p = 0.937), sowie zwischen den Schülern der Treatmentgruppe 1b und den Lernern, die sich mit den Texten auseinandersetzten (Treatment 3) (p = 0.482).

Beim Erinnern und eigenständigen Formulieren konzeptuellen Wissens im Behaltenstest schneidet also zum ersten Mal die Treatmentgruppe 1a am besten ab. Vielleicht hilft es bei der Erklärung von Zusammenhängen und Konzepten, vor allem wenn es darum geht, selbst Antworten zu formulieren und nicht nur die richtige vorformulierte Antwort zu finden, wenn man mehrmals ausführliche Rückmeldungen zu diesen Inhalten erhalten hat und deren Formulierungen wenigstens teilweise verinnerlichen konnte, um diese dann wiederzugeben. Das heißt, wenn es darum geht, schwierige Sachverhalte und Abläufe zu erfassen, wiederzugeben und länger zu erinnern, hilft eine häufige und detaillierte Wiederholung der kompletten Zusammenhänge.
Bei der Durchdringung konzeptuellen Wissens, kommt es ganz anders als beim Erlernen von Fakten, sehr wohl darauf an genauer hinzusehen, warum eine Aufgabe so beantwortet wurde, wie es der Schüler getan hat. Hier kann es sein, dass die falsche Antwort deshalb zustande kam, weil einzelne Zusammenhänge innerhalb eines Wissenskonstrukts nicht verstanden wurden und diese Denkfehler innerhalb der Schülerantwort gilt es offenzulegen.

Dennoch schneidet auch hier das Treatment 2 kaum schlechter ab (mögliche Gründe hierfür, siehe weiter oben), was zeigt, dass der Mehraufwand bei der Konzeption der formativen Leistungstests, der durch die Formulierung ausführlicher Rückmeldungen entsteht, wenig Rechtfertigung findet.

Für den Unterricht wäre es demnach empfehlenswert, das Feedback der Treatmentgruppe 2 ein wenig zu erweitern: Man könnte den Schülern neben der Information, wo sie richtig / falsch lagen, anzeigen lassen können, welche Teilantworten richtig gewesen wären.
Diese Kombination, aus reinem dichotomen Feedback und der Gewissheit, wie die richtige Antwort formuliert ist, hat eventuell den größten Effekt auf den Wissenszuwachs. Zusätzlich dazu scheint es sinnvoll, dass die Lehrkraft durch die Korrektur der formativen Tests, die Fragen isoliert, bei denen die Schüler die meisten Probleme haben, um auf diese in der folgenden Stunde noch einmal genau einzugehen. Eine zusätzliche Rückmeldung der Lehrkraft an die Klasse, lädt auf Schülerseite dazu ein, im Unterrichtsgespräch noch einmal genauer nachzufragen.
Der Umstand, dass die Schüler in der Regel nur daran interessiert waren, was falsch oder richtig war und welche Note dies bedeuten würde, zeigte aber auch, dass den Schülern letztendlich die Einzelleistung an sich wichtiger war, als ihr individueller Lernfortschritt und das Überdenken der Arbeitshaltung bei der Erarbeitung neuen Wissens. Es gab auch keine kritische Hinterfragung auf Schülerseite: „Warum habe ich das nicht verstanden und wie kann ich dies ändern? Was muss ich

tun, um besser zu werden?“ Nicht einmal: „Super – hier steht schon die richtige Antwort, lese ich doch mal nach, wie es richtig gewesen wäre.“

Zusammenfassend lässt sich die Vermutung, dass ausführliches Feedback (auch in Bezug auf konzeptuelles Wissen) effektiver als eine einfache dichotome Rückmeldung ist, nicht generell bestätigen. Auch zeigt sich, dass individuelle Unterschiede der Schüler in Arbeitshaltung, Lernverhalten und Gewissenhaftigkeit (Treatment 1a versus 1b) jeweils eine andere Art von Feedback benötigen. Diejenigen Schüler, die das bereitstehende Feedback auch nutzten, schnitten signifikant besser ab, als diejenigen, die dies nicht taten. Dies ist ein zu erwartendes aber auch wichtiges Ergebnis, führt es doch zu der Erkenntnis, dass man sich nicht nur Gedanken über die Art des Feedbacks machen sollte, sondern als Lehrkraft auch darauf achten muss, ob Schüler die Rückmeldungen überhaupt erhalten, und diese nutzen und verarbeiten. Dieses Ergebnis bestätigt, dass Unterschiede in der Schülerpersönlichkeit bestimmen, inwieweit diese ihre Wissenslücken wahr- und ernstnehmen und diese zu schließen versuchen (Kluger & DeNisi 1996). Diejenigen Schüler, die keine klaren Lernziele verfolgen und nicht gewissenhaft arbeiten, gehören wohl der Gruppe an, die das bereitstehende Feedback nicht nutzen. Eine weitere Erklärung für diese unterschiedliche Herangehensweise bei der Feedbacknutzung, könnte sein, dass Lerner auf verschiedene Art und Weise offen und sensibilisiert für die Annahme von Feedback sind (King & Young 2002).

Weiterhin ist noch anzumerken, dass sich beim Erlernen von konzeptuellem Wissen ein signifikanter Einfluss des Pretest auf den Posttest zeigte. Dies war beim Behaltenstest nicht der Fall. Dafür zeigte sich beim Behaltenstest ein signifikanter Einfluss der Schulleistung. Beides belegt, dass das Vorwissen in irgendeiner Art und Weise einen Einfluss auf das Behalten haben muss.

Abschließend scheint es sinnvoll anzumerken, dass man bei einer Studie, die in den normalen Unterricht integriert ist, nicht verhindern kann, dass Effekte durch den Unterrichtsalltag verwischt werden: So wiederholen Lehrer zu Beginn jeder Stunde bei der Abfrage die Inhalte der vorhergehenden Unterrichtsstunde jedes Mal mit der gesamten Klasse. Dadurch erhält ja schon jeder Schüler, egal welchem Treatment er angehört immer wieder Informationen über die Lerninhalte und wiederholt diese ständig. Den Unterschied, den vereinzelte Tests da beim Wissenszuwachs hervorrufen können, ist wahrscheinlich geringer, als wenn es diese gemeinsamen Wiederholungen nicht

geben würde. Aber diese Bedingungen liegen ja auch vor, wenn man über den Einsatz von formativen Leistungstests im Unterricht nachdenkt.

6.3.3.3 Fehlvorstellungen zur „Evolutionären Anpassung“

Im Kapitel 6.2.6 findet man alle Fragen, die im Pre- und Posttest zum Konzept der evolutionären Anpassung gestellt wurden. Da man in Moodle die Tests sogar auf Teilantwortebene auswerten kann, lässt sich genau nachvollziehen, inwieweit die Lerninhalte und deren Zusammenhänge verstanden wurden und wo noch Wiederholungsbedarf besteht. Auch den Lernfortschritt einzelner Schüler kann man mit einem gewissen Aufwand verfolgen.

Hier soll aber zunächst auf die Ergebnisse hinsichtlich der gegebenen Antworten innerhalb der Teilkonzepte der evolutionären Anpassung eingegangen werden:

Tabelle 44: Schülerantworten zu den Teilkonzepten der "Evolution" vor und nach der UE

Teilkonzept	Weniger Falschantworten	Mehr richtige Antworten
Genetische Variabilität	**-6.2%**	**13%**
Zufällige Genmutation	**-14%**	**31%**
Natürliche Selektion	**-3%**	**13%**
Dauer von Anpassungsvorgängen	**-8%**	**15%**
Evolution endet bei optimaler Anpassung	**-8%**	**20%**

Die Ergebnisse zeigen, dass die größte Änderung im Schülerdenken beim Teilkonzept der „Zielgerichtetheit von Anpassungsvorgängen“ (zufällige Genmutation) von Statten ging. Hier wurden im Posttest 14 % weniger Falschantworten angekreuzt und 31 % mehr der richtigen Antworten als solche erkannt:

Für die Schüler war eine Unterscheidung zwischen individueller und evolutionärer Anpassung in diesem Punkt am leichtesten nachzuvollziehen. Sie haben verstanden, dass sich Lebewesen nicht aus einer Notwendigkeit heraus anpassen, oder weil ein Wunsch in diese Richtung besteht, sondern dass es sich bei der evolutionären Anpassung um Vorgänge handelt, die durch eine zufällige Mutation im Genmaterial angestoßen wird. Auch konnte in diesem Zusammenhang vielen Lernern klar gemacht werden, dass man antrainierte Eigenschaften, wie schnelleres Laufen, oder größere Muskelmasse nicht in den Genen speichern und an seine Nachkommen vererben kann.

Dieser größte Anteil an Veränderungen im Schülerwissen, ist wohl auf verschiedene Faktoren zurückzuführen: Zum einen kann man als Lehrkraft in diesem Bereich sehr viele anschauliche Beispiele heranziehen, um den Schülern diesen Sachverhalt näherzubringen und zu verdeutlichen: Zum Beispiel kommt das Baby eines Bodybuilders nicht mit einer anderen Muskelmasse auf die Welt, als andere Babys. Zum anderen lassen sich durch die Gegenüberstellung von individuellen und evolutionären Anpassungsbeispielen, die Unterschiede sehr gut isolieren und darstellen: Wenn einem Menschen kalt ist, zieht er einfach eine dickere Jacke an, er passt sich also sofort und zielgerichtet den veränderten Außentemperaturen an. Eine Giraffe hingegen kann ihren Hals nicht sofort ausfahren und verlängern, wenn sie die Blätter eines zu hohen Baumes nicht auf Anhieb erreichen kann.

Beim besseren Verständnis der Teilkonzepte „genetische Variabilität“ und „natürliche Selektion“ gab es hingegen größere Probleme.
Das Konzept der natürlichen Selektion (nur 3 % weniger falsche und 13 % mehr richtige Antworten) ist für die Schüler wahrscheinlich schwieriger nachzuvollziehen, weil sich der Vorgang erstens auf einer etwas abstrakteren Gedankenebene befindet und zweitens viele Einzelkomponenten beinhaltet, die man sowohl einzeln als auch im Zusammenspiel verstehen muss, weil sie ineinandergreifen:

- Eine Population besteht aus vielen einzelnen Lebewesen, die bestimmten äußeren Bedingungen ausgesetzt sind (Klima, Nahrungsangebot, etc.).
- Es gibt Merkmale, die im Erbmaterial verankert sind, oder durch eine zufällige Genmutation dort zustande kommen (zum Beispiel: längerer Giraffenhals).
- Sind diese Merkmale vorteilhaft für die Lebewesen einer Population (genau in diesem Lebensraum), überleben eben die Lebewesen, die dieses Merkmal besitzen und vererben es dann an die nachfolgenden Generationen weiter.
- Das besser angepasste Individuum setzt sich durch.

Auch die genetische Variabilität mit nur 6.2 % weniger Falschantworten und ebenfalls 13 % mehr richtigen Antworten im Posttest, birgt ein paar Gedankengänge, die für die Lerner wahrscheinlich schwer nachzuvollziehen sind: Natürlich kann man ihnen die Artenvielfalt relativ schnell näherbringen, denn die können sie selbst jederzeit sehen, wenn sie sich in ihrem direkten Umfeld umschauen: Da gibt es Menschen mit verschiedenen Haut- und Haarfarben, unterschiedlich große

Spinnen, Schmetterlinge in verschieden Farben, eine Vielzahl unterschiedlicher Lebewesen überall auf der Welt und in jedem erdenklichen Lebensraum. Schwieriger hingegen ist es für die Schüler, dass nie alle Lebewesen einer Population gleich aussehen werden. Warum sind nicht alle Lebewesen einer Population, wenn sie sich an ihren Lebensraum optimal angepasst haben, nicht absolut gleich? Hier gilt es, noch mehr herauszustellen, dass es Merkmale gibt, die für das Überleben der Art wichtig sind und eben auch Merkmale (wie die Augenfarbe des Menschen, oder dessen genauer Körpergröße), die dies nicht sind.

Die Veränderungen im Denken zu den Teilaspekten „Dauer von Anpassungsvorgängen“ mit 15 % und „Evolution endet bei optimaler Anpassung“ mit 20 % mehr korrekten Antworten und jeweils 8 % weniger Falschantworten liegt in einem eher mittleren Bereich, was verwunderlich ist, denn gerade diese beiden Aspekte sind nicht sehr komplex, sondern müssen lediglich so hingenommen und nicht weiter hinterfragt werden:

Das Anpassungsvorgänge mehrere Generationen dauern, ist eigentlich logisch und leicht nachzuvollziehen, auch dass Evolution an sich nie endet, ist offensichtlich, denn die Welt ändert sich dauernd und macht eine neuerliche Anpassung an die Veränderten Gegebenheiten unabdingbar. Vielleicht können hier vor allem die niedrige Jahrgangsstufe und das damit einhergehende Alter der Schüler als Erklärungsversuch dienen:

Es bedarf eventuell einer größeren Vorstellungskraft, derartig langsame und andauernde Vorgänge als solche zu erkennen. Die Vorstellung des Zeitraumes von „mehreren Generationen“ ist genauso schwierig, wie Veränderungen in der Umwelt wahrzunehmen, die einen als Kind unmittelbar betreffen. Veränderungen von Lebensräumen, Klima, etc. kann man vielleicht erst dann nachvollziehen, wenn man selbst die Chance hatte diese Vorgänge über längere Zeit zu beobachten.

Insgesamt sind natürlich Verbesserungen im Verstehen des Konzepts der evolutionären Anpassung zu sehen. Aussagen darüber, ob dies an den gewählten Unterrichtsinhalten und -materialien, den formativen Leistungstests, den Rückmeldungen oder dem Unterricht an sich lag, kann man an dieser Stelle nur bedingt machen. Allerdings sieht man bei den Schülern, die die formativen Leistungstests geschrieben und ein ausführliches Feedback erhalten haben, größere Fortschritte beim Abbau von Fehlvorstellungen innerhalb des Konzeptes der evolutionären Anpassung, wenn es darum geht, den Sachverhalt und die Zusammenhänge dieses Konzepts mit eigenen Worten zu schildern.

Im weiteren Verlauf der Datenauswertung wäre es noch interessant zu überprüfen, inwieweit sich die einzelnen Teilaspekte zu den Fehlvorstellungen innerhalb der einzelnen Gruppen geändert haben.

Die Frage: „Lassen sich Fehlvorstellungen bei Schülern niedriger Jahrgangsstufen in Hinblick auf das Konzept der evolutionären Anpassung mit Hilfe von formativen Leistungstests berichtigen?“ muss man differenziert betrachten: Auf einer sehr abstrakten Ebene lassen sich sicherlich Erfolge bei der Überwindung von Fehleinschätzungen zum Konzept der evolutionären Anpassung bei so jungen Schülern verzeichnen. So lange man nicht in die Tiefe geht, und die Schüler einzelne Komponenten nicht zu sehr hinterfragen (Vererbungslehre, Aufbau von genetischem Material, Genmutationen, etc.), sondern diese als gegeben annehmen und akzeptieren, kann man auf einer „oberflächlichen“ Wissensebene einzelne Teilaspekte in das richtige Licht rücken. Ein tieferes Verständnis der einzelnen Vorgänge und eine vollständige Durchdringung, was evolutionäre Anpassung und alle damit einhergehenden biologischen Aspekte bedeuten, kann erst später in höheren Klassen erfolgen. Außerdem ist es notwendig, derart komplexe Zusammenhänge immer wieder aufzugreifen, zu wiederholen und an gegebener Stelle aus allen Blickwinkeln zu beleuchten.

6.3.4 Motivationale Auswirkungen

Wie bereits in der Auswertung der Evaluationsergebnisse beschrieben, hatten folgende Persönlichkeitsmerkmale keinen Einfluss auf das Treatment:

- Positives Selbstkonzept
- negatives Selbstkonzept
- intrinsische Motivation (Interesse)
- intrinsische Motivation (Kompetenzorientiert)
- Instruktionsbezogene Angst (Trait)
- Instruktionsbezogenes Interesse (Trait)
- Instruktionsbezogenes Wohlbefinden (Trait)

Bei der Auswertung der Daten zeigte sich, dass die Veränderung in der Motivation der Schüler vom ersten Messzeitpunkt bis zum Posttest nicht signifikant war:

Dies liegt daran, dass es sich bei den oben genannten Persönlichkeitsmerkmalen um ein stabiles, überdauerndes Konstrukt (trait) handelt. Das Selbstkonzept der Schüler wird sich zum Beispiel nicht innerhalb von 9 Wochen grundlegend ändern.
Weiterhin gab es keine signifikanten Unterschiede zwischen den Treatments bezüglich der situativen Emotionen Langeweile, Wohlbefinden und Interesse. Dies war zu erwarten, denn der Unterricht war in allen drei Treatmentgruppen identisch.

Interessant für diese Arbeit ist bestimmt, dass die Schüler der Treatmentgruppe 1a, die von sich selbst sagten, dass sie das bereitstehende ausführliche Feedback auch genutzt haben, eine hohe Korrelation mit der Gewissenhaftigkeit aus dem Big Five aufwiesen ($r = 0.295$, $p = 0.008$). Dies ist nicht weiter verwunderlich, denn vor allem die gewissenhaften Schüler bringen genügend intrinsische Motivation auf, um sich mit weiterführenden Informationen zu beschäftigen.

Ansonsten gab es in allen Treatmentgruppen keine Unterschiede den Big Five, sowie das (die) Instruktionsbezogene(s)_Interesse / Angst und Wohlbefinden betreffend. Dies liegt daran, dass alle Gruppen vergleichbare Bedingungen vorfanden, wie die zufällige Zuordnung zu den Treatments, den gleichen Lehrer, den gleichen Raum, etc.

Sieht man sich die Kurzskalen genauer an, stellt man fest, dass es auch hier keine signifikanten Unterschiede zwischen den Treatments bezüglich der situativen Emotionen Langeweile, Wohlbefinden und Interesse während der gesamten Unterrichtseinheit gab. Die Kurzskalen wurden an 3 Messzeitpunkten während der Unterrichtseinheit von den Schülern ausgefüllt und bei der Lehrkraft abgegeben. Für die Analyse wurden alle drei Messzeitpunkte gemittelt und zu einem Wert zusammengefasst. Dies bedeutet, dass die Unterrichtseinheit von den Schülern nicht als unterschiedlich empfunden wurde und belegt die Qualität des Untersuchungsdesigns, da alle Schüler denselben Unterricht bekamen, und lediglich das Feedback variiert wurde.

Korrelationen

Untersucht man die Korrelationen zwischen den *situativen Emotionen* „Interesse, Wohlbefinden und Langweile“, sowie der „Bewertung der Unterrichtssequenz“ mit dem *Wissenszuwachs* genauer zeigt sich, dass das Wohlbefinden signifikant positiv mit den nachfolgenden Wissenstests korrelier-

te. Dies bedeutet, dass Schüler, die sich im Unterricht wohler fühlten und eine höhere Leistung erzielten. Langeweile hingegen korrelierte im Behaltenstest negativ mit den Leistungen. Dies belegt, dass Langeweile im Unterricht zu einer schlechteren Leistung führt.

7 Ausblick

In den folgenden Abschnitten soll ein kurzer Überblick gegeben werden, auf welche Dinge bei den nächsten Forschungsprojekten besonderes Augenmerk gelegt werden könnte. Auch sollen ein paar Tipps für die Schulpraxis beim Einsatz von neuen Medien generell und in Hinsicht auf formative Leistungsmessung und den damit einhergehenden Rückmeldungen gegeben werden.

7.1.1 Nächste Schritte in der Forschung

Generell scheint es besser zu sein, häufiger kleinere formative Leistungstests in den Unterricht zu integrieren. Der Übungseffekt ist höher und Defizite auf Schülerseite können zeitnah entdeckt und abgebaut werden. Diese Gegebenheit sollte außerdem mit dem Einsatz elektronischer Tests kombiniert werden: Kurze, vielleicht nur aus 5 - 8 Fragen bestehende Tests sollten am Ende der Erarbeitung jeden Teilaspekts des Lerninhalts stehen, also vielleicht sogar am Ende jeder dritten Stunde. Durch die Kürze der Prüfungen, ist der Zeitaufwand der Durchführung nicht besonders groß – wahrscheinlich können sie in den letzten 10 Minuten der jeweiligen Stunde abgehalten werden. Dieses Vorgehen setzt allerdings mehrere Dinge voraus:
Es ist genügend elektronisches Equipment vorhanden (die gesamte Klasse sollte zeitgleich an den Tests arbeiten können).
Und die Schüler und Lehrer sind es gewohnt mit dem vorhandenen elektronischen Equipment effektiv zu arbeiten. Alle organisatorischen Schritte, wie das Einloggen und Durchführen der Tests müssen auf Schülerseite verinnerlicht sein. Die Lehrkraft sollte ausreichend mit der Technik vertraut und versiert genug sein, technische Probleme möglichst effizient und schnell zu lösen. Außerdem ist es von Vorteil, wenn die Lehrkraft einen möglichst großen Pool an Aufgaben zu allen Unterrichtsthemen bereit hat, um möglichst schnell und effektiv passende elektronische Tests erstellen zu können.

In diesem Zusammenhang wäre es interessant mehrere Dinge genauer zu untersuchen:

- Wie gehen die Lehrer mit diesem Szenario um? Ist es möglich, elektronische Leistungstests so weit zu integrieren, dass sie als normale Unterrichtsroutine angesehen werden (ähnlich, wie Abfragen oder Hefteinträge, etc.)?

- Gibt es Lehrkräfte, die trotz anfänglicher Zweifel bereit sind, sich mit der Technik auseinanderzusetzen und diese zu nutzen?
- Wie sollten die Rückmeldungen bei solchen Kurztests gestaltet werden? Sollte man zusätzlich zum dichotomen Feedback „Korrekt / Falsch“, die richtige Antwort angeben?
- Kann man aus den Kurztestes heraus individuelle Übungsaufgaben / Hausaufgaben, bzw. gleich neue Kurztests für jeden einzelnen Schüler generieren lassen? (Adaptivität der Tests)
- Inwieweit nutzt die jeweilige Lehrkraft die Rückmeldung zum Wissensstand der einzelnen Lerner, die sie durch die Korrektur der Kurztests bekommt, um ihren nachfolgenden Unterricht zu adaptieren?

Auf die beiden letzten Punkte der Aufzählung möchte ich noch einmal gesondert eingehen:
Ich denke, dass der Einsatz eines adaptiven Computersystems die größten Effekte auf die Schülerleistung haben würde, denn nur damit können einzelne individuelle Denkprobleme isoliert und abgebaut werden. Hat ein Schüler zum Beispiel bei der Lösung von Gleichungen immer wieder Schwierigkeiten damit, die Vorzeichenregeln richtig umzusetzen, macht es wenig Sinn, ihn mit den bereits verinnerlichten Klammerregeln und Äquivalenzumformungen zu beschäftigen. Dann wäre es angeraten, diese Problematik zu isolieren und zu eliminieren und zwar mit Hilfe einzelner kurzer Übungsaufgaben, die sich gezielt und ausschließlich mit diesem Problem beschäftigen. Isoliert man diese einzelnen Gedankengänge nicht, wird der Schüler immer wieder von den restlichen Strukturen der Aufgaben, die er dann auch immer mit berücksichtigen muss, vom eigentlichen Problem abgelenkt. Würde man ein adaptives System entwickeln und einsetzen, gäbe es sehr viele Möglichkeiten, die Schüler viel individueller zu unterstützen:
Durch gezielte Haus- und Übungsaufgaben und gut durchdachte spezifische Rückmeldungen. Im besten Fall, könnte sogar der nächste Kurztest für diesen Schüler automatisch generiert werden. Auch eine Darstellung des eigenen Lernverlaufs und –erfolgs auf Schülerseite wäre möglich und das auf einzelne Teilaspekte der Lerninhalte bezogen…
Es wäre sehr spannend, zu untersuchen, was ein derartiges adaptives und adaptierbares System bei der formativen Leistungsmessung bewirken könnte.

Im zweiten Punkt wird die Adaptierbarkeit des Unterrichts und somit auch eine, in der Studie nicht untersuchte Form des Feedbacks angesprochen – des Feedbacks an die Lehrkraft. Ich denke, dass in dieser Art der Rückmeldung das zweite große Potential der elektronischen Leistungstests liegt:

Der Lehrer muss seine formativen Tests nicht mehr mühevoll und zeitraubend handschriftlich korrigieren, sondern kann sofort und auf Knopfdruck sehen, wo die Schüler die meisten Probleme hatten und zeitnah (noch in der gleichen oder der darauffolgenden Unterrichtsstunde) noch einmal auf diese Probleme gezielt eingehen. Dies bringt noch eine weitere Feedbackart ins Spiel: Die verbale Rückmeldung an einzelne Schüler, einzelne Schülergruppen oder die gesamte Klasse während der Unterrichtsgesprächs. Insgesamt wäre es sinnvoll zu untersuchen, wie effektiv diese Feedbackarten hinsichtlich des Lernerfolgs und der Schülermotivation ist.

Im Bereich der Computerunterstützung bei der formativen Leistungsmessung lassen sich noch sehr viele Dinge untersuchen:

- Welche Endgeräte eignen sich am besten? Besonders interessant: Wie kann ich vorhandene Kapazitäten (Schülerhandys) sinnvoll in den Unterricht integrieren und die Möglichkeiten von Smartphones voll ausnutzen. Es ist sehr schade, dass Schülerhandys generell an der Schule verboten werden und nicht genutzt werden. Eine nicht repräsentative Umfrage von Lehrkräften hat ergeben, dass in den Klassen 9 aufwärts, so gut wie jeder Schüler ein Smartphone besitzt, warum dieses also nicht für Recherchen, das Aufzeichnen von englischen Dialogen, die Erstellung und Präsentation von Vorträgen, usw. heranziehen?
- Welche Aufgabenformate bieten sich besonders bei der Vermittlung und der Abfrage von konzeptuellem Wissen an? Lassen sich Begriffskonstrukte und komplexe Zusammenhänge zum Beispiel durch die Verbindung von Drag & Drop und Conceptmapping Aufgaben besonders gut veranschaulichen, üben und verinnerlichen?
- Inwieweit lassen sich noch weiter entwickelte Technologien, wie die der virtuellen und erweiterten Realität, die bisher vor allem beim Erlernen prozeduralen Wissens eingesetzt wurden, auch bei der Erarbeitung von kognitiven Lerninhalten einsetzen? Gibt es eine Möglichkeit, diese Technologien für die formative Leistungsmessung heranzuziehen? Für die Schulpraxis ist diese Frage allerdings noch sehr weit weg, denn im Moment gibt es viele Schulen, die noch nicht einmal über einen normalen Rechnerraum oder ein WLAN verfügen. Dennoch ist diese Frage aus wissenschaftlicher Sicht interessant.

Generell sollten bei weiteren Untersuchungen auch andere Fächer mit einbezogen werden, um genauer zu überprüfen, ob unterschiedliche Fächer, die auch verschiedene Wissensstrukturen und –ebenen mit sich bringen, von erfolgreich erprobten Feedbackvarianten profitieren.

Weiter wäre es sinnvoll zu untersuchen, inwieweit sich das eigene Lernverhalten und die bisher praktizierten Lernstrategien von Schülern mit Hilfe von geeigneten Rückmeldungen und formativen Tests modifizieren lassen. Dies setzt allerdings einen längeren Untersuchungszeitraum voraus, denn die Implementierung von Routinen dauert länger als die kurze Evaluationszeit der Studie.

7.1.2 Empfehlungen für die Schulpraxis

Während des Projekts konnte immer wieder beobachtet werden, wie motiviert und begeistert die Schüler mit den Tablets umgegangen sind. Auch in den iPad Klassen war dies der Fall, so dass man davon ausgehen kann, dass es sich nicht nur um eine anfängliche Begeisterung handelt, sondern dass die Motivation mit elektronischen Geräten umzugehen, auch länger anhält. Durch Hospitationen in unterschiedlichen Fächern zeigte sich das breite Einsatzspektrum dieser Geräte: Nicht nur zum Surfen oder Recherchieren im Internet wurden die Tablets herangezogen, sondern auch um in Englisch Dialoge aufzunehmen, um in Musik Vorträge zu halten, bei denen nicht nur Folien gezeigt wurden, sondern auch an geeigneten Stellen Musikbeispiele oder kurze Filme eingespielt wurden, in Physik, um zu zeigen, wie sich bestimmte Bedingungen innerhalb eines Versuchsaufbaus verändern, wenn man einzelne Parameter umgestaltet, usw. Die Liste der Möglichkeiten lässt sich wohl endlos fortsetzen. Leider gibt es nur sehr wenige Schulen, die Tablet PCs im Unterricht einsetzen.

Dies liegt an mehreren Dingen:

An der technischen Infrastruktur der Schulen, die ungefähr ein Jahrzehnt hinterherhinkt (viele Schulen haben kein WLAN, kein funktionierendes Schulnetz oder sogar nicht einmal einen Computerraum). Ist ein Computerraum vorhanden, wird er aber nicht oder nur rudimentär von Informatiklehrkräften, oder gerade noch Mathematiklehrern genutzt. Oft ist es eine zu große technische Herausforderung für andere Lehrkräfte, sich mit dem System im Computerraum vertraut zu machen und viele Lehrer finden auch, dass sich ein Wechsel in einen anderen Raum, nur um kurz einen Arbeitsauftrag zu erledigen, nicht lohnt und unnötig Unruhe in die Klasse bringt. Könnte man im Klassenzimmer bleiben und alle Kinder wären mit den Tablets vertraut, könnte man diese jederzeit schnell und unkompliziert in den Unterricht integrieren.

An der Frage der Finanzierung. In den iPad Klassen haben die Eltern der Schüler die Geräte selbst finanziert, d.h. nur die Kinder dürfen in die iPad Klasse, deren Eltern sich das auch leisten können. Auch hier zeigt sich, wie an vielen Stellen im deutschen Bildungssystem, dass man noch lange nicht von Gleichberechtigung in Bildungsfragen sprechen kann.

An der Scheu vieler Lehrer, sich mit technischen Neuerungen auseinanderzusetzen. Oft ist die jüngere Generation, die mit derartigen elektronischen Geräte aufgewachsen ist bereit, sich diesen Neuerungen zu öffnen, immer vorausgesetzt, man hat auch sinnvolle Einsatzmöglichkeiten, doch ältere Lehrkräfte möchten sich nicht mehr so gerne in diese Materie einarbeiten. Auch der erst einmal höhere Arbeitsaufwand bei der Erstellung neuer Lehrmaterialien ist für viele abschreckend. Viele sehen nicht, dass man sich dann im Nachhinein wieder viel Arbeit und Ressourcen sparen kann (kleine Änderungen in den Materialien können sehr schnell vorgenommen werden, es sind keine Kopien mehr nötig, man kann den Schülern alles elektronisch zur Verfügung stellen und diese können gleich damit arbeiten, Tests lassen sich per Knopfdruck korrigieren, etc.).

Außerdem haben Viele noch nicht erkannt, wie viele Einsatzmöglichkeiten sich während der Unterrichts bieten würden. Hier wäre es gut, wenn schon während der Lehrerausbildung mehr Seminare im Medienpädagogischen Bereich angeboten würden. Durch die Einbindung von Filmen, Animationen und gutem Bildmaterialien lassen sich bestimmte Sachverhalte und Abläufe besser veranschaulichen und die Schülermotivation, sich mit den Lerninhalten auseinanderzusetzen enorm steigern.

Schulbuchverlage hängen aufgrund des zu geringen Absatzmarktes für digitale und vor allem interaktive Schulbücher noch viel zu sehr an der gedruckten Version ihrer Unterrichtsmaterialien. Man könnte den Kindern gesundheitlich gesehen einen großen Gefallen tun, wenn Sie nicht mehr für jedes Fach ein Buch mit sich herumschleppen müssten, sondern alle zusammen auf einem einzigen Gerät mit sich führen könnten. Auch müssen die Kinder in ihrem späteren Berufsleben immer häufiger mit technischem Equipment umgehen. Das Handling mit und Schreiben auf solchen Geräten sollte also früh geübt werden, um diese später ganz selbstverständlich in den Arbeits- und Lernprozess einbinden zu können.

Vieles spricht für den Einsatz mobiler elektronischer Geräte in den Unterricht, einige Lehrer müssten ihre Ängste in dieser Beziehung abbauen. Es wäre hilfreich, wenn sich in jeder Schule ein Administrator finden würde, der das technische Equipment mit betreut und vielleicht auch schulinterne Weiterbildungen durchführen könnte. Würde das gesamte Lehrerkollegium bei der Gestal-

tung von elektronischen Unterrichtsmaterialien und Testfragen zusammenarbeiten, könnte bald jeder auf einen großen Pool zurückgreifen und müsste nicht alles selbst erstellen. Oft gibt es auch schon viele Materialien im Internet, die man als Lehrkraft nutzen und für die Klasse anpassen kann (bei Moodle zum Beispiel das bayerische Realschulnetz).
Schön wäre auch, wenn mehr in diese Dinge investiert würde: Sowohl in finanzieller Hinsicht, als auch bei der Lehreraus- und -weiterbildung.

Hinsichtlich der durchgeführten Studie lässt sich festhalten, dass es durchaus Sinn macht, formative Leistungstests in elektronischer Form anzubieten. Dabei sollten die Tests möglichst kurz sein, einen hohen Übungscharakter aufweisen, aber auch Transferfragen nicht zu kurz kommen lassen. Nur elektronische Tests ermöglichen eine sofortige Korrektur und Rückmeldung an die Lehrkraft und erlauben dadurch eine sofortige Adaption des Unterrichts an die Bedürfnisse der Lerner und ein individuelles Feedback an die Schüler.
Der Umgang mit den Tablets war sehr unkompliziert, da auf Schülerseite durchgängig eine hohe Medienkompetenz vorhanden war.

8 Literaturverzeichnis

Achtenhagen, F. & Baumert, J. (1999). Alternative Evaluationsverfahren. In: Unterrichtswissenschaft, 27/2, 98-101.

Anderson, K. T., Zuiker, S. J., Taasoobshirazi, G., & Hickey, D. T. (2007). Classroom discourse as a tool to enhance formative assessment and practise in science. International Journal of Science Education, 29/14, 1721-1744.

Aufenanger, S. & Schlieszeit, J. (2013). Tablets im Unterricht nutzen, Computer + Unterricht, 89, S. 6-10

Ayala, C. C., Shavelson, R. J., Ruiz-Primo, M. A., Brandon, P. R., Yin, Y., Furtak, E. M., Young, D. B. & Tomita, M. K. (2008). From Formal Embedded Assessments to Reflective Lessons: The Development of Formative Assessment Studies. Applied Measurement in Education, 21/4, 315-334.

Baalmann W., Frerichs V., Weitzel H., Gropengießer H. und Kattmann U.(2004). Schülervorstellungen zu Prozessen der Anpassung – Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion. Zeitschrift für Didaktik der Naturwissenschaften; S. 7-28

Bairlein, F. (1996). Ökologie der Vögel: physiologische Ökologie, Populationsbiologie, Vogelgemeinschaften, Naturschutz. Jena: G. Fischer.

Bangert-Drowns, R. L., Kulik, C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. Review of Educational Research, 61, 213-238.

Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. Zeitschrift für Erziehungswissenschaft, 9/4, 469-520.

Baumgartner, P., Herber, S., Waba, S. (2011). Szenarien für den Einsatz von Netbooks, Computer + Unterricht 81, 44-45

Beatty, I. D., Gerace, W. J. (2009). Technology-enhanced formative assessment: A research-based pedagogy for teaching science with classroom response technology. Journal of Science Education and Technology, 18/2, 146-162.

Bell & Cowie (2001). The characteristics of formative assessment in science education, Science Education 85/5, 536-553

Bennett, R. E. (2011). Formative assessment: a critical review. Assessment in Education. 18/1, 5-25.

Bergmann, H.-H. (1987). Die Biologie des Vogels: eine exemplarische Einführung in Bau, Funktion und Lebensweise. Wiesbaden: Aula-Verlag.

Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., Wiesemes, R. & Nickmans, G. (2006). A learning integrated assessment system. Educational Research Review, 1/2006, 61-67.

Bishop, B. & Anderson, C. (1990). Student conceptions of natural selection and its role in evolution. Journal of Research in Science Teaching 27/5, 415-427

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education, 5/1, 7-74.

Black, P. & Wiliam D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21/1, 5-31.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company.

Bloom, B. S. (1974). An introduction to mastery learning theory. In J. H. Block (Ed.), Schools, society and mastery learning. New York: Holt, Rinehart & Winston.

Blum, W., Drüke-Noe, C. Hartung, R. & Köller, O. (Hrsg.) (2006). Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen. Berlin: Cornelsen.

Bishop B. und Anderson, C. W. (1986). Evolution by natural selection: A teaching modul; The institute for research on Teaching, Michigan State University 1986

Brandon, P. R., Young, D. B., Shavelson, R. J., Jones, R., Ayala, C. C., Ruiz-Primo, M. A., Yin, Y, Tomita, M. K., Furtak, E. M. (2008). Lessons Learned from the Process of Curriculum Developers' and Assessment Developers' Collaboration on the Development of Embedded Formative Assessments. Applied Measurement in Education, 21/4, 390-402.

Brinkley, E. S., Beatson, J, Wiemken, C. (2001). Vögel. Tessloff Verlag Nürnberg

Broadfoot, P. & Black, P. (2004). Redefining assessment? The first ten years of Assessment in Education. In: Assessment in Education, 11/1, 7-27.

Brown, B. A. (2008). Assessment and academic identity: Using embedded assessment as an instrument for academic socialization in science education. Teachers College Record, 110/10, 2116-2147.

Büchter, A. & Leuders, T. (2005). Mathematikaufgaben selbst entwickeln. Lernen fördern - Leistungen überprüfen. Berlin.

Bühner, M. (2011). Einführung in die Test- und Fragebogenkonstruktion. Pearson Deutschland GmbH. Addison-Wesley Verlag

Burnie, D. (2006). Vögel. Beobachten und experimentieren. Starnberg: Dorling Kindersley

Burnie, D., King, D. & Sturm, L. (2006). Vögel. Starnberg: Dorling Kindersley.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest and performance. Journal of Educational Psychology, 79, 474-482.

Butler, R. (1993). Effects of task- and ego-achievement goals on information seeking during task engagement. Journal of Personality and Social Psychology, 65, 18-31.

Butler, R. (2005). Ein zielorientiertes Modell für Feedbackverfahren im Unterricht [A Goal-Oriented Model of the Functions and Consequences of Evaluation and Information-Seeking] Implikationen für Lernmotivation und Schulstruktur. Unterrichtswissenschaft, 33/2, 122-142.

Bürgermeister, A. (2014). Leistungsbeurteilung im Mathematikunterricht: Bedingungen und Effekte von Beurteilungspraxis und Beurteilungsgenauigkeit, Waxmann Verlag GmbH, Münster

Bürgermeister, A., Klimczak, M., Klieme, E., Rakoczy, K., Blum, W., Leiss, D., Harks, B. & Besser, M. (2011). Leistungsbeurteilung im Mathematikunterricht – Eine Darstellung des Projekts „Nutzung und Auswirkungen der Kompetenzmessung in mathematischen Lehr-Lernprozessen“. In A. Füchter & K. Moegling (Hrsg.), Theorie und Praxis der Schulpädagogik. Band 16. Diagnostik und Förderung (S. 28-51). Immenhausen: Prolog-Verlag.

Chang, C. (2008). Enhancing self-perceived effects using Web-based portfolio assessment. Computers in Human Behavior, 24/4, 1753-1771.

Clark, I. (2012). Formative Assessment: Assessment Is for Self-Regulated Learning. Educational Psychology Review, 24/2, 205-249.

Codding, R. S., Skowron, J. & Pace, G. M. (2005). Back to basics: Training teachers to interpret curriculum-based data and create observable objectives. Behavioral Interventions, 20, 165–176.

Costa, P. T. Jr., & McRae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, Florida: Psychological Assessment Resources, Inc.

Crooks, T. J. (1988). The Impact of Classroom Evaluation Practices on Students. Review of Educational Research 58/4, 438-481.

Deutsche Gesellschaft für Evaluation (2008). Standards für Evaluation. Mainz.

Díaz-Morales, J. F., Delgado Prieto, P., Escribano Barreno, C., Collado Mateo, Ma José & Randler, C. (2012): Sleep beliefs and chronotype among adolescents: the effect of a sleep education program, Biological Rhythm Research, 43/4, 397-412

Ditton, H. (1998). Mehrebenenanalysen, Grundlagen und Anwendungen des Hierarchisch Linearen Modells, Juventa Verlag Weinheim und München.

Dochy, F. (2001). A new assessment era: different needs, new challenges. Research Dialogue in Learning and Instruction, 2, 11–20.

Donovan, W. (2008). An Electronic Response System and Conceptests in General Chemistry Courses. Journal of Computers in Mathematics and Science Teaching, 27/4, 369-389.

Dweck, C. S. (1986). Motivational processes affecting learning. American Psychologist, 41, 1040-1048.

Elphick, Jonathan (op. 2008): Atlas des Vogelzugs. Die Wanderungen der Vögel auf unserer Erde. Hauptverlag, Bern

Engel Clough, E. & Wood-Robinson, C. (1985). How secondary students interpret instances of biological adaptation. Journal of Biological Education, 19/2, 125-130.

Feldman, A., Capobianco, B. M. (2008). Teacher learning of technology enhanced formative assessment. Journal of Science Education and Technology, 17/1, 82-99.

Fraser, B. J., Walberg, H. J., Welch, W. W. & Hattie, J. A. (1987). Syntheses of educational productivity research. International Journal of Educational Research, 11, 145-252.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. School Psychology Review, 33/2, 188–192.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: a meta-analysis. Exceptional Children, 53, 199-208.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B. & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. Exceptional Children, 60, 518–537.

Furtak, E. M., Ruiz-Primo, M. A. (2008). Making Students' Thinking Explicit in Writing and Discussion: An analysis of formative assessment prompts. Science Education, 92/5, 798-824.

Gläser-Zikuda, M., Fuß, S., Laukenmann, M., Metz, K. & Randler, Ch. (2005). Promoting students' emotions and achievement - instructional design and evaluation of the ECOLE-approach. In A. Efklides & S. Volet (Eds.), Feelings and Emotions in the Learning Process. Special Issue: Learning and Instruction, 15/5, 481-495.

Gläser-Zikuda, M., Lindacher, T. & Fuß, S. (2006). Wirksamkeit eines Portfolios im Unterricht zur Förderung von Lernleistungen und Lernstrategien - eine quasi-experimentelle Studie. Empirische Pädagogik, 20/3, 229-244.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. Jr. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 37, 504–528.

Gropengießer, H. (2003). Lebenswelten, Denkwelten, Sprechwelten. Wie man Vorstellungen der Lerner verstehen kann. Beiträge zur Didaktischen Rekonstruktion. Bd. 4, Oldenburg: Didaktisches Zentrum

Gutenberg, U., Schwabe, S. (2013). Tablets für den Schulgebrauch, Computer + Unterricht 89, S. 35-39

Harlen, W. & Winter, J. (2004). The development of assessment for learning: Learning from the case of science and mathematics. Language Testing, 21/3, 390-408.

Harlen, W. (2007). Formative classroom assessment in science and mathematics. In: McMillan, James H. (2007) (Ed.). Formative classroom assessment: theory into practice (pp. 116-135). Teachers College Columbia University New York and London: Teachers College Press.

Harlen, W. (2011). On the relationship between assessment for formative and summative purposes, in: Assessment and learning (Gardner, J. Herausgeber), Sage, S. 87-102

Hattie , J. (2003). Formative and Summative Interpretations of Assessment Information http://www.education.auckland.ac.nz/webdav/site/education/shared/hattie/docs/formative-and-summative-assessment-%282003%29.pdf, Paper, University of Auckland

Hattie, J. & Timperley, H. (2007). The Power of Feedback. Review of Educational Research, 77/1, 81-112.

Helmke, A. (1992). Selbstvertrauen und schulische Leistung [Self-confidence and school achievement]. Göttingen, Germany: Hogrefe.

Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistung. In: F. E. Weinert (Hrsg.), Enzyklopädie der Psychologie D/I/3. Psychologie des Unterrichts und der Schule (S. 71-176). Göttingen: Hogrefe.

Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In: R. Arnold & C. Griese (Hrsg.), Schulmanagement und Schulentwicklung (S. 115-135). Hohengehren: Schneider-Verlag

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? Phi Delta Kappan, 89, 140-145.

Hickey, D. T., & Zuiker, S. J. (2012). Multilevel assessment for discourse, understanding, and achievement. Journal of the Learning Sciences, 21/4, 522-582.

Hickey, D. T., Taasoobshirazi, G., & Cross, D. (2012). Assessment as Learning: Enhancing Discourse, Understanding, and Achievement in Innovative Science Curricula. Journal of Research in Science Teaching, 49/10, 1240-1270.

Hochweber, J. (2010). Was erfassen Mathematiknoten? Korrelate von Mathematik-Zeugniszensuren auf Schüler- und Schulklassenebene in Primar- und Sekundarstufe, Waxmann Verlag GmbH

Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten [Test Anxiety Inventory TAI-G: an expanded and modified version with four components]. Zeitschrift für Pädagogische Psychologie, 5/2, 121-130.

Horstkemper, M. (2004). Diagnosekompetenz als Teil pädagogischer Professionalität. In: Neue Sammlung, 44/2, 201-214.

Hummel, M., Glück, R., Jürgens, C. Weisshaar, J., Randler, C. (2012) Interesse, Wohlbefinden und Langeweile im naturwissenschaftlichen Unterricht mit lebenden Organismen, Zeitschrift für Didaktik der Naturwissenschaften, S. 99 - 116

Hung, P., Lin, y., & Hwang, G. (2010). Formative assessment design for PDA integrated ecology observation. Educational Technology & Society, 13/3, 33-42.

Ingenkamp, K. (Hrsg.) (1995). Die Fragwürdigkeit der Zensurengebung. 9. Auflage, Weinheim: Beltz.

Jensen Murray, Settlage John, Odem Louis (1996). Investigating the Disney effect: Are students reluctant to apply natural selection principles to life forms with which they identify?. Annual Meeting of national association of research in science teaching (St. Louis, April 1996)

Johannsen, M. & Krüger, D. (2005). Schülervorstellungen zur Evolution - eine quantitative Studie; IDB Münster 14, S. 23-48

Jurecka, A. & Hartig, J. (2007). Computer- und netzwerkbasiertes Assessment. In Hartig, J. & Klieme, E. (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Leistungsdiagnostik ⍰ Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung* (S. 37-48). Bonn, Berlin: Bundesministerium für Bildung und Forschung.

Kattmann U., Janßenbartels, A., Müller, M. (2009) Selektion: die Entstehung von Giraffe und Okapi; Unterricht Biologie; Dossier: Evolution; Friedrich Verlag; S.48 - 53

Keogh, B. & Naylor, S. (2000). Concept cartoons in science education. Sandbach, Cheshire: Millgate House Publishers.

King, P.E. &, Young, M.J. (2002). An information processing perspective on the efficacy of instructional feedback. American Communication Journal, 5/2,
http://ac-journal.org/journal/vol5/iss2/articles/feedback.pdf

Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. Educational Measurement: Issues and Practice, 30/4, 28-37.

König, A. (2011). Von genutzten und ungenutzten Potenzialen, Computer + Unterricht 83, S. 6-9

Klafki, W. (1994). Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik. 4. Auflage, Weinheim: Beltz.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H. E. & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards*. Bonn: Bundesministerium für Bildung und Forschung.

Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms bei der DFG [Competence models for assessing individual learning outcomes and evaluating educational processes. Description of a new priority program of the German Research Foundation, DFG]. Zeitschrift für Pädagogik, 52, 876–903.

Kluger, A. N., & DeNisi, A. (1996). The effects of Feedback Interventions on performance: A historical review, a meta-analysis, and a preliminary Feedback Intervention Theory. Psychological Bulletin, 119/2, 254-284.

Kluger, A.N., & DeNisi, A. (1998). Feedback Interventions: Towards the understanding of a double-edged sword. Current directions in psychological science, 7, 67-72.

Koedinger, K. R., McLaughlin, E. A. & Heffernan, N. T. (2010). A quasi-experimental evaluation of an online formative assessment and tutoring system. Journal of Educational Computing Research, 43, 489-510.

Koeppen, K., Hartig, J., Klieme, E. & Leutner, D. (2008). Current Issues in Competence Modeling and Assessment. Journal of Psychology, 216/2, 61–73.

Koretz, D. (2008). Test-Based Educational Accountability. Zeitschrift für Pädagogik, 54/6, 777-790.

Langer, W. (2006). Mehrebenenanalyse - Eine Einführung für Forschung und Praxis. Wiesbaden: VS-Verlag.

Lehker, M. (2009). Rechtschreibung und Lesen computergestützt fördern, Deutschunterricht 62/3, 34-38

Leuders, T. (2004) Selbstständiges Lernen und Leistungsbewertung. Der Mathematikunterricht, 50/3, 63-79.

Leuders, T. (2006). Kompetenzorientierte Aufgaben im Unterricht. In W. Blum, C. Drüke-Noe, R. Hartung, & O. Köller (Eds.), Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen (S. 81–95). Berlin: Cornelsen.

Leutner, D. (2001). Pädagogisch-psychologische Diagnostik. In: Rost, Detlef H. (Hrsg.), Handwörterbuch Pädagogische Psychologie (S. 521-530). Weinheim: Beltz.

Lorenz, J. H. (2005). Zentrale Lernstandsmessung in der Primarstufe: Vergleichsarbeiten Klasse 4 (VERA) in sieben Bundesländern (Central student assessment in primary schools: comparative tests for grade 4 in seven federal states of Germany). Zentralblatt für Didaktik der Mathematik (ZDM), 37,(4), 317-324.

Lüdtke, O., Robitzsch, A., Trautwein U. und Köller O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen; Psychologische Rundschau, 58 (2), 103–117

Ludwig, L. (2013). „Schlagt eure Tablets auf“ – vom Hype zum Zukunftsmodell?, Computer + Unterricht 89, S. 24-26

Maier, U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft, 13*, 293-308.

Maier, U. & Randler, C. (2011). DFG Antrag zum Thema „Effekte formativer Leistungsmessung auf Schülerleistungen und Lernmotivation im naturwissenschaftlichen Unterricht“, Deutsche Forschungsgemeinschaft, 2011

Maier, U. (2012). Lehr-Lernprozesse in der Schule: Studium. Allgemeindidaktische Kategorien für die Analyse und Gestaltung von Unterricht. Bad Heilbrunn: UTB-Klinkhardt.

Maier, U., Hofmann, F., Zeitler, S. (2012). Formative Leistungsdiagnostik; Schulmanagement Handbuch 141, Oldenburg Verlag

Maier, U., Wolf, N. und Randler, C. (2014). Evolutionäre Anpassung am Beispiel von Vögeln: Eine Unterrichtseinheit für den Biologieunterricht der Sekundarstufe I, in: Lehr- Lernprozesse in der Schule: Referendariat; Maier, U. (Herausgeber); Klinkhardt UTB 2014, S. 40 – 65

Maier, U. (2014). Computergestützte, formative Leistungsdiagnostik in Primar- und Sekundarschulen. Ein Forschungsüberblick zu Entwicklung, Implementation und Effekten. Unterrichtswissenschaft, 42/1, 69-86.

Mayring, P. (2003). Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Beltz.

Mayring, Ph. & Gläser-Zikuda, M. (Hrsg.) (2005). Die Praxis der Qualitativen Inhaltsanalyse. Weinheim: Beltz/UTB.

McConnell, D. A., Steer, D. N., Owens, K. D., Knott, J. R., Van Horn, S., Borowski, W., Dick, J., Foos, A., Malone, M., McGrew, H., Greer, L., Heaney, P. J. (2006). Using conceptests to assess and improve student conceptual understanding in introductory geoscience courses. Journal of Geoscience Education, 54/1, 61-68.

Merrell, C. & Tymms, P. (2007). Identifying reading problems with computer-adaptive assessments. Journal of Computer Assisted Learning, 23/1, 27-35.

Nachtigall, W (1987). Vogelflug und Vogelzug. Rasch u. Röhring.

Narciss, S. (2006). Informatives tutorielles Feedback. Münster, Waxmann

Natriello, G. (1987). The impact of evaluation processes on pupils. Eduational Psychologist, 22/2, 155-175.

Nehm Ross H., Schonfeld, I. S. (2008). Measuring Knowledge of Natural Selection: A Comparison of the CINS, an Open-Response Instrument, and an Oral Interview; Journal of research in science teaching, 45/10, 1131–1160

Nehm, R. H. und Reilly, L. (2007). Biology Majors' Knowledge and Misconceptions of Natural Selection *www.biosciencemag.org, 57/3,* BioScience

Ott, O. (2013). Trends bei digitalen Schulbüchern, Computer + Unterricht, 89, S. 46-50

Panzacchi, C. (2009). Vögel, cbj Verlag, München

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). Knowing what students know. The science and design of educational assessment. Washington, DC: National Academic Press.

Rakoczy, K., Klieme, E., Bürgermeister, A. & Harks, B. (2008). The Interplay Between Student Evaluation and Instruction - Grading and Feedback in Mathematics Classrooms. Zeitschrift für Psychologie / Journal of Psychology 216/2, 111–124.

Ramaprasad, A. (1983). On the definition of feedback. Behavioral Science, 28/1, 4-13.

Rammstedt, B., John O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German, Journal of Research in Personality 41, 203 – 212

Randler, C. & Hummel, E. (2011). Vögel in unserer Umgebung ? Wir erforschen die Geheimnisse des Fliegens. RAAbits Realschule Biologie Tiere Beitrag 3, 1-30.

Randler, C., Hummel, E., Gläser-Zikuda, M., Vollmer, C., Bogner, F. X., & Mayring P. (2011). Reliability and validation of a short scale to measure situational emotions in science education. International Journal of Environmental and Science Education, 6, 355-365.

Randler, C. (2012). Field experiments in learning research. In Seel, N.M. (Ed.): Encyclopedia of the Sciences of Learning, Part F: 1293-1297.

Roschelle, J. Rafanan, K., Bhanot, R. Estrella, G., Penuel, B., Nussbaum, M. & Claro, S. (2010). Scaffolding group explanation and feedback with handheld technology: Impact on students' mathematics learning. Educational Technology Research and Development, 58, 399-419

Rosenwald, G. (2011). Lernwerkstatt Vögel. Ein Leben aus der Vogelperspektive; Kohlverlag, Kerpen 2011

Ruiz-Primo & Furtak (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry, Journal of Research in Science Teaching, 44/1, 57–84

Russell, M. K. (2010). Technology-aided formative assessment of learning: New developments and applications. In Andrade, H. L. & Cizek, G. J. (Eds.), *Handbook of formative assessment* (pp. 125-138). New York, London: Routledge.

Sacher, W. (2006). Schulische Leistungsdiagnose. Grundlagen und Probleme. Schulmagazin 5 - 10, 74/3, 5-8.

Sadler, R. (1989). Formative Assessment and the design of instructional systems, Instructional Science 18, 119 – 144

Sadler, R. (1998). Formative Assessment: revisiting the territory, Assessment in Education: Principles, Policy & Practice, 5/1, 77-84

Segers, M., Dochy, F., & De Corte, E. (1999). Assessment practices and students' knowledge profiles in a problem-based curriculum. *Learning Environments Research: An International Journal, 2*, 191-213.

Schlieszeit, J. (2013). Tablets im Unterricht nutzen, Computer + Unterricht 89, S.10-13

Schmidt, E. (2006). Missliche Missverständnisse bei dem Begriff Evolution, sachlogisch/didaktische Lösungs-Vorschläge, IDB Münster, Ber. Inst. Didaktik Biologie 15, S 69-94

Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In: Rost, D.H. (Hrsg.), Handwörterbuch Pädagogische Psychologie (S. 95-100). Weinheim: Beltz.

Schulmeister, R. (2005). Lernplattformen für das virtuelle Lernen, Evaluation und Didaktik, Oldenburg Wissenschaftsverlag, München / Wien

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), Perspectives of curriculum evaluation (Vol. 1, pp. 39-55). Chicago: Rand McNally.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.

Shavelson, R. J. & Ruiz-Primo, M. A. (1999). Leistungsbewertung im naturwissenschaftlichen Unterricht. Unterrichtswissenschaft, Übersetzung von Hella Beister, 27/2, S. 102-127.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M. K., Yin, Y. (2008). On the Impact of Curriculum-Embedded Formative Assessment on Learning: A Collaboration between Curriculum and Assessment Developers. Applied Measurement in Education, 21/4, 295-314.

Sjuts, J. (2006). Unterrichtliche Gestaltung und Nutzung kompetenzorientierter Aufgaben in diagnostischer Hinsicht. In: Blum, Werner; Drüke-Noe, Christina; Hartung, Ralf & Köller, Olaf (Hrsg.), Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen (S. 96-112). Berlin: Cornelsen.

Sjuts, J. (2007). Kompetenzdiagnostik im Lernprozess - auf theoriegeleitete Aufgabenstellung und -auswertung kommt es an. mathematica didacta, 30/2, 33-52.

Smit, R. (2008). Formative Beurteilung im kompetenz- und standardorientierten Unterricht, Beiträge zur Lehrerbildung, 26/3

Smit, R. (2009). Die formative Beurteilung und ihr Nutzen für die Entwicklung von Lernkompetenz. Baltmannsweiler

Stolpmann, B. E., Welling, S. (2011). Tablet-PCs in der Schule, Computer + Unterricht 81, S. 46-47

Struyven, K., Dochy, F., Janssens, S., Schelfhou, W. & Gielen, S. (2006). The Overall Effects of End-of-Course Assessment on Student Performance: A Comparison between Multiple Choice Testing, Peer Assessment, Case-Based Assessment and Portfolio Assessment. Studies in Educational Evaluation, 32/3, 202-222.

Tan, A. L. & Towndrow, P. A. (2009). Catalyzing student-teacher interactions and teacher learning in science practical formative assessment with digital video technology. Teaching and Teacher Education, 25/1, 61-67.

Thissen-Roe, A. Hunt, E. & Minstrell, J. (2004). The DIAGNOSER project: Combining assessment and learning. Behavior Research Methods, Instruments, and Computers, 36, 234-240.

Trumpower, D. L. & Sarwar, G. S. (2010). Effectiveness of structural feedback provided by Pathfinder Networks. Journal of Educational Computing Research, 43, 7-24.

Walter, J. (2008). Adaptiver Unterricht erneut betrachtet: Über die Notwendigkeit systematischer formativer Evaluation von Lehr- und Lernprozessen und die daraus resultierende Diagnostik und Neudefinition von Lernstörungen nach dem RTI-Paradigma. Zeitschrift für Heilpädagogik, 59, 202–215.

Wandersee, J. H., Good R. G. und Demastes S. (1995). Forschung zum Unterricht über Evolution: Eine Bestandsaufnahme; Zeitschrift für Didaktik der Naturwissenschaften, S. 43 – 54

Wang, K. H., Wang, T. H., Wang, W. L., & Huang, S. C. (2006). Learning styles and formative assessment strategy: Enhancing student achievement in web-based learning. Journal of Computer Assisted Learning, 22/3, 207-217.

Wang, T.-H. (2011). Developing web-based assessment strategies for facilitating junior high school students to perform self-regulated learning in an e-learning environment. Computers & Education, 57/2, 1801-1812.

Weitzel H. und Harald Gropengiesser H. (2009). Vorstellungsentwicklung zur stammesgeschichtlichen Anpassung: Wie man Lernhindernisse verstehen und förderliche Lernangebote machen kann, Zeitschrift für Didaktik der Naturwissenschaften

Wild, K.-P., Krapp, A., Schiefele, U., Lewalter, D., & Schreyer, I. (1995). Dokumentation und Analyse der Fragebogenverfahren und Tests. Berichte aus dem DFG-Projekt ‘‘Bedingungen und Auswirkungen berufsspezifischer Lernmotivation’’ Nr. 2 [Documentation and analysis of questionnaires and tests]. Neubiberg, Germany: Universität der Bundeswehr München.

Wilde, M, Bätz, K., Kovaleva, A. & Urhahne, D. (2009). Überprüfung einer Kurzskala intrinsischer Motivation (KIM). Zeitschrift für Didaktik der Naturwissenschaften; 15. Jahrgang (online Zeitschrift). Verfügbar unter: http://www.ipn.uni-kiel.de/zfdn/pdf/15_Wilde.pdf (Abruf am 6.10.2009).

Winter, F. (2003). Auf dem Weg zu einer Feedback-Kultur im Klassenzimmer. In: Lernende Schule, 6/21, 11-13.

Winter, F. (2006). Leistungsbewertung. Eine neue Lernkultur braucht einen anderen Umgang mit den Schülerleistungen. Baltmannsweiler: Schneider Verlag.

Winter, F. (2007). Fragen der Leistungsbewertung beim Lerntagebuch und Portfolio. In: Gläser-Zikuda, M. & Hascher, T. (Hrsg.), Lernprozesse dokumentieren, reflektieren und beurteilen. Lerntagebuch und Portfolio in Bildungsforschung und Bildungspraxis (S. 109-129). Bad Heilbrunn: Klinkhardt.

Yeh, S. S. (2006). Can rapid assessment reduce the pressure? Teachers College Record, 108/4, 621-661.

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., Tomita, M. K., Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. Applied Measurement in Education, 21/4, 335-359.

9 Eigene Veröffentlichungen

Maier, U., Wolf, N. und Randler C. (2014). Evolutionäre Anpassung am Beispiel von Vögeln: Eine Unterrichtseinheit für den Biologieunterricht der Sekundarstufe I, in: Lehr- Lernprozesse in der Schule: Referendariat: Maier, Uwe (Herausgeber); Klinkhardt UTB 2014, S. 40 – 65

10 Abbildungsverzeichnis

Abbildung 1: Formative Assessment - Zyklisches Modell nach Harlen (2011) 16
Abbildung 2: Hinderungsgründe beim Einsatz digitaler Medien im Unterricht (Ludwig (2013) 40
Abbildung 3: Funktionalitäten von Lernplattformen (König 2011) 49
Abbildung 4: Erstellung des Schülerkennworts für die Vorstudie 70
Abbildung 5: Verlaufsplan Vorstudie 72
Abbildung 6: Ablauf der randomisierten Feldstudie 80
Abbildung 7: Schülerinnen bei der Arbeit mit den Tablets 81
Abbildung 8: Wartung der Tablet PCs 84
Abbildung 9: Beispiel einer Single Choice Aufgabe in Moodle 98
Abbildung 10: Beispiel einer Multiple Choice Aufgabe in Moodle 98
Abbildung 11: Beispiel einer Single Choice Aufgabe zur Abfrage konzeptuellen Wissens 99
Abbildung 12: Beispiel einer Lückentextaufgabe in Moodle 101
Abbildung 13: Ausführliches Feedback - Treatmentgruppe 1 102
Abbildung 14: Zuordnungsaufgabe in der Testumgebung von Moodle 103
Abbildung 15: Zugangsdaten der Schüler und LogIn Screen von Moodle im Vergleich 112
Abbildung 16: Auswahl der Feedbackoptionen in Moodle 124
Abbildung 17: Eingabemaske für das Gesamtfeedback in Moodle 125
Abbildung 18: Weiteres Beispiel für das ausführliche Feedback der Treatmentgruppe 1 127
Abbildung 19: Dichotomes Feedback der Treatmentgruppe 2 129
Abbildung 20: Textbeispiel der Kontrollgruppe beim zweiten Messzeitpunkt 130
Abbildung 21: Concept Map zum Auftrieb 131
Abbildung 22: Dokument zum Eintragen des individuellen Lernfortschrittes (Vorstudie) 134
Abbildung 23: Eingabemaske in Moodle - Erstellung von Kursen und Tests 143
Abbildung 24: Moodle - Aktivität anlegen 144
Abbildung 25: Zusammenstellung der Tests - Fragenauswahl 145
Abbildung 26: Nutzerverwaltung in Moodle Navigation 146
Abbildung 27: Verwaltung der Kursteilnehmer in Moodle 146
Abbildung 28: Auswertung der Testergebnisse in Moodle auf Aufgabenebene 147
Abbildung 29: Auswertung der Testergebnisse in Moodle auf Antwortebene 147

Abbildung 30: Auswertung der Tests in Moodle - Statistik 148
Abbildung 31: Auswertung der Motivationstests in Moodle 148
Abbildung 32: Technische Realisierung der Evaluation 150
Abbildung 33: Geschätzte Randmittel (Posttest - Konzeptwissen) 170
Abbildung 34: Geschätzte Randmittel (Behaltenstest - Konzeptwissen) 170
Abbildung 35: Situative Emotion "Interesse" 172
Abbildung 36: Situative Emotion "Wohlbefinden" 172
Abbildung 37: Situative Emotion "Langeweile" 173
Abbildung 38: Screeplot zur Faktorenanalyse (Tab. 32) 182
Abbildung 39: Screeplot zur Faktorenanalyse "Bewertung der Unterrichtseinheit" 188

11 Tabellenverzeichnis

Tabelle 1: Formative assessment – Beurteilungsmethoden (vgl. Bürgermeister 2014) 14
Tabelle 2: Wissensarten in den Naturwissenschaften nach Shavelson & Ruiz-Primo (1999) 29
Tabelle 3: Ergebnisse der Feedbackforschung 35
Tabelle 4: Spektrum digitaler Schulbücher (aus: Ott 2013) 45
Tabelle 5: Fehlvorstellungen innerhalb des Konzeptes "Evolutionäre Anpassung" 54
Tabelle 6: Beispiel einer Testaufgabe zur "Evolutionären Anpassung" 55
Tabelle 7: Aufgabenformat aus den schriftlichen formativen Leistungstests der Vorstudie 65
Tabelle 8: Kontrollbogen Version 1 zum FLT - Vorstudie 66
Tabelle 9: Kontrollbogen Version 2 zum FLT - Vorstudie 67
Tabelle 10: Auswertung- Kontrollbogen 67
Tabelle 11: Kontrollbogen Version 3 zum FLT - Vorstudie 68
Tabelle 12: Übersicht über den Inhalt der Schülerordner aus der Vorstudie 69
Tabelle 13: Übersicht über den Inhalt der Testordner aus der Vorstudie 71
Tabelle 14: Überblick über die Lerninhalte und Messzeitpunkte der Studie 82
Tabelle 15: Aufgabentypen und ihre Eignung 95
Tabelle 16: Kurzskala 116
Tabelle 17: Fragen zum sozioökonomischen Hintergrund der Schüler (Motivationstest 1) 118
Tabelle 18: Big Five 120
Tabelle 19: Lösungsprozess einer Mathematikaufgabe zum Thema "Terme vereinfachen" 132
Tabelle 20: Vergleich der in der Studie verwendeten Tablets 153
Tabelle 21: Hauptkomponentenanalyse der Schulnoten 160
Tabelle 22: Komponentenmatrix der Schulnoten 161
Tabelle 23: Vergleichbarkeit der Treatmentgruppen 162
Tabelle 24: Reliabilität der Tests 164
Tabelle 25: Erstes allgemeines lineares Modell (mulitvariat) 166
Tabelle 26: Univariate Tests zum ersten allgemeinen linearen Modell 166
Tabelle 27: Zweites allgemeines lineares Modell (multivariat) 168
Tabelle 28: Univariate Tests zum zweiten allgemeinen linearen Modell 168

Tabelle 29: Geschätzte Randmittel der vier Treatmentgruppen 169
Tabelle 30: Intrinsische Motivation der Schüler 171
Tabelle 31: Situative Emotionen 171
Tabelle 32: Korrelationen 173
Tabelle 33: Teilkonzepte innerhalb des Konstruktes "Evolutionäre Anpassung" 174
Tabelle 34: Schülerantworten zu den konzeptuellen Fragen zur "Evolutionären Anpassung" 180
Tabelle 35: Ergebnisse zu den Schülerantworten im Bereich "Evolutionäre Anpassung" 181
Tabelle 36: Faktorenanalyse - "Umgang mit den Tablets" und "Feedbacknutzung" 182
Tabelle 37: Komponentenmatrix zu - "Umgang mit den Tablets" und "Feedbacknutzung" 182
Tabelle 38: Korrelationen mit dem Tablethandling 183
Tabelle 39: Faktorenanalyse "Bewertung der Unterrichtseinheit" 188
Tabelle 40: Komponentenmatrix zu "Bewertung der Unterrichtseinheit" 189
Tabelle 41: Unterschiede in den Treatmentgruppen (UE, Tablets, Feedbacknutzung) 189
Tabelle 42: Vor- und Nachteile elektronischer Leistungstests 200
Tabelle 43: Vor- und Nachteile von 7'' Tablets 202
Tabelle 44: Schülerantworten zu den Teilkonzepten der "Evolution" vor und nach der UE 216

12 Inhaltsverzeichnis des Anhangs

Anmerkung:

Alle unten aufgeführten Materialien und Tests können bei der PH Schwäbisch Gmünd angefordert werden.

1) Lehrermaterial
   - Unterrichtssequenz
   - Materialliste
   - Testverlauf
2) Tests
   - Vorstudie
     - Pre-Posttest
     - FLT 1-3 Version 2 (Treatmentgruppe)
     - FLT 1-3 Version 3 (Treatmentgruppe)
     - Kontrollbogen – erste Versuche
     - Kontrollbogen – Version 2
     - Kontrollbogen – Version 3
     - Test 1 der Kontrollgruppe
     - Test 2 der Kontrollgruppe
     - Test 3 der Kontrollgruppe
   - Randomisierte Feldstudie
     - Fragensammlung (das Dokument beinhaltet alle Fragen der elektronischen formativen Leistungstests)
     - Skalendokumentation
     - Texte Kontrollgruppe
     - Auswertung Mot2 Diagramme
     - Behaltenstest
     - Motivationstest 1 (bei der Feldstudie in elektronischer Form in Moodle)
     - Motivationstest 2 (bei der Feldstudie in elektronischer Form in Moodle)
     - Kurzskala
3) Konfiguration des Accesspoints

13 Danksagung

Zuallererst möchte ich mich für die großartige Betreuung durch meinen Doktorvater Herrn Professor Dr. Uwe Maier herzlichst bedanken. Ich bekam während der Promotion jederzeit Antworten auf meine Fragen und fühlte mich immer gut betreut. Vielen Dank für das Vertrauen und die Geduld.
Besonderer Dank gilt auch meinem Zweitbetreuer, Herrn Prof. Dr. Christoph Randler, der mir immer mit Rat und Tat zur Seite stand und mich vor allem bei biologischen und statistischen Fragen großartig unterstützt hat.
Während der gesamten Promotion bestand bei der Zusammenarbeit mit beiden Betreuern eine angenehme und freundschaftliche Atmosphäre.

Auch bei allen teilnehmenden Lehrkräften und Schülern, ohne die die Studie gar nicht möglich gewesen wäre, möchte ich mich herzlich bedanken: Vielen Dank, dass ihr die vielen Tests ohne Murren durchgestanden habt und bereit wart, mich an eurer pädagogischen Arbeit teilhaben zu lassen.

Ein herzliches Dankeschön gilt auch Markus Opferkuch, ohne den die Durchführung der elektronischen Leistungstests nicht möglich gewesen wäre – danke für die unermüdliche Hilfe bei der Konfiguration der Hardware und der Installation der Software.

Ganz besonders danke ich meiner Familie. Meinen Kindern Jonas und Oliver, meinem Mann Heiko, meinen Eltern und meinem Bruder. Es ist schön und wichtig für mich, dass ihr immer für mich da seid. Vielen Dank für eure Geduld, Liebe und Unterstützung.

Weiterhin möchte ich mich bei meinen Mitpromovenden und Arbeitskollegen bedanken: Carolin Ramsteck und Florian Hofmann, Sigrid Zeitler und Eva Fritzsche. Danke, dass ihr nicht nur Kollegen, sondern auch Freunde wart und immer ein offenes Ohr hattet, wenn es darum ging fachliche Themen auszudiskutieren.

Zum Abschluss möchte ich meinen Freunden ganz herzlich danken, weil sie mich über viele Tiefs beim Schreiben hinweggetragen und mich ständig motiviert haben, weiterzumachen. Im Besonderen denke ich da an Brigitte Kruse, Andreas Hölzer, sowie Christine und Matthias Klemm.