

Aus der Klinik für Kinder- und Jugendpsychiatrie  
Geschäftsführender Direktor:  
Herr PD Dr. Matthias Martin (komm. Direktor)

des Fachbereichs Medizin der Philipps-Universität Marburg  
in Zusammenarbeit mit  
dem Universitätsklinikum Gießen und Marburg GmbH,  
Standort Marburg

Titel der Dissertation:

# **Zur Bedeutung von orthographischen Fragmentwissen für die Lese- und Rechtschreibstörung**

Inaugural-Dissertation  
zur Erlangung des Doktorgrades der gesamten Humanmedizin  
dem Fachbereich Medizin der Philipps-Universität Marburg  
vorgelegt von:

Kristina Isabella Klein  
aus Saarbrücken

Marburg, 2008

Angenommen vom Fachbereich Medizin der Phillips-Universität Marburg am:  
03.07.2008

Gedruckt mit Genehmigung des Fachbereichs.

Dekan: Herr Prof. Dr. med. Matthias Rothmund

Referent: Herr Prof. Dr. med. Gerd Schulte-Körne

Koreferent: Frau Prof. Dr. med. Erika Baum

# Inhaltsverzeichnis

1	Einleitung.....	5
2	Theorie.....	8
2.1	Artificial Grammar Learning (AGL).....	8
2.1.1	Grundlagen und Vorgehensweisen des AGL-Experimentes.....	8
2.1.2	AGL – implizite vs. explizite Lernprozesse.....	9
2.1.3	Annahmen über die Art des erworbenen Wissen.....	11
2.1.3.1	Regelbasierend.....	13
2.1.3.2	Ähnlichkeitsbasierend.....	16
2.1.3.3	Ähnlichkeits- und regelbasierend.....	19
2.2	Symptomatik, Definition und Diagnostik der Lese- Rechtschreibstörung.....	24
2.3	Die Bedeutung des orthographischen Fragmentwissens für den Schriftspracherwerb... ..	27
2.3.1	Schriftspracherwerb – Begriffe des orthographischen Wissens und der phonologischen Bewusstheit .....	27
2.3.1.1	Modelle zur Lese- und Rechtschreibentwicklung.....	28
2.3.1.2	Zusammenhang zwischen dem orthographischen Wissen und der AGL-Forschung.....	29
2.3.2	Die Bedeutung von orthographischen Fragmentwissen für den Schriftspracherwerb und sein Zusammenhang mit der AGL-Forschung.....	30
2.3.3	Implizites Lernen in Zusammenhang zur Lese- und Rechtschreibstörung.....	35
3	Zielsetzung und Hypothesen.....	39
3.1	Zielsetzung.....	39
3.2	Hypothesen.....	39
3.3	Fragestellungen.....	41
4	Methode.....	43
4.1	Stichprobe.....	43
4.2	Datenerhebungen der unabhängigen Variablen .....	46
4.2.1	Rechtschreibleistung .....	46
4.2.2	Leseleistung.....	49
4.2.3	Intelligenz.....	51
4.3	Konstruktion des Artificial Grammar Learning- Experimentes.....	53
4.3.1	Konsequenzen aus der Literatur für die Konstruktion des AGL-Experimentes.....	53
4.3.2	Materialien.....	56
4.3.3	Durchführung und Instruktionen.....	59
4.3.3.1	Lernphase.....	60
4.3.3.2	Testphase.....	61
4.4	Datenerhebung der abhängigen Variable.....	63
4.5	Auswertestrategie der Daten.....	63
5	Ergebnisse.....	66
5.1	Testphase.....	66
5.2	Lernphase.....	74
5.2.1	In Abhängigkeit von ihrer Buchstabenlänge falsch geschriebene Items im Gruppenvergleich.....	75
5.2.2	Anzahl der gesehenen Items im Gruppenvergleich.....	76
5.2.3	Anzahl der bei allen vier Malen falsch geschriebenen Items im Gruppen- und Klassenvergleich.....	76
5.2.4	Zusammenfassung .....	78
5.3	Zusammenhang zwischen der Lernphase und der Klassifikationsleistung in der Testphase.....	78
5.4	Weitere Aspekte in der Auswertung der Testphase.....	80
6	Diskussion.....	82
7	Zusammenfassung.....	90
8	Anhang.....	93

9 Literatur.....	95
10 Appendix.....	104
10.1 Verzeichnis der akademischen Lehrer.....	104
10.2 Danksagung.....	105

# 1 Einleitung

Die vorliegende Arbeit untersucht das implizite Lernen von Grundschulkindern im Umgang mit schriftsprachlichen Material. Hierbei wird überprüft, inwiefern implizit orthographisches Fragmentwissen von den Kindern erworben wird und welche Bedeutung dieser Wissenserwerb auf die Lese- und Rechtschreibstörung hat.

Implizites Lernen wird charakterisiert als unbeabsichtigter, beiläufiger Erwerb von neuen Informationen, somit als Fähigkeit unterbewusst zu lernen (z. B. Cleermans, Destrebecqz & Boyer, 1998). Das daraus resultierende Wissen kann weitere Verarbeitungsprozesse beeinflussen, ohne das Bewusstsein, dass dieses Wissen erworben wurde oder dass es gegenwärtig Einfluss nimmt. Darüberhinaus ist es schwierig bzw. nahezu unmöglich, sich bewussten Zugang zu impliziten Wissens zu verschaffen und sich verbal darüber zu äußern.

Es wird angenommen, dass implizites Lernen ein fundamentaler und ubiquitärer Prozess in vielen verschiedenen Bereichen wie Spracherwerb, Erwerb von Lesen und Schreiben und Erwerb motorischer Fähigkeiten ist (z. B. Reber, 1993).

Implizites Lernen ist demnach grundlegend für den Schriftspracherwerb. Kinder erhalten zwar besonders über die Schule und auch über die Eltern durch instruiertes Lernen expliziten Zugang zu Rechtschreibwissen (z. B. über direktes oder regelgeleitetes Lehren von Wörtern). Steffler (2001) nimmt jedoch an, dass es aufgrund der Komplexität der geschriebenen Sprache unwahrscheinlich ist, dass der Schriftspracherwerb allein auf einem expliziten – einem bewusst herbeigeführten – Lernprozess beruht. Nach Steffler (2001) könnte implizites Lernen auch beim Schriftspracherwerb hinsichtlich der Anwendung von Rechtschreibfähigkeiten beteiligt sein.

Die Lese-Rechtschreibstörung (LRS) ist eine der häufigsten kinderpsychiatrischen Störungen (Prävalenz zwischen 3 und 5 %), die die Betroffenen nicht nur in der Kindheit, sondern auch als Erwachsene beeinträchtigt. Nach Haffner et al. (1998) ist trotz genügender nonverbaler Intelligenz der Besuch weiterführender Schulen selten, da

schriftsprachliche Fähigkeiten große Auswirkungen auf das Bildungsniveau haben. Die Betroffenen erreichten im Vergleich zu ihren kognitiven Fähigkeiten ein signifikant geringeres Schul- und Bildungsniveau (Esser & Schmidt, 1993). Nach Heiervang et al. (2001) zeigen Kinder mit einer Lese-Rechtschreibstörung umfangreiche Verhaltensauffälligkeiten. Der größte Anteil der Studien zur Ursachenforschung (Übersicht in Habib 2000) ist hauptsächlich defizitorientiert, v. a. im Bereich der auditiven und visuellen Informationsverarbeitung. Über den impliziten Anteil dieses Wissenserwerbes ist dagegen kaum etwas bekannt.

Cassar & Treimann (1997) und Pacton, Perruchet, Fayol & Cleeremans (2001) wiesen nach, dass Wissen über orthographische Regelmäßigkeiten schon sehr früh von Kindern ohne explizite Instruktionen von ihren Eltern oder Lehrern erworben werden kann. Für Cassar & Treimann (1997) beruht dieses Wissen auf Erfahrungen der Kinder mit Schriftzügen im Alltag. Auch Pacton et al. (2001) nehmen an, dass ein großer Teil orthographischen Wissens unter inzidentellen (beiläufigen) Lernbedingungen ständig im alltäglichen Kontakt mit Schriftmaterial aller Art erworben wird, also ein bedeutender Anteil von implizitem Lernen am Schriftspracherwerb beteiligt ist. Aaron, Wilczynski & Keetay (1998) vermuten, dass die regelgeleitete Komponente des wortspezifischen Gedächtnisses wahrscheinlich als Wissen von Buchstabeneinheiten in Bi- und Trigrammen vertreten ist, welches von der Häufigkeit ihres Vorkommens in der Schriftsprache bedingt wird, sogenanntes orthographisches Fragmentwissen.

Das Lernen zu Schreiben beinhaltet das Abstrahieren von Strukturen und Regelmäßigkeiten der Schriftsprache und Verwendung dieses Wissens zu einem späteren Zeitpunkt (Steffler, 2001). Die Forschung zum implizitem Lernen, die überwiegend auf Laborstudien mit dem Artificial Grammar Learning basiert, befasst sich mit dieser Fähigkeit Regeln (z. B. Reber, 1967,1989) bzw. ähnliche Strukturen (Fragmente) (Knowlton & Squire, 1994,1996a; Meulemanns & Van der Linden, 1997) in einer komplexen Form zu abstrahieren. Im Allgemeinen werden diese Experimente mit Artificial Grammar Learning (AGL) mit normalen Erwachsenen durchgeführt. Nur zu einem geringeren Anteil werden klinische Gruppen, wie Amnesie-, Alzheimer- und Parkinsonpatienten als Vergleich herangezogen (Steffler, 2001). Um die

Sprachentwicklung zu verstehen ist es essentiell, sich mit dem Entwicklungsaspekt der impliziten Wahrnehmung zu befassen (Steffler, 2001). Allerdings gibt es im Vergleich zu Untersuchungen an Erwachsenen kaum Studien, die sich mit dem impliziten Lernen von Kindern befassen. Insbesondere wurde ein Experiment mit Artificial Grammar Learning (AGL) nach Stand der Literatur gerade unter diesem Aspekt des Schriftspracherwerbs noch nicht mit Kindern durchgeführt<sup>1</sup>. Es ist möglich anhand von AGL-Experimenten festzustellen, wie die Kompetenz zur Rechtschreibung eventuell durch das implizite Anwenden neuer Kenntnisse beeinflusst wird. Ich frage mich, ob Kinder ohne ein Rechtschreibdefizit vielleicht in der Lage sind Regeln und Fragmente in einer Sprache besser zu abstrahieren, sich Regelwissen und sogenanntes orthographisches Fragmentwissen besser anzueignen und dieses implizit erworbene Wissen besser anzuwenden als Kinder mit einem Rechtschreibdefizit?

Um dieser Frage nachzugehen, untersuche ich in dieser Studie Grundschul Kinder mit Hilfe des Artificial Grammar Learning-Experimentes, die sich in ihrem Alter und der Lese- und Rechtschreibleistung unterscheiden.

Zunächst stelle ich das Artificial Grammar Learning-Experiment in Vorgehensweisen und theoretischen Annahmen sowie die Bedeutung des phonologischen und orthographischen Wissens vor, um deren Relevanz für den Erwerb der Schriftsprache und Schlussfolgerungen für das Untersuchungsdesign dieser Studie aufzuzeigen. In dem darauffolgenden methodischen Teil erläutere ich die Vorgehensweise der Untersuchung. Die im Anschluss dargestellten Untersuchungsergebnisse werden schließlich diskutiert, um unter anderem auch auf offene Fragen und weitere Forschungsmöglichkeiten einzugehen.

---

<sup>1</sup> Es wurden AGL-Experimente mit Kleinkindern durchgeführt (z. B. 12 Monate alte Kinder in Gomez & Gerken 1999, Gomez & Lakusta 2004), allerdings nicht auf Basis von schriftsprachlichen Material mit dem Hintergrund des Schriftspracherwerbes, sondern zur Untersuchung der Sprachentwicklung mit Hilfe von Lauten.

## 2 Theorie

### 2.1 Artificial Grammar Learning (AGL)

#### 2.1.1 Grundlagen und Vorgehensweisen des AGL-Experimentes

Artificial Grammar Learning (AGL) wird in vielen Laborstudien verwendet, um der Frage nach der Art und der Anwendung von implizit erworbenem Wissen nachzugehen.

In dem von Reber (1967) entworfenen klassische Experiment des AGL prägen sich die Probanden eine Anzahl von Buchstabenreihen („Lernitems“), die nach einem künstlichen grammatischen Regelsystem (einer so genannten „artificial grammar“) konstruiert wurden, anhand verschiedener Methoden, klassischerweise mit Hilfe eines „paper-and-pencil-test“ ein, also beispielsweise durch Präsentation und Nachschreiben der Items auf Papier. Erst im Anschluss an diese Lernphase wird den Probanden mitgeteilt, dass die Items nach einem bestimmten grammatischen Regelsystem, über welches keine genaueren Konstruktionsprinzipien bekannt gegeben werden, generiert wurden.

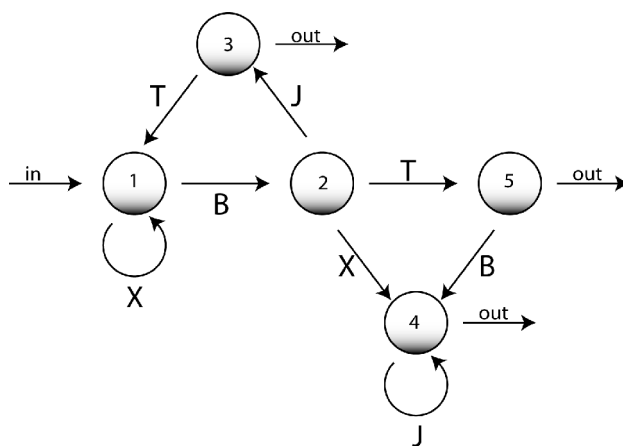


Abbildung 1

In dieser Abbildung ist eine von Abrams & Reber (1989) entwickelte Grammatik gezeigt, wobei im Original der Buchstabe B ein V ist. Die hier abgebildete Grammatik wurde in der vorliegenden Studie verwendet. Die Buchstabenreihen (Items) werden generiert, indem man vom „in“-Pfeil beginnend von Pfeil zu Pfeil die Buchstaben der Reihe nach anfügt, bis man zu einem „out“-Pfeil gelangt. Die Items haben hier eine Länge von 2 bis maximal 6 Buchstaben. Beispiele für richtig gebildete Items sind: XBT, XXBXJJ, BXJ... Nicht der Grammatik entsprechende, also falsch gebildete Items werden üblicherweise produziert, indem die Verknüpfungsregeln an mindestens einer Stelle verletzt werden, z. B. durch Vertauschen von Buchstaben oder dem zu frühen Abbruch der Reihe. Beispiele für falsch gebildete Items sind: XBXB, XXB, XBXX...



Reber (1967, 1989) nahm an, dass unter diesen indirekten, beiläufigen Lernbedingungen – dem Einprägen der Lernitems ohne das Wissen, dass die Items nach einem Regelsystem konstruiert wurden – implizite Lernvorgänge stattfinden.

Die Probanden werden nun dazu aufgefordert, in der folgenden Testphase neue „grammatische“, d. h. nach dem Regelsystem konstruierte Items, von „nichtgrammatischen“ Items, welche Regeln der Grammatik verletzten, also nach ihrer sogenannten 'Grammatikalität' zu unterscheiden. Die Unterscheidung der grammatischen von nichtgrammatischen Items wird im Folgenden Klassifikation genannt. Typischerweise werden die Testitems besser als nach Zufall klassifiziert, also über bzw. unter dem Zufallslevel von 50 %, obwohl die Probanden nicht in der Lage sind, die Regelstruktur verbal zu beschreiben oder die Gründe für ihre Entscheidungen in der Klassifikation der Testitems zu erklären.

Nach Reber basiert das Lernen also auf der Fähigkeit die durch die Grammatik vorgegebenen Regelmäßigkeiten in den Items herauszufinden und sie anzuwenden, ohne sich derer bewusst zu sein und ohne expliziten Zugang zu diesem Wissen zu haben, was die Annahme auf einen impliziten Lernprozess festigt (Reber 1967, 1989; Dienes, Altmann, Kwan & Goode 1995).

### **2.1.2 AGL – implizite vs. explizite Lernprozesse**

Es gibt allerdings auch Kritiker, die argumentieren, dass kein impliziter, sondern ein expliziter Lernvorgang stattfindet. Dunlany, Carlson & Dewey (1984, 1985) zeigten beispielsweise, dass die Probanden während der Präsentation der Buchstabenreihen explizites Wissen über eine eigene Grammatik, die mit der ursprünglichen Grammatik nur in Form von Mikroregeln korreliert, erwerben können, und dass dieses Wissen ausreichend erschien, um bewusst die Items in der Testphase zu klassifizieren.

Knowlton, Ramus & Squire (1992) unterstützen dagegen die Annahme eines impliziten Lernvorganges von Reber. Sie fanden heraus, dass Amnesiepatienten die Items in der Testphase genauso gut klassifizierten wie nicht von einer Amnesie betroffene Teilnehmer. Obwohl diese Patienten schwere Defizite haben, explizites Wissen

abzuspeichern – sie erzielten beispielsweise in einem Test, der die Regeln der Grammatik explizit lehrt, wesentlich schlechtere Ergebnisse als die Nichtbetroffenen – sind Amnesiepatienten in der Lage, implizit zu lernen und nicht-deklaratives Wissen<sup>2</sup> abzuspeichern. Dass Amnesiepatienten die Klassifikation der Testitems genauso gut durchführten wie Nichtbetroffene, sehen Knowlton et al. (1992, 1994, 1996) als Beweis dafür, dass explizites Wissen der Grammatik nicht maßgeblich zur Klassifikation der Testitems beiträgt.

In einem weiteren Experiment waren Nichtbetroffene im Vergleich zu den Amnesiepatienten in der Klassifikation der Testitems erfolgreicher, wenn sie vorher instruiert wurden, jedes Testitem bei der Klassifikation auf Ähnlichkeiten zu den Lernitems zu untersuchen. Diese Ergebnisse, so interpretierten Knowlton, Ramus & Squire (1992), würden zeigen, dass Nichtbetroffene explizites, bzw. deklaratives Wissen über spezifische Lernitems anwenden können, wenn sie dementsprechend vorher instruiert wurden. Gomez (1997) vermutet, dass je nach Intensität der Lernphase (ob das Wissen beispielsweise durch Erinnerungstests nochmals reflektiert wird), explizite Prozesse bei komplexeren Lernleistungen anzunehmen sind.

Hieran ist die Problematik zu erkennen, wie intensiv die Lernphase gestaltet werden kann, um noch implizite Lernvorgänge annehmen zu können und welche Instruktionen den Probanden gegeben werden sollten.

Die Lernphase in den meisten AGL-Experimenten ist so gestaltet, dass die Probanden dazu aufgefordert werden, die ihnen präsentierten Lernitems nachzuschreiben - früher mit Papier und Stift, also in Form eines „paper-and-pencil-test“ (z. B. Knowlton & Squire 1994, 1996), mittlerweile mittels Bildschirm und Tastatur (z. B. Gomez, 1997; Kinder & Assmann 2000; Gomez, Gerken & Schvaneveldt 2000; Kinder, Shanks, Cook & Tunney 2003).

Neben den Überlegungen, wie die Probanden sich mit den Lernitems auseinandersetzen, ergibt sich auch die Frage, mit welcher Intensität die Lernphase stattfindet, also wieviele von den aus der Grammatik maximal möglichen konstruierbaren Items den Probanden

---

2 Seit Anderson (1976) wird zwischen deklarativem Wissen, einem Wissen über Sachverhalte wie Faktenwissen und komplexe Zusammenhänge, und prozeduralem Wissen, einem Wissen über kognitive und psychomotorische Fertigkeiten, unterschieden.

als Lernitems präsentiert werden und wie oft diese Lernitems gezeigt werden sollen. Zu diesen Fragestellungen gibt es im Vergleich von AGL-Experimenten untereinander keine genauen Vorschriften. Es wird Kritik über Experimente geäußert, die auf ungewöhnlich langen Lernphasen beruhen – z. B.: Redington & Charter (1996) über ein Experiment von Mathews et al. (1989), das auf mehreren hundert Lerndurchführungen in einem Zeitraum von vier Wochen basierte, oder Knowlton & Squire (1994) über ein Experiment von Vokey & Brooks (1992). In einigen Experimenten werden die Lernitems beispielsweise insgesamt nur zweimal präsentiert, jedes falsch geschrieben Item wird noch einmal direkt danach gezeigt, bis zu maximal drei möglichen Versuchen, es richtig zu schreiben (Knowlton & Squire 1996). Ein ähnliches Vorgehen hatten Kinder, Shanks, Cock & Tunney (2003), die in ihrem Experiment die Lernitems allerdings insgesamt viermal präsentierten. Dagegen werden in anderen Experimenten alle Lernitems 4 – 6 mal wiederholt, und zusätzlich wird von den Probanden verlangt, ein falschgeschriebenes Item so oft nachzuschreiben, bis es mindestens einmal richtig geschrieben wurde, um davon ausgehen zu können, dass alle Lernitems richtig gelernt wurden (z. B. Gomez 1997; Gomez, Gerken & Schvaneveldt 2000).

Die Bedeutung dieser unterschiedlichen Anforderungen für die Durchführung einer Lernphase eines AGL-Experimentes mit Kindern und wie dies umgesetzt wurde, erläutere ich im methodischen Teil dieser Arbeit unter 4.3.1.

### **2.1.3 Annahmen über die Art des erworbenen Wissen**

Neben der Frage, ob im AGL-Experiment implizite oder explizite Lernprozesse stattfinden, gibt es viele Unstimmigkeiten über die Art des Wissens, welches während eines AGL-Experimentes erworben wird. Dies bezieht sich insbesondere auf das Wissen, welches den Probanden ermöglicht, die Testitems über dem Zufallslevel von 50 % zu klassifizieren. Ein Zufallslevel von 50 % ist beim Klassifizieren der Testitems zu erwarten, wenn die Probanden kein Wissen in der vorausgehenden Lernphase erworben haben oder dieses Wissen beim Klassifizieren der Items nicht anwenden konnten.

Die meisten Annahmen schlagen vor, dass die Informationen, die die Probanden verwenden, um die Testitems zu klassifizieren, alleine auf Wissen, welches sie sich in der Lernphase angeeignet haben, basieren.

Im Gegensatz dazu nehmen Whittlesea (1997) Überlappungen in der Verarbeitung der Lern- und Testitems an. Diese Annahme fordert alle anderen Annahmen, die sich einzig auf die Aneignung von Wissen während der Lernphase konzentrieren, heraus (Whittlesea & Dorken, 1993). Basierend aus den Ergebnissen ihrer Untersuchung (1993, 1997), schlagen Whittlesea et al. vor, dass Wissen, welches auf Verarbeitung basiert, zusätzlich zu Wissen aus der Lernphase erworben wird (d. h. dass die Probanden noch während der Testphase lernen), dass Instruktionen vor der Lernphase vorschreiben, welche Aspekte der Struktur der Lernitems verschlüsselt sind, und dass Probanden ihr Wissen explizit und implizit anwenden können, je nachdem, ob sie den Zusammenhang zwischen dem Verarbeitungsfluss und dem sich auf verschiedenen Wegen erworbenen Wissen in der Lernphase verstanden haben. Werden Probanden beispielsweise dazu aufgefordert, in der Lernphase Regeln zu identifizieren, führt dies zu einem aktiven und bewussten Bemühen und damit zu expliziten Prozessen (Whittlesea, 1997; Whittlesea & Dorken, 1993; Johnstone & Shanks, 2001). Nach Turner & Fischler (1993) ist die Klassifikationsleistung der Probanden geringer (bei schnellen Anforderungen in der Testphase), wenn sie in der Lernphase instruiert werden, aktiv nach Regeln zu suchen.

Neben der Bedeutung der Testinstruktionen ist auch interessant, was die Probanden implizit lernen, wenn diese speziellen Anforderungen an die Instruktionen erfüllt werden.

Hierüber gibt es verschiedene Ansatzpunkte, die einerseits eher eine Verarbeitung auf der Basis von abstrahierten Regeln („regelbasierend“, näher diskutiert in 2.1.3.1) befürworten oder andererseits eine Verarbeitung von Ähnlichkeiten der Items untereinander („ähnlichkeitsbasierend“, 2.1.3.2) vermuten. Daraus folgend gibt es auch Annahmen, dass beide Arten von Verarbeitung während des Experimentes stattfinden und damit zur Klassifikation regelbasierende und auch ähnlichkeitsbasierende Eigenschaften der Items eine Rolle spielen („ähnlichkeits- und regelbasierend“, 2.1.3.3).

Ich gehe nun näher auf diese einzelnen Annahmen ein, um danach die Schlußfolgerungen für das Design der hier vorliegenden Untersuchung zu erläutern.

### **2.1.3.1 Regelbasierend**

Bei der regelbasierenden Verarbeitung wird angenommen (z. B., Reber 1967, 1989; Mathews et al. 1989; Mathews & Roussel, 1997), dass das in einem AGL-Experiment erworbene Wissen vollständig auf Regeln basiert. Dieses Wissen besteht nach Reber (1967, 1989) aus sogenanntem **abstraktem Regelwissen** („abstract rule knowledge“), während Mathews et al. (1989) und Mathews & Roussel (1997) das sogenannte **kontinuierliche Regelwissen** befürworten.

Zunächst gehe ich näher auf die Annahme von Reber (1967,1989), das abstrakte Regelwissen, ein. Während des impliziten Lernvorgang in der Lernphase entsteht eine abstrahierte, geistige, unterbewusste Repräsentation eines komplexen Regelsystem, ein sogenanntes abstraktes Regelwissen, welches dann in der Testphase zur Klassifikation der Items unterbewusst herangezogen wird. Reber vermutete, dass die Probanden entweder wissen, ob ein Item grammatisch ist oder nicht, oder dass sie raten. Die Anwendung des Wissens findet somit nach Reber in Form eines Alles-oder-Nichts-Prozesses statt, entweder erfüllt das Testitem die Bedingungen der Regeln, die in Form von implizitem Wissen abstrakt repräsentiert werden, oder nicht. Diese Repräsentation enthält wenig, wenn überhaupt, Informationen zu spezifischen Itemeigenschaften, sondern basiert hauptsächlich auf der strukturellen Beziehung der einzelnen Items zueinander (Manza & Reber, 1997).

Zusammengefasst vermuten Manza & Reber (1997) also, dass keine spezifischen Eigenschaften der oberflächlichen, äußeren Struktur gelernt werden, sondern die Eigenschaften der Items untereinander implizit abstrahiert und gelernt werden.

Neben der Annahme eines Alles-oder-Nichts-Prozesses von Reber, schlugen Mathews & Roussel (1997) eine kontinuierlichere Anwendung des Regelwissens bei der Klassifikation der Testitems vor. Mathews & Roussel (1997) vermuten, dass die Regeln dabei untereinander in ihrem Einfluss auf die Klassifikationsentscheidungen konkurrieren können. Die Bedeutung einer Regel ist davon abhängig, wie oft und wie

sie sich in ihrer Wirkungsweise bewährt hat und wie viele andere Regeln zu der gleichen Entscheidung führen. Je nachdem wie bedeutend eine Regel ist, wie oft sie verwendet werden konnte und sich dabei bewährt hat, desto besser ist sie verfügbar und damit in ihrer Stärke, der sogenannten Regelstärke, größer. Mathews & Roussel (1997) vermuten, dass die Regeln an sich über das Gedächtnis abrufbar sind. Die Regelstärke jedoch wird implizit in der Klassifikationsentscheidung umgesetzt.

Mathews & Roussel (1997) nehmen also an, dass sich die Probanden bewusst sind über ihre Regeln, besonders über auffällige und häufige Buchstabenfragmente, und dieses Wissen zeigen, bzw. verbalisieren können, was expliziten Prozessen entspricht, während jedoch die relative Stärke der Regeln in einem impliziten Verarbeitungsprozess umgesetzt wird.

Die Annahme, dass Probanden auf der Basis von abstraktem Regelwissen in der Testphase klassifizieren, wurde desweiteren anhand von sogenannten Transferexperimenten, in denen Lern- und Testitems unterschiedliche oberflächliche Strukturen haben (z. B. andere Buchstaben, geometrische Figuren oder Tonsequenzen) aber mit der gleichen unter-schweligen abstrakten Struktur (mit dem gleichen grammatischen Regelsystem) konstruiert wurden, demonstriert. Die Transferleistung der Probanden wird beispielsweise in einer Studie von Reber (1967) nachgewiesen, in der die Probanden Testitems, die aus einem anderen Satz von Buchstaben bestanden, aber mit der gleichen Grammatik wie die Lernitems konstruiert wurden, im weiteren bezeichnet als Transferitems, erfolgreich klassifizierten, während schlechtere Ergebnisse erzielt wurden, wenn die Testitems anhand einer neuen Grammatik generiert wurden. Die direkte Schlussfolgerung von Reber aus diesen Transferleistungen war, dass die Probanden etwas über die tiefe unter-schwellige Struktur der Lernitems, unabhängig von ihren spezifischen, äußeren Merkmalen, gelernt haben mussten.

Allerdings zeichnen sich viele dieser Studien durch ungewöhnlich intensive Lernphasen aus, die zum Teil auf mehreren hundert Durchführungen in einem Zeitraum von vier Wochen basieren (Mathews et al., 1989). Redington & Charter (1996) wiesen in den Transferexperimenten auf methodische Defizite, unter anderem auch auf das Fehlen von

Kontrollgruppen<sup>3</sup> hin, die die Interpretation dieser Experimente erheblich erschweren. In einigen neueren Studien wurde die Transferleistung der Probanden unter gleichen Bedingungen in der Testphase, aber mit weniger intensiveren Lernphasen gefunden (Gomez & Schvaneveldt, 1994; Whittlesea & Dorken, 1993; Knowlton & Squire, 1996, Experiment 3), ebenso wurden auch einige unter Kontrollbedingungen durchgeführt (Gomez & Schvaneveldt, 1994; Altmann, Dienes & Goode, 1995).

Die einzige Gemeinsamkeit zwischen den Items der Lern- und Testphase in Transferexperimenten ist die unterschwellige abstrakte Struktur. Da allerdings die Probanden unter klassischen Testbedingungen ohne erforderliche Transferleistung im Vergleich zu entsprechenden Transfertests erfolgreicher klassifizierten, lässt vermuten, dass zusätzlich zu dem Einfluss der abstrakten Informationen über die Struktur der grammatischen Regeln noch weitere spezifische Informationen ausgehend von den Lernitems für die Klassifizierung der Testitems von Bedeutung sein könnten (z. B. Knowlton & Squire, 1996). Tatsächlich lässt die Annahme über abstraktes Regelwissen außer Acht, dass auch Kenntnisse über einige oberflächliche Merkmale, insbesondere der Anfang und das Ende der Items und mehrmals auftretende auffällige Teile der Items, erworben werden könnten. Diese Vermutung wurde in weiteren Transfertests insbesondere von Knowlton & Squire (1996, Experiment 3) überprüft, welche durch ihre Ergebnisse die Annahmen, dass im AGL-Experiment Kenntnisse über spezifischere weniger abstrakte Informationen zusätzlich erlangt werden könnten, unterstützen. Als wichtig für diese Merkmale werden insbesondere Buchstabenwiederholungen als auffällige Buchstabenschemata erachtet (Mathews & Roussel, 1997; Brooks & Vokey, 1991; Gomez, Gerken & Schvaneveldt, 2000). Gomez, Gerken & Schvaneveldt (2000) demonstrierten anhand einer Grammatik, die keine Wiederholungselemente generierte, dass trotz erfolgreicher Klassifikation von Testitems (ohne Transferanforderung) nur

---

3 Kontrollgruppen lernen die Lernitems nicht, sondern absolvieren nur die Testphase. Sie sollten dabei die Testitems der Transferexperimente lediglich mit Hilfe des Hinweises über die Verteilung der grammatischen und nicht grammatischen Items diese nur auf Zufallsebene klassifizieren, um zu zeigen, dass ohne die Konfrontation mit den Lernitems, die grammatischen Testitems nicht erfolgreich, also über oder – bei vertauschter Zuordnung – unter Zufall, klassifiziert werden können. Redington & Charter (1996) räumten diesbezüglich allerdings auch ein, dass Probanden aus diesen Kontrollgruppen durch ihre Hilflosigkeit eher nach diesen auffälligen Merkmalen suchen, als Probanden, die sich schon Kenntnisse über die Items erwerben konnten. Es sind nur wenige Transferexperimente unter diesen Kontrollbedingungen durchgeführt worden, beispielsweise von Altmann, Dienes & Goode (1995) und Gomez & Schvaneveldt (1994), deren Ergebnisse Redington & Charter (1996) als verlässlich erschienen und einen ziemlich schlüssigen Beweis eines echten Transfereffektes anbieten.

eine geringe Klassifikationsleistung gegenüber Transfertestitems erfolgte. Da die Probanden allerdings eine bessere Transferleistung demonstrierten, wenn Wiederholungselemente vorhanden waren, vermuten Gomez, Gerken & Schvaneveldt (2000), dass eher die Struktur von sich wiederholenden Elementen unter Transferbedingungen abstrahiert wird, als dass sich die Probanden abstraktes Regelwissen aneignen.

Anhand dieser Hinweise kann vermutet werden, dass in der Klassifikation nicht nur abstrahiertes Regelwissen, sondern auch weitere Eigenschaften der Items (z. B. Wiederholungselemente oder andere auffällige Fragmente) bedeutend für implizite Lernprozesse sind. Auf diesen Aspekt gehe ich nun in den ähnlichkeitsbasierenden Ansichten ein.

### **2.1.3.2 Ähnlichkeitsbasierend**

In den ähnlichkeitsbasierenden Ansichten befürworten Brooks (1978), Brooks & Vokey (1991) und Vokey & Brooks (1992) den Erwerb eines **Wissens über spezifische, komplette Items** („specific whole item knowledge“), während Perruchet & Pacteau (1990), Redington & Charter (1996) und Servan- Schreiber & Anderson (1990) den Erwerb von sogenanntem **Fragmentwissen** („fragment knowledge“) annehmen.

Die Annahme des Wissens über spezifische, komplette Items beinhaltet, dass sich Probanden an spezifische komplette Lernitems erinnern, während sie die Testitems nach der gegenüber den Lernitems bestehenden Ähnlichkeit klassifizieren (Brooks & Vokey, 1991; Vokey & Brooks, 1992). Hierbei stellten Vokey & Brooks (1992) den Satz der Testitems so zusammen, dass dieser aus einer Hälfte aus Items bestand, die die Regeln der Grammatik befolgten, im folgenden grammatische Items genannt, und zu der anderen Hälfte Items beinhaltete, die die Regeln der Grammatik verletzen, im folgenden nicht grammatische Items genannt. Außerdem wurde je die Hälfte der grammatischen und nicht grammatischen Testitems in ähnlich (Unterschied zu dem entsprechenden Lernitem nur ein Buchstabe) und unähnlich (Unterschied zum entsprechenden Lernitem zwei oder mehr Buchstaben) geteilt, so dass insgesamt vier Gruppen von Testitems entwickelt wurden. Dieses Prinzip der Aufteilung in vier verschiedene Gruppen wird



auch in der vorliegenden Untersuchung angewendet. Anhand dieses Testmaterials fanden Vokey & Brooks (1992) heraus, dass beide Eigenschaften der Testitems, die Ähnlichkeit zu den Lernitems und die Zugehörigkeit zur Grammatik, Einfluss auf die Klassifizierung der Testitems haben. Vokey & Brooks (1992) zeigten in einem weiteren Experiment, dass auch unter Transferbedingungen die Ähnlichkeit der Items bei der Klassifikation eine Rolle spielt. Vokey & Brooks (1992) vermuteten, dass sich die Probanden spezifische, komplette Items eingeprägt haben und sich an diese in der Testphase erinnerten.

Diese Erkenntnisse überprüften Knowlton & Squire (1994), indem sie das Experiment nochmals durchführten. Knowlton & Squire (1994) kritisierten, dass die Ergebnisse des Transferexperiments von Vokey & Brooks (1992) auch abhängig von ungewöhnlich intensiven und langen Lernphasen sein könnten. Außerdem fanden Knowlton & Squire (1994) anhand ihrer Ergebnisse heraus, dass der Effekt der Ähnlichkeit auf eine andere Variable, der Fragmentstärke, zurückzuführen sein könnte. Um den Begriff Fragmentstärke zu erläutern, gehe ich zunächst auf die Annahme des Fragmentwissens ein.

Die Annahme des Fragmentwissens schlägt vor, dass Artificial Grammar Learning nicht auf der Ähnlichkeit von kompletten Items untereinander basiert, sondern auf einem „Vertrautwerden“ mit von der Grammatik erlaubten Fragmenten der Lernitems in Form von Buchstaben-Bi- oder Trigrammen. Je mehr dieser Fragmente mit denen der Testitems übereinstimmen, desto eher werden sie als grammatisch klassifiziert (Perruchet & Pacteau, 1990; Perruchet, 1994; Perruchet, Gallego, & Pacteau, 1992).

Perruchet & Pacteau (1990) zeigten erstmals unter diesem Aspekt, dass eine gleichwertige Klassifikation der Testitems erfolgen kann, wenn die Probanden in der Lernphase nur mit einer Auflistung der möglichen Bigramme (Buchstabenpaare) der Lernitems konfrontiert werden im Vergleich zu einer Kontrollgruppe, die in der Lernphase die kompletten Lernitems sahen. Perruchet & Pacteau (1990) legten damit die Grundlage anhand ihrer Annahme von Bigrammwissen, dass Fragmentwissen eine bedeutende Rolle bei der Klassifizierung der Testitems spielt.

Gomez & Schvaneveldt (1994) demonstrierten allerdings, dass zwei verschiedene Arten von Bigramm-Verletzungen existieren. Die Probanden, die nur die Bigramme in der Lernphase erlernten, konnten auch nur die Testitems mit unbekanntem (nichtgrammatischen) Bigrammen identifizieren, auch Abhängigkeitsverhältnisse erster Ordnung genannt, während die Probanden, die mit kompletten Items trainierten, auch Testitems als nichtgrammatisch identifizieren konnten, die nach der Grammatik zwar mögliche Bigramme enthielten, welche sich aber an der falschen Position befanden, auch Abhängigkeitsverhältnisse zweiter Ordnung genannt. Zusätzlich konnten Gomez & Schvaneveldt (1994) hier zeigen, dass Wissen über Abhängigkeitsverhältnisse erster Ordnung nicht ausreichend für eine erfolgreiche Transferleistung ist, während anhand von Wissen über Abhängigkeitsverhältnisse zweiter Ordnung eine Transferleistung erfolgreich erbracht werden konnte. Gomez & Schvaneveldt (1994) nahmen daraufhin an, dass Transfer eine Abstraktion von Abhängigkeitsverhältnissen zweiter Ordnung reflektiert.

Die Vermutungen von Perruchet & Pacteau (1990) unterstützten auch Redington & Charter (1996). Sie entwickelten ein Computermodell (sogenanntes „toy-model“), welches den Erwerb und das Anwenden von Wissen über Fragmente (Bi- und Trigrammen) in AGL-Experimenten simulierte. Sie demonstrierten damit, dass es möglich ist, die Ergebnisse von Transferexperimenten auch nur durch Kenntnisse der Probanden über Fragmente (Bi- und Trigramme) der Lernitems und dem Finden von Analogien in Fragmenten von Testitems zu erklären, ohne dass abstraktes Regelwissen erforderlich sei. In ihrer Untersuchung bestätigten Redington & Charter (1996), dass alleiniges Bigrammwissen nicht immer ausreichen schien, um die Ergebnisse zu erklären, wie bereits Gomez & Schvaneveldt (1994) schon vermutet hatten. Dagegen waren interessanterweise Modelle, die Trigrammeigenschaften verwendeten und Wissen über Trigramme simulierten in der Lage, die Transferleistungen vorherzusagen.

Servan-Schreiber & Anderson (1990) schlugen bereits eine ähnliche Idee in einem formaleren Modell („competitive chunking model of artificial grammar learning“) vor, in dem sie im Gegensatz zu Perruchet & Pacteau und Vokey & Brooks nicht nur die Art des im AGL-Experiment erworbenen Wissens beschrieben, sondern auch darlegten, wie

sich dieses Wissen aufbaut und verstärkt. Testitems werden in dem Ausmaß nach grammatisch oder nichtgrammatisch klassifiziert, wie sie die Fragmente (Bi- und Trigramme oder auch längere Fragmente) der Lernitems enthalten, so genannte Fragmentstärke. In der Testphase werden die Testitems implizit nach ihrer Vertrautheit beurteilt, je vertrauter, desto eher werden sie als grammatisch klassifiziert. Die Vertrautheit zu einem Item hängt davon ab, wie stark die Repräsentationen seiner Subitems (Fragmente) sind, und folglich, wie oft sie in der Lernphase auftraten. Anhand dieses Modells war es allerdings noch nicht möglich, Transferleistungen von Probanden zu erklären, da hier keine Abstraktionsleistungen angenommen wurden.

Dieser erste Ansatz zur Fragmentstärke von Servan-Schreiber & Anderson (1990) wurde von Knowlton & Squire (1994) weiterentwickelt. Diese Weiterentwicklung erläutere ich im folgenden Abschnitt, da die Fragmentstärke zur näheren Charakterisierung der Items auch in dieser Untersuchung verwendet wird.

### **2.1.3.3 Ähnlichkeits- und regelbasierend**

Die ähnlichkeits- und regelbasierenden Annahmen gehen davon aus, dass sich die Probanden während dem AGL-Experiment **Regelwissen und Fragmentwissen** (Knowlton & Squire, 1994,1996; Meulemanns & Van der Linden, 1997) aneignen und anwenden. Kinder & Assmann (2000) und Johnstone & Shanks (1999) widerlegten den Erwerb und die Anwendung von Regelwissen, in dem sie mit Hilfe einer bestimmten Auswertungsmethode, einer **Multiplen Regressionsmethode**, zeigten, dass allein ähnlichkeitsbasierendes Wissen für die Klassifikation der Items erklärbar sein könnte.

Zunächst gehe ich jedoch auf die Annahmen, die das Regel- und Fragmentwissen befürworten, näher ein. Knowlton & Squire (1994, Experiment 2b) entwickelten die Idee des „competitive chunking model of artificial grammar learning“ von Servan-Schreiber & Anderson (1990) weiter und überprüften mit dem gleichen Experiment die Ergebnisse von Vokey & Brooks (1992).

Knowlton & Squire (1994) berechneten die Fragmentstärke<sup>4</sup> und führten somit erstmals ein Maß für die Ähnlichkeit ein. Basierend auf dem Material von Vokey & Brooks (1992), konnten sie mit Hilfe der Berechnung der Fragmentstärke zeigen, dass die Klassifizierung der Testitems nach ihrer Grammatikalität (grammatisch vs. nichtgrammatisch) unabhängig von der Ähnlichkeit (Unterschied des Testitems zum entsprechenden Lernitem nur ein Buchstabe) der Testitems zu den Lernitems erfolgt, wenn die Fragmentstärke der Testitems, die ähnlich zu den Lernitems waren, gleichgroß war wie die Fragmentstärke der Testitems, die unähnlich zu den Lernitems waren. Ein Effekt der Ähnlichkeit der Testitems zu den entsprechenden Lernitems konnte nur gezeigt werden, wenn die Fragmentstärke der Testitems unbalanciert war. Sie bezweifelten somit die Annahme, dass die Ähnlichkeit zwischen kompletten Testitems und den entsprechenden Lernitems einen signifikanten Einfluss auf die Klassifizierung der Testitems hat.

Anhand dieser Ergebnisse konnten Knowlton & Squire (1994) zeigen, dass Wissen über die Fragmentstärke einen wichtigen Einfluss auf die Klassifikation der Testitems hat und dies implizit gelernt wird (im weiteren bezeichnet als implizites Fragmentwissen).

Um nun eine Aussage über den Einfluss von Fragmentwissen und Regelwissen unabhängig voneinander treffen zu können, konstruierten Knowlton & Squire (1996, Experiment 1) die Testitems so, dass der Durchschnitt der Fragmentstärke der grammatischen und nichtgrammatischen Testitems in etwa gleich war, wobei die Testitems je zur Hälfte eine hohe bzw. eine niedrige Fragmentstärke hatten. Es entstanden so vier Arten von Testitems: grammatisch mit hoher Fragmentstärke (g/hFs), grammatisch mit niedriger Fragmentstärke (g/nFs), nichtgrammatisch mit hoher Fragmentstärke (ng/hFs) und nichtgrammatisch mit niedriger Fragmentstärke (ng/nFs). Normalerweise haben die grammatischen Items eine höhere Fragmentstärke als nichtgrammatische Items, weil nichtgrammatische Items oft aus Fragmenten bestehen, die überhaupt nicht in den Lernitems vorkommen, also eine Fragmentstärke von 0 haben. Das Ergebnis von Knowlton & Squire (1996, Experiment 1) war, dass beide

4 Die Fragmentstärke jedes einzelnen Testitems wird berechnet, indem gezählt wird, wie oft die Fragmente des Testitems in den Lernitems vorkommen, und daraufhin dieser Wert durch die Gesamtanzahl der im Testitem enthaltenen Fragmente geteilt wird. Beispiel: In dem Item BXJ sind die drei Fragmente BX, XJ und BXJ enthalten. Das Bigramm BX kommt zwölf mal, das Bigramm XJ und das Trigramm BXJ je acht mal in den Lernitems vor, also beträgt die Fragmentstärke  $9,33$  ( $12 + 8 + 8 = 28$ ;  $28 : 3 = 9,33$ ).

Aspekte, Fragmentstärke und Grammatikalität (Zugehörigkeit zur Grammatik), die Klassifikation der Testitems beeinflussen. Die signifikante Interaktion zwischen Fragmentstärke und Grammatikalität wies darauf hin, dass diese beiden Variablen die Klassifikation der Testitems nicht unabhängig voneinander beeinflussen. Ihre Ergebnisse zeigten sogar, dass die Grammatikalität anscheinend eine größere Rolle in der Klassifizierung der Testitems spielt. Sie räumten allerdings ein, dass der Effekt der Fragmentstärke größer hätte sein können, wenn der Unterschied zwischen hoher und niedriger Fragmentstärke größer gewesen wäre.

In der vorliegenden Untersuchung wird auf der Basis dieses Designs von Knowlton & Squire (1996, Experiment 1) ein ähnliches Testmaterial mit Verwendung der selben Grammatik zusammengestellt, wobei jedoch der Unterschied zwischen Testitems mit hoher Fragmentstärke und Testitems niedriger Fragmentstärke größer war (siehe auch 4.3.2).

Dieses Experiment 1 (1996) wurde mit einer Gruppe von Amnesiepatienten und einer Kontrollgruppe von nicht von einer Amnesie betroffenen Personen durchgeführt. Die Ergebnisse der Amnesiepatienten entsprachen nahezu denen der Kontrollgruppe, so dass Knowlton & Squire davon ausgingen, dass sich Amnesiepatienten auf die gleiche Art von Informationen beriefen, um die Testitems zu klassifizieren und genauso sensitiv auf die zwei Variablen reagierten wie die Kontrollgruppe. Knowlton & Squire nahmen deswegen an, dass konkretere Informationen zu den Lernitems, wie das Fragmentwissen, gleichermaßen implizit gelernt werden können wie andere Informationen, die auf Regelwissen basieren. In zwei weiteren Experimenten unterstützten Knowlton & Squire (1996, Experiment 2 und 3) mit weiteren Ergebnissen ihre Annahme vom impliziten Lernen abstrakter und itemspezifischer Informationen.

Meulemann & Van der Linden (1997) lieferten zusätzlich noch den interessanten Aspekt, dass nach dem Lernen von 32 Buchstabenitems (Experiment 2a) die Probanden die Testitems eher anhand von Fragmentwissen klassifizierten, wohingegen nach 125 Lernitems (Experiment 2b) auf der Basis von Regelwissen klassifiziert wurde.

Zusammengefasst vermuten Knowlton & Squire (1996) und Meulemans & Van der Linden (1997), dass in dem AGL-Experiment eine Kombination von regel- und

ähnlichkeitsbasierendem Lernen stattfindet und beide Arten von impliziten Wissen verwendet werden, um die Testitems zu klassifizieren.

Auf dieser Basis vermute ich auch in der vorliegenden Untersuchung, dass die Kinder während dem AGL-Experiment implizit lernen werden und sich Fragmentwissen, sowie mit der Annahme, dass eher grammatische Testitems als zugehörig klassifiziert werden, sich auch Regelwissen aneignen.

Kinder & Assmann (2000) zeigten allerdings mit Hilfe einer Multiplen Regressionsmethode, ursprünglich entwickelt von Lorch & Myers (1990), dass es letztendlich gelingen kann, wenn die Items in genügend oberflächlichen Eigenschaften analysiert werden, dass der Einfluss der Grammatik bei der Klassifikation verschwindet und damit nicht davon ausgegangen werden kann, dass Regelwissen erworben wird.

Im folgenden wird nun kurz auf diese Herausforderung eingegangen, um deren Problematik und Relevanz für die vorliegende Untersuchung zu erläutern. Kinder & Assmann (2000) kritisierten, dass es anhand der gängigen Versuchsdesigns nicht möglich sei, die Items in allen oberflächlichen Eigenschaften, die möglicherweise Entscheidungen über die Zugehörigkeit der Testitems zur Grammatik beeinflussen, auszubalancieren. Sie befürworteten eine weitere mögliche Annäherung an die Art des erworbenen Wissens mit Hilfe der Multiplen Regressionsmethode, die als erstes von Johnstone & Shanks (1999) auf AGL-Daten angewendet wurde. Diese Methode erfasst durch Einführung weiterer vorhersagender Variablen, welche Auswirkungen verschiedenartige Eigenschaften der Lernitems auf die Klassifizierung der Testitems haben. Vorteil dieser Methode ist, dass Einflüsse auf die zahlreichen Eigenschaften der Testitems isoliert werden können, ohne sie vorher daraufhin auszubalancieren. Meinung von Kinder & Assmann (2000) ist, dass, wenn alle oberflächlichen Eigenschaften, die man sich vorstellen kann, als Voraussage mit eingeschlossen werden, und der grammatische Status der Items dann immer noch eine signifikante Rolle spielt, dies darauf hinweisen würde, dass sich die Probanden Wissen über die Grammatik, welches unabhängig von den oberflächlichen Strukturen ist, angeeignet haben. Nach Replikation des Experimentes 1 von Knowlton & Squire (1996) und Analyse ihrer gewonnenen Daten mit neun vorhersagenden Variablen auf Basis der Multiplen Regressionsmethode,

konnten Kinder & Assmann (2000) anfangs vier signifikante Variablen finden, worin allerdings neben Vertrautsein mit den Anfangsbuchstaben, globaler Fragmentstärke, Vertrautsein mit dem Schema der Buchstabenwiederholungen, immer noch der grammatische Status signifikanten Einfluss auf die Klassifikation der Testitems hatte. Dieser verlor seine statistische Signifikanz erst in einer weiteren Regressionsanalyse bei der zusätzliche Variablen, die Informationen über einzelne Buchstaben erfassten – es wurden Items, die den Buchstaben „J“ enthielten eher abgelehnt, als Items ohne „J“ –, mit berechnet wurden. Kinder & Assmann (2000) schlossen hieraus (und aus einer weiteren Analyse, auf die in diesem Zusammenhang nicht näher eingegangen wird), dass doch eher eine vollständig ähnlichkeitsbasierende Verarbeitung der Items während eines AGL-Experimentes stattfindet.

Offen bleibt für mich hier die Problematik: Was ist noch oberflächige Struktur, was ist schon grammatische Struktur? Werden von der Grammatik nicht oberflächige, äußere Strukturen mit vorgegeben (wie z. B. Buchstabenwiederholungen)? Es ist nachvollziehbar, dass wenn ausreichend viele grammatische Eigenschaften von der Analyse oberflächiger Strukturen erfasst werden, die übrig bleibenden grammatischen Eigenschaften der Items keinen signifikanten Einfluss mehr auf die Klassifikation haben.

Entscheidend für die vorliegende Untersuchung ist, dass in diesen vorherigen Untersuchungen (Knowlton & Squire, 1996; Kinder & Assmann, 2000) ein Testmaterial für das AGL-Experiment verwendet wurde, welches sich auch beim Replizieren als zuverlässig erwies. Dieses Testmaterial wurde auch in einem Experiment von Pothos & Kirk (2004), auf welches in 2.3.3 noch eingegangen wird, verwendet. Für die vorliegende Untersuchung ist primär wichtig, dass es gelingt, das AGL-Experiment mit Kindern durchzuführen und der Frage nachzugehen, ob sich rechtschreibgestörte Kinder von Kindern ohne Rechtschreibstörung unterscheiden.

In dem folgenden Abschnitt erläutere ich nun die Symptomatik, Definition und Diagnostik der Lese-Rechtschreibstörung, um im Anschluss darauf einzugehen, inwiefern mit Hilfe eines AGL-Experimentes Rückschlüsse auf das implizite Lernen von orthographischen Fragmentwissen und auf dessen Relevanz für den

Schriftspracherwerb bei Kindern mit einer Rechtschreibstörung im Vergleich von nicht betroffenen Kindern gezogen werden können.

## **2.2 Symptomatik, Definition und Diagnostik der Lese-Rechtschreibstörung**

Ich stelle die Symptome der **Lese-Rechtschreibstörung** kurz auf Basis der in den AWMF-Leitlinien der Deutschen Gesellschaft für Kinder- und Jugendpsychiatrie und -psychotherapie beschriebenen Leitsymptomatik vor. Die Symptomatik lässt sich in die Bereiche des Lesens und des Schreibens unterteilen.

Die Symptomatik der Lesestörung äußert sich durch eine niedrige Lesegeschwindigkeit, die begleitet wird von Startschwierigkeiten und langem Zögern beim Vorlesen, Verlust der aktuellen Leseposition und Fehler beim Auffinden des Zeilenbeginns. Worte oder Wortteile werden ausgelassen, ersetzt, verdreht oder hinzugefügt. Worte im Satz und Buchstaben in Wörtern werden vertauscht und Sätze falsch gegliedert. Betroffene sind nicht in der Lage Gelesenes zu wiederholen und aus dem Gelesenen Zusammenhänge zu erkennen, bzw. Schlüsse zu ziehen. Es bestehen also Defizite im Leseverständnis.

Die Symptomatik der Rechtschreibstörung äußert sich in Schwierigkeiten, Buchstaben und Wörter zu schreiben. Es werden Buchstaben oder Wortteile verdreht, ausgelassen oder falsch eingefügt. Buchstaben werden durch gleich klingende Buchstaben ersetzt. Es entstehen Fehler beim Anwenden von Regeln, wie z.B. Dehnungsfehler, Fehler in Groß- und Kleinschreibung und sog. "Wahrnehmungsfehler" (d-t, g-k usw. werden verwechselt). Die Fehler können dabei inkonstant sein: ein und dasselbe Wort wird auf unterschiedliche Arten falsch geschrieben.

Rechtschreibfehler treten insbesondere beim spontanen Schreiben oder in einem Diktat auf, während das Abschreiben von Wörtern meist fehlerlos erfolgt. Betroffene können ein Wort korrekt artikulieren, es aber trotzdem falsch schreiben. Sie sind selbst meist nicht in der Lage ihre Fehler zu erkennen und zu korrigieren. Betroffene mit höherer Intelligenz oder solche, die leicht auswendig lernen, können bis zu einem Grad der Anforderungen (z. B. bis zum Erreichen einer bestimmten Klassenstufe) ihre Lese- und



Rechtschreibstörung kompensieren. Im Vorschulalter können Schwierigkeiten bestehen Buchstaben korrekt zu benennen, Laute zu unterscheiden (gestörtes Lautbewusstsein), einfache Wortreime zu bilden und das Alphabet aufzusagen.

Nach ICD-10 beinhaltet die Diagnose von umschriebenen Entwicklungsstörungen schulischer Fertigkeiten (F 81.-), zu denen die Lese- und Rechtschreibstörung (F 81.0) und die Isolierte Rechtschreibstörung (F 81.1) gehören, das Vorhandensein einer „umschriebenen und bedeutsamen Beeinträchtigung“ in der Entwicklung schulischer Leistungen (wie z. B. Lese- und Rechtschreibfertigkeiten), die aufgrund von Intelligenz, Alter und Beschulung nicht erklärbar sind. Begleitende Störungen im emotionalen und Verhaltensbereich sind häufig. Um die Diagnosen der Lese- und Rechtschreibstörung (F 81.0), die dem in Deutschland verbreiteten Störungsbegriff „Legasthenie“ entspricht, und der Isolierten Rechtschreibstörung (F 81.1) stellen zu können, sind desweiteren neurologische und emotionale Störungen, periphere Seh- und Hörstörungen und unangemessene Beschulung, bzw. mangelnde Lerngelegenheit, auszuschließen, wenn diese als ursächlich für die Schulleistungsstörungen angesehen werden können. Das DSM-IV hat im Vergleich ähnliche Kriterien für die Diagnose der Schulleistungsstörungen, worunter die Lesestörung (315.00) und die Störung des schriftlichen Ausdrucks (315.2) fallen. Allerdings werden neurologische und emotionale Erkrankungen als Ausschlusskriterium nicht erwähnt.

Das ICD-10 befürwortet somit eine Diskrepanz zwischen Lese-, bzw. Rechtschreibleistung und IQ, sowie eine Diskrepanz zwischen Rechtschreibleistung und Alter. Die hierbei geforderte „bedeutsame“ Diskrepanz zwischen schulischer Fertigkeit und Intelligenz wird im ICD-10 in den klinischen Kriterien nicht quantifiziert. Ebenso wird beispielsweise nicht genau definiert, wie groß die Beeinträchtigung der Leseleistung zusätzlich zu der Rechtschreibstörung sein müsste, um eine Lese-Rechtschreibstörung zu diagnostizieren. Es gibt Unstimmigkeiten darüber, inwiefern das IQ-Diskrepanzkriterium als valide bewertet werden kann (Übersicht in Deimel, 2002; S.120-123). Studien, die sich mit dieser Thematik auseinandersetzen, versuchen Unterschiede zwischen Gruppen zu finden, die diese Diskrepanz zwischen der

Intelligenz und der Lese-, bzw. Rechtschreibleistung erfüllen oder nicht. Die Gruppen unterscheiden sich also nicht in ihrer Lese- bzw. Rechtschreibleistung, wohl aber in ihrer Intelligenz. Probanden, die die Diskrepanz zwischen Intelligenz und Lese-, bzw. Rechtschreibleistung erfüllen, kennzeichnen sich durch eine spezifische Leistungsschwäche (SLS), während Probanden der anderen Gruppe mit schlechter Lese- bzw. Rechtschreibleistung ohne IQ-Diskrepanz als allgemein leistungsschwach (Allgemeine Leistungsschwäche, ALS) gelten. Ergebnisse dieser Untersuchungen sind unterschiedlich, auch je nachdem in welchen Eigenschaften die beiden Gruppen miteinander verglichen werden. Beispielsweise wurden Auffälligkeiten in kognitiven Tests (Meyer, 2000) und in Kleinhirnfunktionstests (Fawcett, Nicolson & Maclagan, 2001) in der Gruppe der schlechten Leser mit IQ-Diskrepanz (SLS) gefunden, während sich nach Shaywitz et al. (1992) sich die beiden Gruppen (SLS und ALS) nicht in ihren sprachlichen Fähigkeiten, ihrer visuellen Wahrnehmung und ihrer Händigkeit unterschieden. In den meisten Studien werden keine Unterschiede zwischen beiden Gruppen gefunden (Deimel, 2002). Aaron (1997) ist der Meinung, dass es keine Hinweise auf Unterschiede zwischen den Gruppen gibt, während Meyer (2000) diese Thematik als ungelöst ansieht.

Ich gehe im methodischen Teil dieser Arbeit näher darauf ein, welche Konsequenzen sich für die Auswahlkriterien der am AGL-Experiment teilnehmenden Kinder für diese Untersuchung ergeben, und wie diese hinsichtlich der Kriterien für eine Lese-Rechtschreibstörung und hinsichtlich der eben erläuterten Diskussion um den IQ umgesetzt werden.

Ich werde in den nun folgenden Abschnitten erläutern, inwiefern ein AGL-Experiment dazu geeignet ist, als Methode im Laborexperiment, Rückschlüsse auf das implizite Lernen von orthographischen Fragmentwissen und auf dessen Relevanz für den Schriftspracherwerb bei Kindern mit einer Rechtschreibstörung im Vergleich von nicht betroffenen Kindern zu ziehen.

## **2.3 Die Bedeutung des orthographischen Fragmentwissens für den Schriftspracherwerb**

### **2.3.1 Schriftspracherwerb – Begriffe des orthographischen Wissens und der phonologischen Bewusstheit**

Der Schriftspracherwerb ist ein Prozess des Abstrahierens von phonologischen und orthographischen Informationen von Schriftbild und Sprache und des Nutzens dieser Kenntnisse zur fehlerfreien und beständigen Rechtschreibung (Steffler 2001). Die hier vorliegende Untersuchung bezieht sich auf den Bereich des visuellen orthographischen Wissens.

Orthographisches Wissen wird als Fähigkeit bezeichnet, visuelle Wortbilder zu erkennen, und beinhaltet das Wissen von orthographischen Regelmäßigkeiten (Schulte-Körne, 2001). Phonologische Bewusstheit ist die Fähigkeit des Erkennens und der Anwendung phonologischer Informationen in Schriftbild und Sprache. Als übergeordneter Begriff beinhaltet die phonologische Bewusstheit mehrere Prozesse in der Lautanalyse, die sich bezüglich kognitiver Leistungen und Verarbeitungen unterschiedlich komplexer Strukturen (z. B.: Silbe, Phonem) unterscheiden (Schulte-Körne, 2001).

Beide dieser Fähigkeiten, das orthographische Wissen und die phonologische Bewusstheit, sind grundlegend für den Erwerb von Lese- und Rechtschreibfertigkeiten. Im Folgenden gehe ich kurz darauf ein, welche Bedeutung das orthographische Wissen und die phonologische Bewusstheit für einen gestörten Erwerb von Lese- und Rechtschreibfertigkeiten hat.

Nach Lennox & Siegel (1996) haben Kinder mit einer unterdurchschnittlichen Rechtschreibleistung im Vergleich zu Kindern mit einer durchschnittlichen Rechtschreibleistung, einen größeren Erfolg in dem Nutzen von orthographischen / visuellen Strategien als in dem Nutzen von phonologischen Strategien beim Schreiben. Sie haben anscheinend mehr Schwierigkeiten mit Phonem-Graphem-Zuordnungen als normale Kinder. Kinder mit einer unterdurchschnittlichen Rechtschreibleistung weisen

auch eine andere Entwicklung im Schriftspracherwerb auf als normale Kinder (Lennox & Siegel, 1996). Bereits Rack (1985) beschrieb diesen erschwerten Zugang der Leserechtschreibgestörten zu dem phonologischen Kodieren insbesondere bei visueller Präsentation, während sie allerdings in der Lage zu sein scheinen, dies mittels eines gesteigerten Nutzens des visuellen / orthographischen Kodierens zu kompensieren. Auch Schulte-Körne et al. (1997) unterstützen die große Bedeutsamkeit der phonologischen Dekodierfähigkeit für die Rechtschreibfähigkeit.

Ich stelle nun kurz einige ausgewählte Entwicklungsmodelle zur Lese- und Rechtschreibfähigkeit vor, um später auf ihre Relevanz, insbesondere im Forschungsbereich des orthographischen Wissens, zurückzukommen.

### **2.3.1.1 Modelle zur Lese- und Rechtschreibentwicklung**

In dem klassischen Zwei-Wege-Modell nach Coltheart (1978) wird beim Leseprozess der direkte lexikalische Weg von dem indirekten, phonologischen Weg unterschieden. Der direkte Weg ist ein ausschließlich visueller Prozess, der keine phonologische Analyse benötigt, sondern die Erkennung des Wortes durch das Abrufen von Gedächtnisrepräsentationen erfolgt. Der indirekte, phonologische Weg beruht auf Phonem-Graphem-Zuordnungen<sup>5</sup>.

Es gibt auch Ansichten über eine phasenartige Entwicklung des Schriftspracherwerbs, die annimmt, dass zu Beginn der Entwicklung des Schriftspracherwerbs Schreibanfänger in der semiphonetischen und phonetischen Phase erstmal primär Phonem-Graphem-Zuordnungen und anfangs die Aussprache von Buchstabennamen (d. h. semiphonetisch) anwenden, ohne auf Reihenfolge oder Regelmäßigkeiten der Buchstaben zu achten (Ehri, 1986). Eine ähnliche Ansicht, dass orthographische Regeln erst in späteren Phasen, bzw. höheren Stufen, beim Rechtschreiben verwendet werden, vertrat Frith (1985) in dem Drei-Stufen-Modell. Frith (1985) unterteilt den Lese- und Rechtschreiberwerb in eine logographische, alphabetische und orthographische Stufe. In der logographischen Stufe werden nur bekannte Wörter anhand auffälliger graphischer

---

<sup>5</sup> Als Phoneme werden die kleinsten, sequentiellen lautlichen Bestandteile eines Wortes bezeichnet. Phoneme zeichnen sich durch ihre Funktion, Unterschiede in der Bedeutung von Wörtern zu signalisieren, aus. Sie werden im Schriftbild von einer geringeren Anzahl von Graphemen repräsentiert (Schulte-Körne, 2001).

Merkmale erkannt, ohne dass eine Lautanalyse und- synthese stattfindet. Diese auffälligen graphischen Merkmale bestimmen auch die Schreibweise des Wortes. Die alphabetische Stufe zeichnet sich durch Wissen über Grapheme und Phoneme und deren Zusammenhänge aus, so dass nun unbekannte Wörter gelesen werden können. Beim Schreiben wird nur der Wortklang ohne Beachtung von Regeln umgesetzt. Erst in der orthographischen Stufe werden Wörter direkt anhand orthographischer Merkmale ohne phonologische Prozesse erkannt und beim Schreiben orthographische Regeln verwendet.

Zunächst erläutere ich, inwiefern Annahmen über die inhaltliche Struktur des orthographischen Wissens und der Wissenserwerb in AGL-Experimenten in Zusammenhang stehen. Später werde ich auf Erkenntnisse aus der Forschung im Bereich des orthographischen Wissens eingehen, die insbesondere die eben erläuterte Annahme von Ehri (1986) und Frith (1985), dass Wissen über orthographische Regeln im Prozess des Schriftspracherwerbs erst spät erworben und angewendet werden, in Frage stellen.

### **2.3.1.2 Zusammenhang zwischen dem orthographischen Wissen und der AGL-Forschung**

Die Entwicklung des orthographischen Wissens ist im Vergleich zu der des phonologischen Wissens nicht so gründlich untersucht worden, weil es in typischen Rechtschreibaufgaben schwierig ist, phonologische und orthographische Informationen überhaupt voneinander zu trennen (Steffler 2001).

Es gibt mehrere Annahmen über die inhaltliche Struktur des orthographischen Wissens (Übersicht in Steffler 2001, 178-182). Steffler (2001) sieht in diesen Ansätzen einen klaren Zusammenhang zu den theoretischen Annahmen in dem AGL-Forschungsbereich des impliziten Lernens, welchen ich nun erläutere.

Eine Annahme vermutet, dass orthographisches Wissen als abstraktes Regelwissen basierend auf sprachlichen Mustern und Regelmäßigkeiten vertreten ist. In AGL-Experimenten wird, damit vergleichbar, der Erwerb von abstraktem Regelwissen von Reber (1967,1989) befürwortet. Es wird auch eine Annahme über ganzes-Wort-Wissen

(wortspezifisches Gedächtnis genannt) geäußert. Diesbezüglich wird in der AGL-Forschung von Brooks et al. (1978, 1992,1992) der Erwerb von Wissen über spezifische, komplette Items vermutet. Abschließend beinhaltet eine Annahme, dass Wissen über Fragmenthäufigkeiten, bzw. Wahrscheinlichkeiten der Zusammengehörigkeit von Buchstabenkombinationen, orthographisches Wissen vertritt. Dies ist vergleichbar mit den Vermutungen von Perruchet & Pacteau (1990), Redington & Charter (1996) und Servan- Schreiber & Anderson (1990), dass Fragmentwissen in AGL-Experimenten erworben wird.

Hinsichtlich dieses Zusammenhangs zwischen den Annahmen zu Wissen über Fragmenthäufigkeiten als Vertreter des orthographischen Wissens und Fragmentwissen, welches in AGL-Experimenten erworben wird, liegt eine Bedeutung von orthographischen Fragmentwissen für den Schriftspracherwerb nahe, auf welche ich nun eingehen werde.

### **2.3.2 Die Bedeutung von orthographischen Fragmentwissen für den Schriftspracherwerb und sein Zusammenhang mit der AGL-Forschung**

Um die Bedeutung des impliziten orthographischen Fragmentwissens auf den Schriftspracherwerb zu erläutern, gehe ich im Folgenden näher auf die Untersuchungen von Aaron, Wilczynski & Keetay (1998), Cassar & Treiman (1997) und Pacton et al. (2001) ein.

Bereits Spoehr & Smith (1975) haben herausgefunden, dass reguläre Buchstabensequenzen, die als in einer Sprache typischen Kombinationen von Buchstaben definiert sind, leichter zu erkennen sind, als irreguläre. Aaron, Wilczynski & Keetay (1998) beziehen dies auf einen regelgeleiteten stochastischen Prozess („rule-governed stochastic process“). Sie nahmen an, dass die Abfolge von Buchstaben in einem Wort eine systematische Beziehung untereinander hat, wie die einzelnen Wörter als Satzglieder in einem Satz. Wenn also z. B. der Buchstabe „q“ auftaucht, folgt immer ein „u“ nach, usw.. Sie untersuchten dabei anhand dem Vergleich einer Kontrollgruppe zu tauben Probanden, inwiefern das wortspezifische Gedächtnis ein Gedächtnis für

Wörter als Ganzes ist, oder ein Gedächtnis bezüglich des regelgeleiteten stochastischen Prozesses für häufige, innerhalb eines Wortes vorkommende Buchstabenschemata reflektiert. Sie benutzten hierfür nichtaussprechbare und aussprechbare Nichtwörter. Die nichtaussprechbaren Nichtwörter bestanden aus in der englischen Sprache seltenen Buchstabenkombinationen, während die aussprechbaren Wörter aus im Englischen häufig vorkommenden Bi- und Trigrammen zusammengesetzt waren. Die Forscher gingen von der Hypothese aus, dass die Tauben, die nicht von den phonologischen Merkmalen der Wörter profitieren können, beide Arten von Nichtwörter gleich gut schreiben könnten – wenn für die Rechtschreibung das Ganze-Wort-Wissen ausschlaggebend sei. Beide Gruppen konnten jedoch mehr aussprechbare Nichtwörter korrekter reproduzieren, so dass Aaron, Wilczynski & Keetay (1998) dies als Beweis für das Gedächtnis eines regelgeleiteten stochastischen Prozesses sahen. Sie schlugen vor, dass die regelgeleitete Komponente des wortspezifischen Gedächtnisses wahrscheinlich als Wissen von Buchstabeneinheiten in Bi- und Trigrammen vertreten ist, welches von der Häufigkeit ihres Vorkommens in der Schriftsprache bedingt wird.

Dieses Wissen wird im weiteren „orthographisches Fragmentwissen“ genannt.

Von zentraler Bedeutung von implizitem Lernen und orthographischem Fragmentwissen für den Schriftspracherwerb im Bereich der normalen Entwicklung sind die Ergebnisse der Studien von Cassar & Treiman (1997) und Pacton et al. (2001).

Cassar & Treiman (1997) untersuchten die Entwicklung des kindlichen Wissens über orthographische Regelmäßigkeiten in der englischen Schriftsprache. Wie bereits anhand früherer Untersuchungen vermutet wurde, konnten sie zeigen, dass Kinder in ihrer Entwicklung schon früher als vorher angenommen anfangen, sich orthographisches Wissen anzueignen und zwar schon bevor sie über den Buchstaben-Laut-Zusammenhang unterrichtet wurden. Diese Kinder zeigten bei einer Auswahl zwischen zwei Alternativen eines erfundenen Wortes („nnus – nuss“) eine Sensitivität gegenüber verdoppelten Buchstaben, an welcher Stelle des Wortes diese dem üblichen Gebrauch von Verdopplungen in der englischen Sprache folgend auftreten dürfen (z. B. am Ende) oder nicht (z. B. am Anfang). Cassar & Treiman (1997) erbrachten zusätzlich den Nachweis, dass auch Kinder, die bereits der phonetischen Phase von Ehri (1986)

zuzuordnen sind, sich nicht nur auf Phonem-Graphem-Zuordnungen berufen, sondern auch Wissen über einfache orthographische Regelmäßigkeiten haben. Damit bezweifelten Cassar & Treiman (1997) die traditionelle Ansicht, dass orthographische Regeln erst in späteren Phasen (Ehri, 1986) , bzw. höheren Stufen (Frith 1985), des Schriftspracherwerbs verwendet werden (siehe 2.3.1.1).

Pacton et al. (2001) stellten fest, dass Cassar & Treiman (1997) nie Bezug auf allgemeine Lerntheorien (z. B. implizites Lernen) in ihrem Artikel genommen hatten und dies ein auffallender Mangel auch in anderen Studien, die speziell konstruiert wurden um Zusammenhänge zwischen impliziten Lernen und Alltagslernen zu untersuchen, ist. Sie nahmen sich vor, diese Lücke zwischen experimenteller Laborforschung und natürlichen Lernsituationen zu schließen. Nach Meinung von Pacton et al. (2001) ist implizites Lernen im Sprach- und Schriftspracherwerb involviert. Da Spracherwerb allerdings in einem Entwicklungszeitraum erfolgt, der besonders herausfordernd für empirische Untersuchungen ist, und zusätzlich auch die Möglichkeit besteht, dass teilweise ein instruierter Lernprozess stattfindet, konzentrierten sich Pacton et al. (2001) auf besser einzugrenzende Fähigkeiten im Schriftspracherwerb, die zuverlässig noch nicht explizit gelehrt wurden. Sie untersuchten die Entwicklung des impliziten Wissens über bestimmte orthographische Regelmäßigkeiten basierend auf nicht explizit unterrichteten Erfahrungen über die französische Schriftsprache von in diesem Sprachraum aufgewachsenen Kindern. Die Ergebnisse der Experimente von Pacton et al. bestätigten und erweiterten die Erkenntnisse aus der früheren Untersuchung (Cassar & Treiman, 1997). Französischsprachige Erstklässler, die bisher kaum Unterricht im Schriftspracherwerb erhalten hatten, zeigten eine Sensitivität gegenüber der Häufigkeit von bestimmten Doppelkonsonanten. Die Kinder lehnten erfundene Wörter mit in der französischen Schriftsprache nicht vorkommenden Doppelkonsonanten eher ab, als zur Alternative stehende erfundene Wörter mit erlaubten Doppelkonsonanten, wobei alle Buchstaben generell annähernd gleich häufig als Einzelbuchstaben auftraten. Vorteil der französischen Sprache im Vergleich zur englischen Sprache und damit im Vergleich zur Untersuchung von Cassar & Treiman (1997) war, dass sich die Verwendung von Doppelkonsonanten im Französischen nicht auf die Aussprache der betroffenen Wörter



auswirkt (außer die Verdopplung „ss“, die aber auch nicht verwendet wurde), und damit phonologische Faktoren dieses Wissen nicht erklären konnten. Zusätzlich waren die Kinder sensitiv gegenüber der Position von Verdopplungen in den erfundenen Wörtern (die einzige legale Position in der Mitte des Wortes wurde am meisten akzeptiert) und der Tatsache, dass Vokale im Französischen nie verdoppelt werden und zeigten in gewisser Hinsicht auch ein abstraktes orthographisches Wissen bei Wörtern, die nur illegale Doppelkonsonanten an unterschiedlichen Positionen enthielten. Pacton et al. (2001) verglichen dieses Versuchsdesign mit dem üblichen Aufbau von Transferexperimenten in AGL, in denen sich Lern- und Teststimuli in der äußeren Struktur unterscheiden aber die gleiche unterschwellige abstrakte Struktur haben.

Insgesamt bestätigten auch Pacton et al. (2001), dass implizites Lernen von orthographische Fragmentwissen den Schriftspracherwerb von Kindern früher beeinflusst und dies früher geschieht als Entwicklungsmodelle (Pacton et al., 2001, 416) voraussagen würde.

Mit Hilfe des AGL-Experimentes ist es möglich, das Lernen von Fragmentwissen unter Laborbedingungen zu untersuchen. Desweiteren kann auch überprüft werden, inwiefern eine regelgeleitete Komponente von den Kindern bei der Auseinandersetzung mit schriftsprachlichem Material gelernt wird. Das Lernen kompletter Items im AGL-Experiment wird wie das Lernen kompletter Wörter, also die Annahme eines wortspezifischen Gedächtnisses von Aaron, Wilczynski & Keetay (1998) und den meisten AGL-Forschern eher als unwahrscheinlich angenommen. Unter diesem Aspekt kann das durch implizite Lernen erworbene Fragmentwissen für den von Coltheart (1978) in seinem Zwei-Wege-Modell beschriebenen direkten, lexikalischen Weg, bei dem auf der Basis eines visuellen Prozesses die Erkennung des Wortes durch das Abrufen von Gedächtnisrepräsentationen erfolgt, von Bedeutung sein. Das Einprägen kompletter Wörter ist nach Aaron, Wilczynski & Keetay (1998) aufgrund der enormen Kapazität, die dafür erforderlich wäre, unökonomisch. Außerdem kann durch das Lernen kompletter Wörter nach Aaron, Wilczynski & Keetay (1998) nicht die Fähigkeit, neue Wörter oder Pseudowörter zu schreiben, erklärt werden. Es kann auch nicht erläutern, warum mehr reguläre Wörter als irreguläre Wörter richtig geschrieben

werden. Diese Art von erworbenem Wissen spielt also nach Aaron et al. (1998) im Schriftspracherwerb, wie auch nach den meisten AGL-Forschern im Wissenserwerb in AGL-Experimenten, nur eine untergeordnete Rolle.

Wahrscheinlicher ist, dass eine Kombination von regelgeleiteten und ähnlichkeitsbasierenden Prozessen stattfindet, die zu einem Wissenserwerb in Schriftsprache und im AGL-Experiment führt. Pacton et al. (2001) vertraten die Meinung, dass Annahmen über regelbasierendes Wissen von dem Phänomen der verminderten Klassifikationsleistung der Probanden in Transfertests herausgefordert werden ("transfer decrement", Dienes & Altmann, 1997; Gomez, 1997; Mathews et al. 1989; Shanks, Johnstone & Staggs, 1997; Whittlesia & Wright, 1997). Pacton et al. argumentieren, dass unter der Annahme, dass nur abstraktes Regelwissen beteiligt wäre, diese schlechtere Klassifikationsleistung sich möglicherweise dann am ehesten durch eine inkomplette Abstraktion der Regeln erklären ließe, da zusätzlich die Entwicklung dieses abstrakten Wissens erheblich mehr Zeit in Anspruch nimmt, als normalerweise in den Experimenten zur Verfügung steht. In ihren eigenen Experimenten zeigten sie jedoch, dass trotz intensiven Lernens über einen Zeitraum von fünf Jahren die verminderte Transferleistung persistierte. Sie vermuteten deswegen, dass selbst nach einer beachtlichen Vertrautheit mit Schriftmaterial, das orthographische Verhalten der Kinder nicht bereitwillig als alleine regelbasierend, in dem Sinne wie vorher definiert, qualifiziert werden kann. Pacton et al. (2001) bestätigten diese Vermutung durch erfolgreiches Simulieren ihrer Ergebnisse mit SRN („simple recurrent network“), welches mit den häufigsten 1000 französischen Wörtern trainiert wurde. Das Computermodell – ein einfaches Verknüpfungsmodell – rechnet allein auf Basis von einfachen Assoziationen zwischen Input-Items, und verwendet somit keinerlei regelbasierende Abstraktionen. Die Analyse mit diesem Modell unterstützte deutlich die Annahme, dass abstraktes Regelwissen nicht nötig ist, um die Leistungen der Kinder zu erklären. Es kann allerdings nach Pacton et al. (2001) nicht ausgeschlossen werden, dass eine regelgeleitete Komponente mit beteiligt ist.

Zusammenfassend befürworteten Aaron, Wilczynski & Keetay (1998) und Pacton et al. (2001) eher das implizite Lernen orthographischen Fragmentwissens im

Schriftspracherwerb, wie auch bei den Klassifikationsentscheidungen in einem AGL-Experiment eher Fragmentwissen von Bedeutung zu sein scheint (z. B. nach Kinder & Assmann, 2000), als das Lernen komplett auf Basis von abstrakten Regeln. Es ist aber möglich, dass in diesem impliziten Lernprozess auch ein abstrakterer regelbasierender Prozess mitbeteiligt ist. Insgesamt kann also ein AGL-Experiment als eine geeignete Untersuchungsmethode zur Erfassung des impliziten Lernens von orthographischen Fragmentwissens mit oder ohne einer regelgeleiteten Komponente im Labor angesehen werden.

Im Folgenden stelle ich nun Untersuchungen vor, in denen das implizite Lernen in Bezug auf die Lese-Rechtschreibstörung erforscht wurde. Mir ist keine Untersuchung aus dem Forschungsbereich der Lese- und Rechtschreibstörung bekannt, die das implizite orthographische Fragmentwissen untersuchte.

### **2.3.3 Implizites Lernen in Zusammenhang zur Lese- und Rechtschreibstörung**

Es gibt einige Untersuchungen, die das implizite Lernen im Zusammenhang mit der Lese-Rechtschreibstörung untersuchten, allerdings mit anderen Experimenten und nicht anhand des herkömmlichen AGL-Experimentes mit Buchstabenmaterial. Diese widersprechen sich untereinander teilweise in ihren Ergebnissen.

Vicari, Marotta, Meneghini, Molinari & Petrosini (2003) untersuchten implizites Lernen von Lese-Rechtschreibgestörten (Durchschnittsalter: 10 Jahre) anhand eines SRT ("serial reaction time task"). In diesem Test wurden den Probanden auf einem Computerbildschirm einzeln eine Reihe von farbigen Kreisen (grün, blau und rot) gezeigt, die nach Zufall in verschiedenen Zeitabständen gezeigt wurden (zwischen 0,5 und 2 sec). Dadurch sollte vermieden werden, dass, wenn das Target (grünen Kreise) immer in konstanten Zeitintervallen erscheinen würde, eher ein Rhythmus als eine Sequenz gelernt werden würde. Die Probanden wurden dazu aufgefordert, sich die Kreise anzuschauen und immer bei Erscheinen eines grünen Kreises so schnell wie möglich die Leertaste zu drücken. Ein Computerprogramm erfasste Reaktionszeiten und Fehler der Probanden. Vicari et al. fanden heraus, dass die Lese-Rechtschreibgestörten

im Vergleich zur Kontrollgruppe im Verlauf des Tests langsamer reagierten. Im Gegensatz zur Kontrollgruppe, die bei jedem Durchgang schneller reagierte, wurden die Reaktionszeiten der Lese-Rechtschreibgestörten sogar größer. Auch wenn später eine Reihenfolge, in der die farbigen Kreise erscheinen, explizit gelernt wurde, reagierten die Kontrollkinder schneller als die betroffenen Kinder, obwohl allen Kindern nicht aufgefallen war, dass im letzten Durchgang die gelernte Reihenfolge wieder verwendet wurde. Schlussfolgerungen von Vicari et al. waren, dass Lese-Rechtschreibgestörte beeinträchtigt sind, eine Reihenfolge zu erkennen und implizites Wissen über diese anzuwenden. Weil keines der Kinder Regelmäßigkeiten bemerkt hatte, schlossen Vicari et al., dass kein explizites Wissen erworben wurde.

Waber et al. (2003) untersuchten ebenso mit Hilfe eines SRT Experimentes Kinder hinsichtlich Lesefähigkeit, Intelligenz und Aufmerksamkeitsproblemen. Sie konnten allerdings keinen Zusammenhang zwischen Lesefertigkeit oder der Intelligenz und dem impliziten Wissenserwerb demonstrieren. Es konnte nur gezeigt werden, dass allein die Intelligenz – nicht die Lesefähigkeit oder Aufmerksamkeitsdefizite – die Reaktionszeit der Antworten und explizites Lernen beeinflusst, die Fehlerfreiheit der Antworten allerdings von Intelligenz, Lesefähigkeit und Aufmerksamkeitsproblemen beeinflusst wurde. D.P. Waber et al. (2003) vermuteten, dass auf dieser grundlegenden und wahrscheinlich automatischen Ebene das implizite Lernen während dieses Experimentes gleich gut bei Kindern mit und ohne Leseproblemen zu sein scheint. Probleme und damit festzustellende Unterschiede zwischen den beiden Gruppen beim impliziten Lernen könnten ihrer Meinung nach eventuell eher bei der Verarbeitung höherer Ordnung von Symbolen und linguistischem Material auftreten.

Pothos & Kirk (2004) untersuchten anhand eines AGL-Experimentes, indem sie statt Buchstabenmaterial geometrische Symbole (Kreis, Quadrat, Sechseck und Raute) verwendeten und als unterschiedliche Stimuli in zwei AGL-Experimenten lese-rechtschreibgestörten Erwachsenen und Kontrollpersonen präsentierten. Die Stimuli waren in dem einen Experiment, wie normale Buchstabenitems aufeinanderfolgend (aufeinanderfolgende Stimuli, „sequenced stimuli“), also beispielsweise statt „VJ“ „○□“. Im anderen Experiment wurden die Symbole jedoch ineinander als ein Einheit

(eingebettete Stimuli, „embedded stimuli“) präsentiert, also statt „□“ wurde der Kreis in dem Quadrat gezeigt. Pothos & Kirk (2004) „übersetzten“ dabei mit Hilfe der entsprechenden Grammatik die gleichen Items, die Knowlton & Squire (1996, Experiment 1) entwickelt hatten. Die Ergebnisse der Untersuchung von Pothos & Kirk (2004) zeigten, dass Lese-Rechtschreibgestörte beide Arten von Stimuli (eingebettete und aufeinanderfolgende) gleich gut klassifizierten, wohingegen die Kontrollgruppe die eingebetteten Stimuli gleich gut klassifizierten wie die Lese-Rechtschreibgestörten, die aufeinanderfolgende Stimuli jedoch nach Auffassung von Pothos & Kirk wesentlich (um 5%) schlechter klassifizierte. Pothos & Kirk stellten für diesen Unterschied die Vermutung auf, dass die Kontrollgruppe bei der Konfrontation mit aufeinanderfolgenden Stimuli primär die einzelnen aufeinanderfolgenden Elemente verarbeiteten, und so – wahrscheinlich durch explizites Bemühen, die Struktur der Stimuli zu verstehen – das Lernen hemmten. Pothos & Kirk vermuteten, unterstützt von Feststellungen von Facchetti & Molteni (2001), dass Lese-Rechtschreibgestörte weniger dazu neigen, individuelle Elemente der aufeinanderfolgenden Stimuli, sondern Stimuli eher als Ganzes (wie eingebettete Stimuli) zu verarbeiten. Pothos & Kirk erwarteten, dass der Lernprozess der Lese-Rechtschreibgestörten in beiden Experimenten primär implizit ist, wohingegen das Lernen der eingebettete Stimuli bei der Kontrollgruppe primär explizit sei, aber für die aufeinanderfolgenden Stimuli eine Mischung aus impliziten Lernen und explizitem Bemühen, die Stimuli zu verstehen, vorliegt. Anstatt Vermutungen aufzustellen, ob Lese-Rechtschreibgestörte nun im impliziten Lernen beeinträchtigt sind, schlossen Pothos & Kirk aus ihrer Untersuchung, dass Lese-Rechtschreibgestörte in manchen Situationen davon abgehalten werden explizite Strategien einzusetzen, weil sie nicht in der Lage sind individuelle Elemente der Stimuli zu verarbeiten.

Anhand dieser beschriebenen Untersuchungen wird deutlich, dass sich die aktuelle Forschung mit der Fragestellung beschäftigt, inwiefern implizites Lernen bei Lese-Rechtschreibgestörten im Vergleich zu Nichtbetroffenen beeinträchtigt ist. Die letzten beiden erläuterten Untersuchungen konnten hier keine Beeinträchtigung nachweisen,

wobei in der Untersuchung von Waber (2003) Kinder nur hinsichtlich der Lesefähigkeit, aber nicht der Rechtschreibfähigkeit untersucht wurden. Waber et al. (2003) deuten an, dass möglicherweise eine Beeinträchtigung beim Verwenden von linguistischem Material festzustellen ist. In der hier vorliegende Untersuchung wird ein Experiment durchgeführt, mit dessen Hilfe festgestellt werden kann, ob Kinder mit und ohne einer Rechtschreibstörung im impliziten Umgang mit schriftsprachlichem Material Fragmentwissen mit oder ohne eine regelgeleitete Komponente unter Laborbedingungen erlernen. Dies steht im Gegensatz zu dem einfacheren Design des SRT, in dem implizites Lernen nur anhand Reaktionszeiten und Fehlantworten bezüglich von Symbolen und Farben untersucht wurde, und auch im Gegensatz zu dem AGL-Experiment von Pothos & Kirk (2004), in dem nur Symbole verwendet wurden. Durch das Untersuchen rechtschreibgestörter Kinder im Vergleich zu nicht rechtschreibgestörten Kindern können Rückschlüsse auf das Störungsbild der Isolierten Rechtschreibstörung und der Lese- und Rechtschreibstörung gezogen werden. Damit trägt die hier vorliegende Untersuchung zur Ursachenforschung der Lese- und Rechtschreibstörung bei.

Im weiteren gehe ich nun auf Zielsetzungen und Hypothesen der vorliegenden Untersuchung ein.

## **3 Zielsetzung und Hypothesen**

### **3.1 Zielsetzung**

Zusammengefasst ist das AGL-Experiment eine geeignete Versuchsanordnung zur Erfassung des impliziten Lernens und der visuell orthographischen Verarbeitung von schriftsprachlichem Material von Kindern im Laborexperiment.

Das hierbei erworbene Wissen bezieht sich auf Buchstabensequenzen einer künstlichen Grammatik. Es ergibt sich daraus die Hypothese, dass bei der Verarbeitung von schriftsprachlichem Material von einer Grammatik vorgegebene Eigenschaften und visuelle orthographische Informationen über die Häufigkeitsverteilung von Buchstaben und Fragmenten implizit gelernt werden. Dieses implizite Wissen fließt in die Verarbeitung von neuem Material ein, ohne dass die Anwendung dieses Wissens bewusst wird. Da anzunehmen ist, dass dieser implizite Lernvorgang und das Anwenden des erworbenen impliziten Wissens auch für den Schriftspracherwerb bedeutend ist, ist das Ziel dieser Untersuchung herauszufinden, inwiefern Kinder mit und ohne einer Rechtschreibstörung in der Versuchsanordnung eines AGL-Experimentes implizit lernen und dieses implizit erworbene Wissen auf neues Material anwenden. Dabei wird der Prozess des impliziten Lernens der Kinder mit und ohne eine Rechtschreibstörung im Querschnitt vergleichend untersucht. Zielgröße ist in dieser Untersuchung das Ergebnis der Testphase im AGL-Experiment als abhängige Variable, während die Rechtschreibstörung der Kinder, die Fragmentstärke und die Grammatikalität der Items die unabhängigen Variablen bilden.

Anhand der Literatur und eigenen Überlegungen lassen sich folgende Hypothesen und Fragestellungen ableiten.

### **3.2 Hypothesen**

1. Die nicht von einer Rechtschreibstörung betroffenen Kinder haben eine bessere Klassifikationsleistung als die betroffenen Kinder.

2. Die Testitems mit hoher Fragmentstärke werden eher als zugehörig klassifiziert als Items mit niedriger Fragmentstärke.
3. Testitems, die die Regeln der Grammatik befolgen (= grammatische Items), werden eher als zugehörig klassifiziert als Items, die die Regeln der Grammatik verletzen.
4. Kinder, die in höheren Klassenstufen sind, haben bessere Klassifikationsleistungen als Kinder in unteren Klassenstufen.

Zu 1.) Aus eigenen Überlegungen wäre anzunehmen, dass die nicht von der Rechtschreibstörung betroffenen Kinder die Testitems besser klassifizieren, wenn das implizite Lernen die Rechtschreibstörung mit verursacht.

Zu 2. und 3.) Die Auswertung des Klassifikationsverhalten der Kinder kann bezüglich der Fragmentstärke nach hoch oder niedrig und der grammatischen Eigenschaft der Testitems nach grammatisch oder nichtgrammatisch in einer Vierfeldertafel (siehe Ergebnisse) veranschaulicht werden. Die Hypothesen über das Klassifikationsverhalten der Kinder basieren auf den Erkenntnissen der Untersuchungen von Knowlton & Squire (1994, 1996) und Kinder & Assmann (2000).

Für die nichtgrammatischen Items mit niedriger Fragmentstärke wird angenommen, dass sie eher als nicht zugehörig („nein, gehört nicht dazu“) klassifiziert werden und damit unter dem Zufallslevel von 50 % einsortiert werden. Während bei den grammatischen Items mit hoher Fragmentstärke davon auszugehen ist, dass sie eher als zugehörig („ja, gehört dazu“) klassifiziert werden und dadurch über dem Zufallslevel von 50 % liegen. Es ist dagegen schwierig vorauszusagen, wie die grammatischen Items mit niedriger Fragmentstärke, bzw. die nichtgrammatischen Items mit hoher Fragmentstärke klassifiziert werden, da angenommen wird, dass die beiden Eigenschaften Fragmentstärke und Grammatik die Klassifikation beeinflussen. So klassifizierten die Probanden im Experiment 2 b von Knowlton & Squire (1994) grammatische Items niedriger Fragmentstärke mit 60,4 % und nichtgrammatische Items hoher Fragmentstärke mit 44,8 % über und unter dem Zufallslevel von 50 %, während sie im Experiment 1 von Knowlton & Squire



(1996) grammatische Items niedriger Fragmentstärke mit 62,5 % und nichtgrammatische Items hoher Fragmentstärke mit 54,9 % nur über dem Zufallslevel klassifizierten.

Zu 4.) Da die älteren Kinder mehr expliziten und impliziten Kontakt zu Schriftmaterial hatten als jüngere Kinder – allein durch den Faktor Zeit – ist anzunehmen, dass ältere Kinder über ein umfangreicheres implizites Wissen über Schriftmaterial verfügen als jüngere Kinder. Ob jedoch das Erwerben und das Anwenden impliziten Wissens während eines AGL-Experimentes von älteren Kindern dem Erwerben und Anwenden impliziten Wissens von jüngeren Kindern überlegen ist, wurde noch nicht erforscht. In den bisher durchgeführten AGL-Experimenten ist die Klassifikationsleistung in Abhängigkeit von der Klassenstufe bzw. dem Alter noch nicht untersucht worden. Bezüglich des Klassifikationsverhaltens in AGL-Experimenten basiert diese Hypothese auf eigenen Überlegungen. Hinweise in der Literatur, dass bei impliziten Lernprozessen auch das Alter von Bedeutung ist, finden sich in der Untersuchung von Murphy, McKone & Slee (2002), die die allgemein bestehende Annahme widerlegten, dass die Leistungen des impliziten Gedächtnis in einer umfangreichen Alterspanne (ab 3 Jahren bis ins Erwachsenenalter) stabil bleiben und damit früh heranreift, während sich das explizite Gedächtnis bekanntermaßen bis in die Adoleszenz verbessert und sich vergleichsweise später entwickelt (z. B. Naito, 1990). Sie konnten zeigen, dass implizite Gedächtnisleistungen bis über das 10. Lebensjahr hinaus zunehmen.

### **3.3 Fragestellungen**

#### **Weitere Fragestellungen**

1. Welchen Einfluss hat die Leseleistung der Kinder auf die Leistungen der Kinder in der Testphase?

Zu 1.) Wenn die Leseleistung implizites Lernen der Kinder beeinflusst, vermute ich, dass Kinder mit einer guten Leseleistung besser klassifizieren als Kinder mit einer schlechteren Leseleistung.

## **Fragestellungen zu der Lernphase**

2. Klassifizieren Kinder, die viele Items in der Lernphase falsch nachschreiben, schlechter als Kinder mit weniger Fehlern?
3. Ist die Anzahl der Fehler bei Rechtschreibgestörten Kindern höher?
4. Hat das Alter der Kinder Einfluss auf die Anzahl der Fehler?

In AGL-Experimenten ist es üblich, dass die Probanden die Lernitems in der Lernphase mehrere Male wiederholen und nachschreiben. Üblicherweise werden falsch geschriebene Items öfter dargeboten als richtig geschriebene Items. Dadurch sieht ein Proband, der mehr Fehler macht, die Items öfter, als eine Person die weniger Fehler macht. Teilweise wird auch verlangt, dass die Items solange nachgeschrieben werden, bis jedes Item mindestens einmal richtig geschrieben wurde (z. B. Gomez 1997; Gomez, Gerken & Schvaneveldt, 2000). Diese Experimente wurden allerdings mit Erwachsenen durchgeführt. Bei Kindern ist die Gefahr groß, dass sie von den Anforderungen in der Lernphase überfordert werden. Die Lernphase dieses AGL-Experimentes wurde deswegen so konstruiert, dass die Kinder die Lernitems dreimal nachschrieben und sobald sie ein Item einmal oder mehrere Male falsch schrieben, daraufhin (nur) noch ein viertes Mal dargeboten bekamen (nähere Erläuterungen in 4.3.1). Ich vermute, dass Kinder mit einer Rechtschreibstörung und Kinder aus unteren Klassenstufen mehr Probleme haben die Items richtig nachzuschreiben, als Kinder ohne Rechtschreibstörung und aus höheren Klassenstufen. Mir ist kein AGL-Experiment auf Basis von schriftsprachlichem Material bekannt, das nur mit Kindern durchgeführt wurde. Die Frage, inwiefern die Anzahl der falsch geschriebenen Items einen Einfluss auf die Klassifikationsleistung in AGL-Experimenten hat, wurde in dieser Form – soweit mir bekannt ist – noch nicht überprüft.

## 4 Methode

Im folgenden gehe ich nun darauf ein, wie die vorliegende Untersuchung unter Betrachtung dieser eben erläuterten Zielsetzungen, Hypothesen und Fragestellungen methodisch umgesetzt wurde.

### 4.1 Stichprobe

An einer unausgelesenen Stichprobe von 264 Kindern der 1. bis 4. Klasse zweier hessischer Grundschulen wurden standardisierte Tests zur Erfassung der Rechtschreibleistung, Leseleistung und Intelligenz der Kinder durchgeführt.

Die Auswahl der am AGL-Experiment teilnehmenden Kinder erfolgte zunächst anhand der Leistungen in den Rechtschreibtests. Als Kriterium für die Aufnahme in die Gruppe der rechtschreibgestörten Kinder wurden Leistungen der Kinder mit mindestens einer Standardabweichung unterhalb des Mittelwertes ( $T\text{-Wert} \leq 40$ ) angesehen. Für die Aufnahme in die Kontrollgruppe waren mindestens durchschnittliche Leistungen ( $T\text{-Wert} \geq 50$ ) im Rechtschreibtest erforderlich. Die in dieser Untersuchung geforderten Rechtschreibfertigkeiten der Kinder von mindestens einer Standardabweichung unter dem Mittelwert ist als Maß für unterdurchschnittliche Leistungen gebräuchlich.

Jedes rechtschreibgestörte Kind wurde mit einem Kind aus der Kontrollgruppe anhand der Klassenstufe und der Intelligenz parallelisiert, wobei die Differenz zwischen den beiden T-Werten der Intelligenz nicht größer als 5 sein durfte.

Dadurch entstehen zwei Extremgruppen, die zusätzlich durch die Parallelisierung der Kinder nach IQ und Klassenstufe in weiteren Variablen kontrolliert werden. Durch das Erfassen des Intelligenzquotienten der Kinder kann gewährleistet werden, dass evtl. festzustellende Unterschiede im Gruppenvergleich auf einen Einfluss der Rechtschreibleistung und nicht auf Einflüsse intellektueller Fähigkeiten der Kinder zurückzuführen sind. Nach Reber et al. (1991) korreliert der Intelligenzquotient nicht hoch mit dem Klassifikationsverhalten in einem AGL-Experiment, also in Form einer impliziten Anforderung, im Gegensatz zu der Leistungsfähigkeit in einer expliziten

Aufgabenstellung, die hoch mit dem Intelligenzquotienten korreliert. Durch die Parallelisierung der Kinder wird gewährleistet, dass der Gruppenvergleich nicht durch den IQ und die Klassenstufe verzerrt wird. Durch diesen Extremgruppenansatz und das dadurch bedingte Nichterfassen des mittleren Bereiches in der Rechtschreibleistung (T-Werte im Rechtschreibtest zwischen 40 bis 50) können Gruppenunterschiede besser erfasst werden.

Es ist einzuräumen, dass durch diesen Extremgruppenansatz, mit dessen Hilfe Einflüsse von IQ und Alter im Voraus kontrolliert werden, in den Auswahlkriterien nicht das IQ-Diskrepanzkriterium, welches die ICD-10 befürwortet, erfüllt wird. Da dieses allerdings auch umstritten ist (nähere Erläuterungen, siehe 2.2), wurde im Rahmen dieser Untersuchung entschieden die Intelligenz, ebenso wie die Klassenstufe, durch Parallelisierung der Kinder nach IQ und Klassenstufe zu kontrollieren, damit der Gruppenvergleich nicht durch diese beiden Größen verzerrt wird, und um die Ergebnisse des AGL-Experimentes uneingeschränkter auf die Rechtschreibfertigkeit zurückführen zu können.

Anhand der Leistungen in den Rechtschreibtest wurden somit zunächst 20 Kinder mit schlechter Rechtschreibung (T-Wert  $\leq 40$ ) ausgewählt. Zusätzlich wurden sieben rechtschreibgestörte Kinder aus einer Familienuntersuchung zur Genetik der Rechtschreibstörung der Stichprobe hinzugefügt. Die rechtschreibgestörten Kinder wurden mit 27 Kindern mit guter Rechtschreibung (T-Wert  $\geq 50$ ) unter Berücksichtigung der Intelligenz und der Klassenstufe, wie beschrieben, parallelisiert. Die Gruppe der rechtschreibgestörten Kinder und die Kontrollgruppe – Schulkinder mit guter Rechtschreibleistung – umfassten also jeweils 27 Kinder. Voraussetzung für die Teilnahme war, dass die Kinder nicht zweisprachig aufwuchsen und ihre Muttersprache deutsch war.

Die mittlere Rechtschreibleistung der rechtschreibgestörten Gruppe ( $n = 27$ ) lag somit bei 36,07 T-Werten ( $\pm 4,29$  Standardabweichung), die mittlere Rechtschreibleistung der Kontrollgruppe ( $n = 27$ ) betrug 55,78 T-Werte ( $\pm 5,79$  Standardabweichung). Der mittlere T-Wert der Intelligenz der rechtschreibgestörten Kindern war 50,96 ( $\pm 5,67$

Standardabweichung), der mittlere T-Wert der Intelligenz der Kontrollgruppe lag bei 51,93 ( $\pm 5,71$  Standardabweichung) (siehe Tabelle 1).

Bei der Auswahl der Kinder für die Gruppe der rechtschreibgestörten Kinder konnte die Leseleistung der Kinder nicht als Kriterium berücksichtigt werden, da die Leseleistungen der Kinder zu gut waren, und damit, durch Ausschluss der zu guten Leser, die Gruppe der rechtschreibgestörten Kinder zu klein geworden wäre. Es wird auch von der ICD-10 nicht genau definiert wie stark die Leseleistung zusätzlich zu der Rechtschreibleistung beeinträchtigt sein müsste, um die Diagnose Lese-Rechtschreibstörung zu stellen. Die Leseleistung der Kinder mit einer Rechtschreibstörung ist zwar im Durchschnitt geringer. Die mittlere Leseleistung der Rechtschreibgestörten betrug 40,04 T-Werte ( $\pm 7,14$  Standardabweichung), die der Kontrollgruppe betrug 53,37 T-Werte ( $\pm 8,42$  Standardabweichung). Die Leseleistung der rechtschreibgestörten Kinder erfüllt aber nicht das für unterdurchschnittliche Leistungen gebräuchliche Maß von mindestens einer Standardabweichung unter dem Mittelwert. Somit wird die Gruppe der Betroffenen auch nur als „rechtschreibgestört“ bezeichnet.

Klasse		1	2	3	4	insgesamt
Anzahl je Gruppe		n = 5	n = 6	n = 8	n = 8	n = 27
RST	KK	52,6 $\pm$ 1,34	57 $\pm$ 7,24	55,13 $\pm$ 4,19	57,5 $\pm$ 7,50	55,78 $\pm$ 5,79
	RS	38,8 $\pm$ 1,09	34,83 $\pm$ 5,38	36,25 $\pm$ 4,13	35,13 $\pm$ 4,70	36,07 $\pm$ 4,29
LES	KK	51,6 $\pm$ 7,70	48,83 $\pm$ 11,27	55,25 $\pm$ 4,23	56 $\pm$ 9,44	53,37 $\pm$ 8,42
	RS	39,8 $\pm$ 4,92	37,17 $\pm$ 9,09	40,88 $\pm$ 8,1	41,5 $\pm$ 6,37	40,04 $\pm$ 7,14
CFT	KK	55,2 $\pm$ 5,89	48,83 $\pm$ 6,91	51 $\pm$ 5,68	53,13 $\pm$ 4,05	51,93 $\pm$ 5,71
	RS	54,4 $\pm$ 5,86	48,5 $\pm$ 7,39	48,75 $\pm$ 4,89	52,88 $\pm$ 3,68	50,96 $\pm$ 5,67

Tabelle 1: Die Tabelle zeigt die durchschnittlichen Leistungen der Kinder in den Rechtschreib-, Lese- und Intelligenztests der Gesamtgruppe und unterteilt in die einzelnen Klassenstufen. RST = Rechtschreibtest, LES = Lesetest, CFT = Culture-Fair-Intelligenz-Test; **KK** = Kontrollkinder, RS = rechtschreibgestörte Kinder; in diesen Spalten sind die Mittelwerte und Standardabweichungen von T-Werte angegeben. Eine genaue Auflistung der individuellen Ergebnisse der einzelnen Paarlinge in den Tests befindet sich im Anhang (Tabelle 20).

## 4.2 Datenerhebungen der unabhängigen Variablen

Die unabhängigen Variablen wurden anhand standardisierter Tests, die die individuelle Rechtschreib-, Lesefähigkeit und Intelligenz der Kinder erfassen, ermittelt. Diese Tests sind mittels sehr großer Stichproben normiert worden, so dass für die Testauswertung – die Übertragung der Rohwerte – Normtabellen, die sich auf bestimmte Auswertungszeiträume beziehen, vorliegen. Ich führte jeden Test in Form eines Gruppentests klassenweise vormittags durch.

Alle verwendeten Tests erfüllen die Testgütekriterien von Objektivität, Reliabilität und Validität, so dass sie als objektives, zuverlässiges und gültiges Instrument zur Erfassung der individuellen Rechtschreib-, Leseleistung und Intelligenz angesehen werden.

### 4.2.1 Rechtschreibleistung

Die Rechtschreibleistung der Kinder wurde mit Hilfe von verschiedenen standardisierten Rechtschreibtests abhängig von Alter, Klassenstufe und Zeitpunkt der Testdurchführung<sup>6</sup> gemessen. Die Tests wurden als Gruppentests pro Klasse innerhalb einer Schulstunde durchgeführt. Für die Klassenstufe 1 und einen Teil der Klassenstufe 2 wurde der Weingartener Grundwortschatz Rechtschreibtest (WRT) 1+ verwendet. Die restlichen Zweitklässler wurden zu einem späteren Zeitpunkt mit dem Weingartener Grundwortschatz Rechtschreibtest (WRT) 2+ getestet. In der dritten Klasse wurde der Diagnostische Rechtschreibtest (DRT) 3+ durchgeführt. Die Klassenstufe 4 wurde mit der Hamburger Schreib-Probe (HSP) 4/5 getestet.

Im Folgenden erläutere ich die verwendeten Rechtschreibtests näher.

Der Weingartener Grundwortschatz Rechtschreibtest (WRT) 1+ erfasst die Rechtschreibfähigkeit von Schülern am Ende der ersten und am Anfang bzw. Mitte der zweiten Klasse. In diesem Test wird den Kindern ein Lückentext vorgelegt, in den sie 25 fehlende Grundwortschatz-Wörter in die entsprechenden Lücken hinein schreiben. Nach Birkel (1995) erfüllen diese Wörter fünf der von Rückemann (1991 in Birkel

---

<sup>6</sup> Die Tests wurden in den Schulen am 31.03.2004, 01.04.2004 und 05.05.2004 durchgeführt. Diese Daten entsprachen der 26. bzw. 29. Unterrichtswoche. Die 29. Unterrichtswoche war die 11. Woche vor Schuljahresende.

1995) aufgestellten sieben Kriterien für einen Grundwortschatz. Die Testdurchführung dauert maximal 45 min. Zur Ermittlung des Rohwertes werden die richtig geschriebenen Wörter gezählt (pro Wort ist somit maximal ein Fehler möglich). Die Auswertung<sup>7</sup> der Rohwerte erfolgt anhand verschiedener Normtabellen, die sich auf die letzten 12 Wochen des ersten Schuljahres bzw. die 15. – 27. Unterrichtswoche des zweiten Schuljahres beziehen. Die für die Normtabellen verwendeten Eichstichproben umfassen beim WRT 1+ insgesamt 193 Klassen aus 15 der 16 Bundesländer.

Der Weingartener Grundwortschatz Rechtschreibtest (WRT) 2+ erfasst die Rechtschreibleistung von Kindern am Ende der zweiten Klasse und Anfang/Mitte der dritten Klasse. Der Weingartener Grundwortschatz Rechtschreibtest (WRT) 2+ entspricht in Durchführung und Auswertung dem Weingartener Grundwortschatz Rechtschreibtest (WRT) 1+. Es wird hier allerdings, auf den Stand der zweiten und dritten Klasse angepasst, ein anderer Lückentext mit 43 fehlenden Grundwortschatz-Wörtern verwendet. Die Auswertung<sup>8</sup> der Rohwerte erfolgt anhand von Normtabellen, die auf einer Eichstichprobe von insgesamt 163 Klassen aus allen Bundesländern basieren.

Nach Birkel (1995) werden beide Tests (WRT 1+ und 2+) als ein objektiv messendes Instrument zur Feststellung der Rechtschreibfähigkeit angesehen. Auch ist nach Birkel (1995) den Tests eine recht gute Meßgenauigkeit zuzuschreiben. Im WRT 1+ wird ein Testhalbierungskoeffizient (odd-even Methode) von  $r = 0,91$  ( $n = 500$ , Teilstichprobe der Eichstichprobe), ein Konsistenzkoeffizient von  $r = 0,92$  ( $n = 500$ , Teilstichprobe der Eichstichprobe) und ein Paralleltest-Reliabilitätskoeffizient von  $r = 0,89$  ( $n = 1306$ , Ende der 1. Klasse) angegeben, im WRT 2+ betragen die Testhalbierungskoeffizienten  $r = 0,94$  bzw.  $r = 0,93$  ( $n = 450$ ), der Konsistenzkoeffizient  $r = 0,94$  ( $n = 450$ ) und der Paralleltest-Reliabilitätskoeffizient  $r = 0,93$  ( $n = 2929$ , 2. Klasse). Alle Koeffizienten signalisieren nach Birkel (1995) eine sehr gute Reliabilität des Testes. Nach Birkel (1995) kann in beiden Tests eine Inhaltsvalidität angenommen werden. Die Übereinstimmungsvalidität wurde anhand Leistungseinschätzungen der Lehrer in einer

7 Bei sieben der insgesamt zwölf getesteten Erstklässlern wurde der WRT 1+ eineinhalb Unterrichtswochen zu früh durchgeführt und war dadurch für die Kinder eher etwas zu schwer. Dies kann dazu führen, dass die T-Werte minimal zu niedrig ausgefallen sind (um 1 bis 2 T-Punkte). Es stand leider kein alternativer Test zur Verfügung. Der Teil der Zweitklässler, die an diesem Test teilnahmen, wurden in der 26. Unterrichtswoche getestet, die sich innerhalb des Auswertungsbereiches befand.

8 Der WRT 2+ wurde mit dem anderen Teil der Zweitklässler vier Wochen später in der 29. Unterrichtswoche innerhalb des Auswertungsbereiches, den letzten zwölf Wochen des zweiten Schuljahres, durchgeführt.

zehnstufigen Rangskala überprüft. Es werden Validitätskoeffizienten zwischen  $r = 0,63$  ( $n = 1245$ ) und  $r = 0,80$  ( $n = 1064$ ) auf Basis der Eichdaten je nach Anfang, Mitte oder Ende der Klassenstufe angegeben. Im WRT 2+ betragen die entsprechenden Werte zwischen  $r = 0,65$  und  $r = 0,69$ .

Beide Tests sind somit objektive, reliable und valide Erhebungsinstrumente für die individuelle Rechtschreibfertigkeit.

Der Diagnostische Rechtschreibtest (DRT ) 3+ erfasst die Rechtschreibleistung von Kindern in den letzten vier Monaten des dritten und den ersten drei Monaten des vierten Grundschuljahres. In diesem Test liegen den Kindern 44 Sätze vor, in denen jeweils ein Wort als Lücke fehlt, welches die Kinder eintragen. Die Wörter nehmen nach Müller (1997) in ihrer Schwierigkeit zu. Der Test kann innerhalb von 25-35 min durchgeführt werden. Neben einer quantitativen Auswertung ist es möglich, den Test zusätzlich qualitativ auszuwerten, was bei der hier vorliegenden Zielsetzung nicht notwendig war. Zur Ermittlung des Rohwertes eines Kindes werden in der quantitativen Auswertung<sup>9</sup> alle falsch geschriebenen Wörter gezählt. Für diese quantitativen Rohwerte liegen Prozentrangnormen beruhend auf einer Eichstichprobe von 103 Klassen aus alten und neuen Bundesländern, vor. Sowohl die Durchführung und die quantitative Auswertung des Testes sind nach Müller (1997) objektiv. Es wird in der Reliabilitätsprüfung des Testes (quantitativen Auswertung) ein Korrelationskoeffizient für die Paralleltestzuverlässigkeit von  $r = 0,92$  ( $n = 179$ ) und ein Korrelationskoeffizient für die Halbierungszuverlässigkeit (odd even, vgl. Müller 1997, 40) von  $r = 0,95$  ( $n = 200$ , korrigiert nach Spearman-Brown; 3. Auflage 1997) angegeben. Nach Müller 1997 ist eine Inhaltsvalidität gegeben, ebenso ist nach Müller die Übereinstimmung zwischen Lehrerurteil (Zensuren zwischen 1 – 6) und Test recht hoch, der Zusammenhang wird im Mittel mit  $r = 0,78$  (Mittel der Korrelationskoeffizienten von 156 Klassen) angegeben.

Insgesamt kann auch der DRT 3+ als ein objektives, reliables und valides Erhebungsinstrument für die Rechtschreibfertigkeit angesehen werden.

---

<sup>9</sup> Alle Drittklässler wurden innerhalb des vorgegebenen Auswertungszeitraumes getestet.



Die Hamburger Schreib-Probe (HSP) 4/5 kann unter der Voraussetzung von zugrundeliegenden Normstichproben für die Auswertung in der Mitte und am Ende der vierten Klasse und am Anfang der fünften Klasse durchgeführt werden. Die Hamburger Schreib-Probe umfasst 16 Einzelwörter und 5 Sätze, wobei die die Bedeutung der Wörter und Sätze durch Illustrationen veranschaulicht wird. Durch die Bildvorgabe wird nach May (2000) das Erfassen der Bedeutung als wichtige Bedingung des Rechtschreibens gesichert. Der Test kann in weniger als einer Schulstunde durchgeführt werden. Als Rohwert wird die Zahl richtig geschriebener Wörter ermittelt. Es ist möglich den Test nicht nur auf Wortebene, sondern zusätzlich auf Graphemebene auszuwerten, was allerdings über die hier vorliegende Zielsetzung hinausgeht. Die Auswertung auf Wortebene ist nach May (2000) als objektiv anzusehen. Die für die Auswertung<sup>10</sup> verwendeten Normtabellen repräsentieren die 15.-23. Unterrichtswoche und die 33.-40.Unterrichtswoche. Die Reliabilität (Auflage 5, 2000) wird mit einem Zuverlässigkeitskoeffizienten zwischen  $r = 0,92$  und  $r = 0,99$  für die interne Konsistenz der Gesamtergebnisse angegeben. Die Stabilität der Gesamtergebnisse liegt (je nach Abstand der Erhebungszeitpunkte) zwischen  $r = 0,52$  und  $r = 0,93$  (Auflage 5, 2000). Es wird eine Übereinstimmungsvalidität zwischen der Rechtschreibleistung in der HSP und der Rechtschreibleistung in Deutschaufsätzen mit den Korrelationen von  $r = 0,78$  und  $r = 0,81$  (Auflage 6, 2002) angegeben.

Insgesamt stellt auch die HSP 4/5 ein objektives, zuverlässiges und gültiges Instrument zur Erfassung der individuellen Rechtschreibfähigkeit dar.

## 4.2.2 Leseleistung

Neben den Rechtschreibtests wurden zusätzlich standardisierte Lesetests in der gleichen Schulstunde jeweils in den einzelnen Klassen durchgeführt. In der 1. Klasse wurde hierfür die Würzburger-Leise-Leseprobe (WLLP) verwendet. Die restlichen

---

<sup>10</sup> Die für die Auswertung verwendeten Normtabellen repräsentieren die 15.-23. Unterrichtswoche (Zeitpunkt der Testdurchführung von 11 der Viertklässler: 26.Unterrichtswoche) und die 33.-40.Unterrichtswoche (Zeitpunkt der Testdurchführung von 3 der Viertklässler: 31.Unterrichtswoche). Von der HSP ist bei vielen Anwendern bekannt, dass dieser Test allgemein zu leicht ist (S.151, Deimel 2002). Daher wurden alle Werte (auf der Basis von umfangreichen eigenen Datenerhebungen) um fünf T-Punkte nach unten korrigiert, damit die HSP-Werte mit den Werten der anderen Testverfahren vergleichbar sind. Der größte Teil der Viertklässler wurde außerdem drei Wochen zu spät getestet. Dies kann dazu führen, dass die T-Werte nochmals minimal zu hoch ausgefallen sind.

Klassenstufen wurden mit dem Salzburger Lese Screening (SLS 1-4) getestet. Die Stichprobengewinnung erfolgte jedoch unabhängig von den Leistungen der Kinder im Lesetest.

Im Folgenden erläutere ich die beiden verwendeten Lesetests näher.

In der Würzburger-Leise-Leseprobe (WLLP) werden zu einem geschriebenen Wort vier Bildalternativen in einer Reihe dargeboten. Das Kind muss das zu dem geschriebenen Wort passende Bild anstreichen. Unter den vier bildlichen Auswahlmöglichkeiten befinden sich Wörter ähnlicher phonologisch-orthographischer Zusammenhänge (z. B. Knopf – Kopf) oder semantischer Zusammenhänge (z. B. Knopf – Hose). Insgesamt können 140 Reihen bearbeitet werden, was allerdings durch ein Zeitlimit von 5 min eingeschränkt ist. Testrohwert sind alle korrekt bearbeitete Reihen. Es liegen für die erste Klasse Prozentrangnormen vor, die für den Zeitraum Juni/Juli gelten<sup>11</sup>. Die Würzburger-Leise-Leseprobe kann nach Küspert & Schneider (1998) sowohl in Durchführung als auch Auswertung als völlig objektiv bezeichnet werden. Die Reliabilität wird anhand der Paralleltestmethode, bei der ein Korrelationskoeffizient von  $r = 0,97$  ( $n = 175$ ) angegeben ist, und anhand der Retestmethode, bei der ein Korrelationskoeffizient von  $r = 0,94$  ( $n = 513$ ) angegeben ist, überprüft. Der Test ist nach Küspert & Schneider (1998) inhaltlich valide, da auf nichts anderes als die Lesefertigkeit geschlossen werden kann. Die Validitätsprüfung erfolgt unter anderem anhand einer Lehrerbeurteilung auf Basis einer 7-stufigen Skala. Hier ist für die erste Klasse ( $n = 74$ ) ein Spearman'scher Rang-Korrelationskoeffizient von  $r = 0,75$  angegeben.

Insgesamt kann die Würzburger-Leise-Leseprobe (WLLP) als ein objektives, reliables und valides Erhebungsinstrument für die Lesefertigkeit angesehen werden.

Das Salzburger Lese Screening (SLS 1-4) umfasst 70 kurze sehr einfache richtige und falsche Sätze, die innerhalb von 3 min möglichst schnell gelesen und nach ihrer Richtigkeit beurteilt werden sollen. Die Sätze werden unter lesetechnischer Hinsicht immer schwieriger, so dass sie vom Ende der ersten Klasse bis Ende der vierten Klasse

---

<sup>11</sup> Der Test wurde zu früh (Ende März, Anfang April und Anfang Mai) im Vergleich zu dem Auswertungszeitraum der Norm von Juni/ Juli durchgeführt. Für die erste Klasse steht allerdings in diesem frühen Zeitraum kein anderer Test zur Verfügung.

verwendet werden können. Separate Normen ( $n = 1867$ ) liegen für das Ende der ersten Schulstufe bzw. den Beginn der 2. Schulstufe und für jedes Halbjahr der 2. bis 4. Klassenstufe vor<sup>12</sup>. Mit Hilfe der Normwerte (Anzahl der richtig bearbeiteten Sätze) wird ein Lesequotient (LQ) ermittelt, der wie der Intelligenzquotient skaliert ist (Mittelwert von 100, Standardabweichung von 15). Auswertung und Durchführung des Testes werden von Mayringer & Wimmer (2003) als objektiv betrachtet. Die Paralleltest-Reliabilitäten liegen zwischen  $r = 0,90$  und  $r = 0,92$ . Zur Validierung des SLS 1-4 wurde der Subtest Textlesen des SLRT herangezogen. Korreliert wurden die Lesezeit beim Textlesen mit der Anzahl korrekt bearbeiteter Sätze im Lese-Screening. Dabei ergaben sich Validitätskoeffizienten von  $r = 0,81$  in der 2. und 3. Schulstufe sowie  $r = 0,76$  in der 4. Schulstufe.

Insgesamt stellt auch das Salzburger Lese Screening (SLS 1-4) ein objektives, zuverlässiges und gültiges Instrument zur Erfassung der individuellen Lesefähigkeit dar.

### **4.2.3 Intelligenz**

Die Intelligenz wurde in der ersten und zweiten Klasse anhand des Culture-Fair-Intelligenz-Test CFT 1 und in der dritten und vierten Klasse anhand des Culture-Fair-Intelligenz-Test CFT 20 gemessen. Beide Intelligenztests sind sprachfrei und als Gruppentest im Klassenverband durchführbar.<sup>13</sup> Für beide Intelligenztests ist das Intelligenzkonzept von Catell (1968 in Weiß 1987) grundlegend. Dieses unterscheidet die zwei Allgemeinfaktoren, die General-Fluid-Ability (flüssiger Intelligenzfaktor) – die Fähigkeit der Wahrnehmung und Erfassung komplexer Beziehungen in neuartigen Situationen, und die General-Crystallized-Ability (kristallisierter Intelligenzfaktor) – Fähigkeiten, die normalerweise in der Schule gelehrt werden (Weiß 1987, 28). Der flüssige Intelligenzfaktor hat bei beiden Intelligenztests einen stärkeren Anteil, in dem sie als Wahrnehmungstest konstruiert wurden, einer Testart, die sich nach Catell (1968 in Weiß 1987) am stärksten als kulturfrei erwiesen hat.

Im Folgenden gehe ich näher auf die beiden Intelligenztests ein.

---

<sup>12</sup> Alle Testdurchführungen befanden sich im Zeitraum des Auswertungsbereiches.

<sup>13</sup> Die Intelligenztests wurden klassenweise in jeweils einer Schulstunde am 12., 13. und 14.05.04 durchgeführt. Die Testdurchführungen lagen innerhalb der vorgegebenen Auswertungszeiträume.

Der Culture-Fair-Intelligenz-Test CFT 1 umfasst 5 Subtests (Substitution, Labyrinth, Klassifikation, Ähnlichkeiten, Matrizen) mit 5 verschiedenen Aufgabenstellungen. Der CFT 1 ermöglicht die Bestimmung der Grundintelligenz von 5-9 jährigen Kindern mittels der Bestimmung der Fähigkeit des Kindes in neuartigen Situationen und anhand von sprachfreiem, figuralem Material, Denkprobleme, Regeln und Merkmale zu erfassen und Beziehungen zu entwickeln. In der ersten Klasse dauert der Test etwa 40 min. Für die Rohwerte liegen Alters- und Klassennormen in Form von Prozentwerten, T-Werten und Intelligenzquotienten vor. Ermittelt wurde in der vorliegenden Untersuchung als Gesamtrohwert nur die Summe aller Rohwerte der 5 Subtests. Es ist anhand der Normtabellen, die auf einer Eichstichprobe von insgesamt 6078 Schülern basieren, neben der Übertragung des Gesamtrohwertes zusätzlich möglich, summierte Rohwerte ausgewählter Subtests zu übertragen. Bezüglich der Reliabilität sind Zuverlässigkeitskoeffizienten der Subtests zwischen  $r = 0,85$  und  $r = 0,91$  (Testhalbierungsmethode, korrigiert nach Spearman-Brown) angegeben. In der Retestanalyse unter Zusammenschau von Einzel- und Klassenretests wird ein Korrelationskoeffizient von  $r = 0,84$  für die Summe 1 angegeben. Bei der Validitätsbestimmung ist es nach Weiß (1987) primär wichtig festzustellen, welcher Art intellektuellen Verhaltens der CFT 1 misst und nicht seine prognostische Funktion. Für jeden Subtest wurden hierfür mittels einer Faktorenanalyse ermittelt, welchen Anteil ein Subtest am General-Ability-Factor (sogenannte „g-Sättigung“) hat und welche Ladungsanteile in Sekundärfaktoren bei dem Subtest vorhanden sind. Dabei wurde festgestellt, dass die ermittelten „g-Sättigungen“ gut mit vergleichbaren von Catell ermittelten Werten übereinstimmen (siehe Weiß 1987, 28). Die Korrelation (berechnet nach Spearman) der CFT 1-Gesamtleistung mit dem HAWIK-Gesamtergebnis ist mit  $r = 0,66$  nach Weiß (1987) als befriedigend bis gut zu bezeichnen. Der CFT 1 korreliert hoch mit konstrukt-nahen Variablen (z. B. HAWIK-Handlungsteil:  $r = 0,66$ , DVET:  $r = 0,59$ , BT  $r = 0,51$ ; berechnet nach Spearman) und niedrig mit konstrukt-fernen Variablen (z. B. Rechtschreibtes DRT 2:  $r = 0,3$ ; berechnet nach Spearman).

Der Culture-Fair-Intelligenz-Test CFT 20 besteht aus zwei Testteilen, von denen der erste Teil als Kurzform zur Erfassung der Intelligenz in einem Altersbereich von 8 Jahren und 7 Monaten bis 70 Jahren verwendet werden kann. Beide Testteile bestehen

aus je vier Subtests (Series, Classification, Matrices, Topologies nach Weiß, 1987, 10) mit je 46 Aufgabenreihen zunehmender Schwierigkeit. Die Kurzform kann innerhalb von ca. 37 min durchgeführt werden. Die reine Testzeit beträgt 14 min. Die Rohwerte können anhand von Normtabellen, basierend auf einer Eichstichprobe von 4400 Schülern der 3.-10.Klasse, für die entsprechenden Altersgruppen in IQ-Punkten, T-Werte und Prozentränge umgewandelt werden. Die Objektivität ist sowohl für die Durchführung als auch die Auswertung nach Weiß (1987) gegeben. Für den gesamten Test wird der Reliabilitätskoeffizient bei einer Testwiederholung mit der Parallelform mit  $r = 0,77$  ( $N = 70$ ), der Konsistenzkoeffizient (split-half) mit  $r = 0,95$  angegeben. Bezüglich der Validitätsprüfung korreliert der CFT 20 zu komplexen Begabungstests (z. B. PSB, LPS, FAT) durchschnittlich mit  $r = 0,64$ , mit konstrukt-nahen Tests zu  $r = 0,54$  und mit konstrukt-fernen Tests zu  $r = 0,29$ . Zu Schulnoten wird eine Korrelation von  $r = 0,54$  für die Mathematiknote und von  $r = 0,29$  für die Deutschnote angegeben.

Insgesamt können somit beide Tests als objektives, reliables und valides Erhebungsinstrument für die Intelligenz angesehen werden.

## **4.3 Konstruktion des Artificial Grammar Learning-Experimentes**

### **4.3.1 Konsequenzen aus der Literatur für die Konstruktion des AGL-Experimentes**

Aufgrund von Erkenntnis und wichtigen Aspekten aus veröffentlichten AGL-Experimenten zur Konstruktion von AGL-Experimenten und der besonderen Herausforderung ein AGL-Experiment mit Kindern durchzuführen, ergeben sich für den Aufbau und die Durchführung dieses AGL-Experimentes folgende Konsequenzen.

Um implizite Lernvorgänge annehmen zu können, ist die Lernphase, bezüglich ihrer Intensität und der Instruktion zu Anfang der Lernphase von Bedeutung.

Gomez (1997) nahm an, dass je nach Intensität der Lernphase (beispielsweise durch das Reflektieren des erworbenen Wissens mit Hilfe von Erinnerungstests) explizite Prozesse

bei komplexen Lernleistungen anzunehmen sind. Es ergibt sich hier also die Frage, wieviele von den aus der Grammatik maximal möglichen konstruierbaren Items den Probanden als Lernitems präsentiert werden, wie oft diese Items gezeigt werden sollen und wie sich die Probanden mit den Items auseinandersetzen (z.B. durch Nachschreiben). Beim Vergleich von AGL-Experimenten untereinander, gibt es diesbezüglich keine genauen Vorschriften.

Das Testmaterial des AGL-Experiment dieser Untersuchung ist an das Testmaterial des Experimentes 1 von Knowlton & Squire (1996) angelehnt. Da Knowlton & Squire (1996) 23 Lernitems von 46 möglichen Items den Probanden präsentierten, wurden in dem hier vorliegenden Experiment den Kindern auch 23 Lernitems präsentiert.

In den meisten AGL-Experimenten werden die Probanden dazu aufgefordert, die ihnen präsentierten Lernitems nachzuschreiben – mittlerweile mittels Bildschirm und Tastatur. Auch in dieser Studie werden die Kinder in der Lernphase aufgefordert, die präsentierten Items auf einer Tastatur nachzuschreiben, was damit dem Vorgehen der meisten AGL-Experimente entspricht. Um dies den Kindern zu vereinfachen, werden mit Hilfe einer Schablone nicht benötigte Tasten abdeckt, damit es für die Kinder leichter ist die Buchstaben zu finden. Eine ähnlichen Ansatz hatten Witt, Nühsman & Deuschl (2002), die nicht benötigte Tasten von der Tastatur entfernten, um Parkinsonpatienten (mit Tremor) das Schreiben zu erleichtern.

Durch Kritik an Experimenten, die auf ungewöhnlich langen Lernphasen beruhen (z. B. mehrere hundert Lerndurchführungen in einem Zeitraum von vier Wochen – nähere Erläuterungen, siehe 2.1.2), wird deutlich, dass diese Fragestellung, wie oft die Lernitems gezeigt werden sollen, von Bedeutung für die Konstruktion der Lernphase ist. Im Vergleich sind die Anforderungen in AGL-Experimenten an die Probanden bezüglich dem Nachschreiben der Lernitems sehr unterschiedlich. Kurz zusammengefasst (nähere Erläuterungen, siehe 2.1.2), werden die Items etwa 2-6 mal präsentiert, jedes falschgeschriebene Item wird direkt ein weiteres Mal gezeigt, entweder bis zu maximal drei Versuchen, es richtig zu schreiben, oder ein falschgeschriebenes Item wird so oft gezeigt, bis es mindestens einmal richtig geschrieben wurde. Gerade letzteres Vorgehen ist unmöglich von den Kindern zu

verlangen, mit der Überlegung, dass Kinder teilnehmen, die größte Schwierigkeiten haben können Items nachzuschreiben. Auch besteht die Gefahr die Kinder zu demotivieren, wenn sie direkt mit ihren Fehlern konfrontiert werden. Aus diesen Gründen werden die Lernitems den Kindern insgesamt dreimal präsentiert, und ein falschgeschriebenes Lernitem wird im weiteren Verlauf der Lernphase nur ein weiteres Mal wiederholt.

Um Unterschiede zwischen guten und schlechten Nachschreibern möglichst gering zu halten, die Kinder nicht zu überfordern oder zu demotivieren und trotzdem zu gewährleisten, dass sich die Kinder mit den Lernitems ausreichend auseinandergesetzt haben, werden in dieser Untersuchung die Kinder also dazu aufgefordert, 23 Lernitems von insgesamt 46 konstruierbaren Items ohne Fehler je dreimal und mit (mindestens einem) Fehler je viermal in zufälliger Reihenfolge nachzuschreiben.

Desweiteren ist die Instruktion der Probanden zu Beginn der Lernphase wichtig für die Art des Wissenserwerbes während dem AGL-Experiment. Nach Meinung von Whittlesia (1997), Whittlesia & Dorken (1993) und Johnstone & Shanks (2001) kann die Aufforderung der Probanden in der Lernphase Regeln zu identifizieren, zu einem aktiven und bewussten Bemühen und damit zu expliziten Prozessen führen. Nach Turner & Fischler (1993) ist die Klassifikationsleistung der Probanden geringer, wenn sie in der Lernphase instruiert werden, aktiv nach Regeln zu suchen (siehe auch 2.1.3).

Diese Erkenntnisse sind für die vorliegende Untersuchung insofern wichtig, dass den Kindern durch die Instruktion vor der Lernphase nicht bewusst wird, dass Eigenschaften der Lernitems für die sich anschließende Testphase relevant sein können, um bei den Kindern keine expliziten, aktiven Bemühungen hervorzurufen. Deswegen werden die Kinder vor der Lernphase lediglich darüber informiert, dass sie eine 'Geheimsprache' lernen und damit dies gelingt, die Wörter dieser Geheimsprache nachgeschrieben werden. Erst nach der Lernphase wird den Kindern gesagt, dass nun (in der Testphase) weitere Wörter folgen, welche entweder zu der Geheimsprache gehören oder auch nicht, und sie nun für jedes einzelne Wort entscheiden sollen, ob es zu der Geheimsprache gehört oder nicht.

Die Kinder werden vor der Lern- und Testphase so instruiert, dass anzunehmen ist, dass die Instruktionen keine Auswirkungen auf das Lernen und die Klassifikationsleistungen der Kinder haben (nähere Erläuterungen, siehe 2.1.3).

Insgesamt gesehen, ist die Durchführung des AGL-Experimentes in dieser Untersuchung somit vergleichbar zu Durchführungen von anderen veröffentlichten Untersuchungen, in denen angenommen wurde, dass implizites Lernen stattfand (Knowlton & Squire, 1994, 1996; Gomez 1997; Kinder, Shanks, Cook & Tunney 2003).

### 4.3.2 Materialien

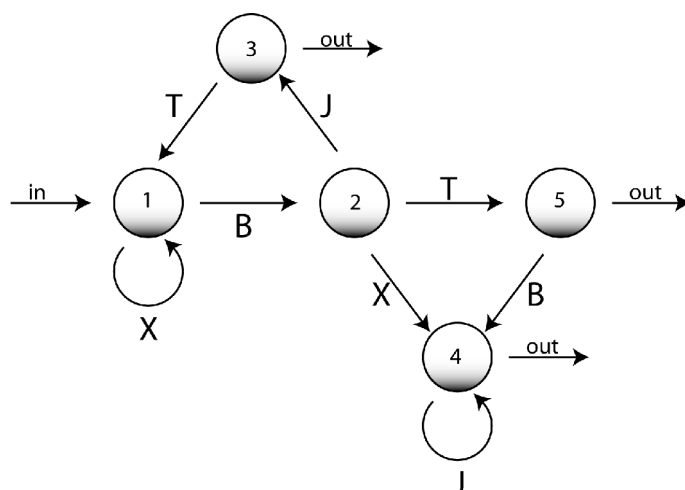


Abbildung 2 Diese Abbildung zeigt ein künstliches grammatisches Regelsystem („artificial grammar“), entwickelt von Abrams & Reber (1989).

Die Items wurden nach dem von Abrams & Reber (1989) entwickelten künstlichen Regelsystem konstruiert. Im Original ist der Buchstabe B ein V. Die Buchstabenreihen werden generiert, indem man von Pfeil zu Pfeil die Buchstaben der Reihe nach anfügt, bis man zu einem „out“-Pfeil gelangt. Die Reihen haben eine Länge von zwei bis maximal sechs Buchstaben. Es können so mit dieser Grammatik insgesamt 46 Reihen gebildet werden. Von diesen wurden, angelehnt an die im Experiment 1 von Knowlton & Squire (1996) verwendeten Items, 23 Lern- und 16 Testitems ausgewählt. Die Hälfte der insgesamt 32 Testitems wurden also nach der Grammatik konstruiert (grammatisch



= G), während die andere Hälfte die Regeln der Grammatik an mindestens einer Stelle verletzte (nichtgrammatisch = NG).

Anhand der 23 Lernitens wurde die Fragmentstärke für alle weiteren, in der Testphase verwendeten Testitens, berechnet. Einige Beispiele für Lernitens sind: XBT, XBJTBJ, BXJJJJ, BXJJ, BTBJ.

Ein Fragment („chunk“) ist definiert als Bi- oder Trigramm, das in den Items enthalten ist. Die Fragmentstärke (Fs) – „chunk strength“ von Knowlton & Squire (1994,1996) entwickelt – jedes einzelnen Testitens wird berechnet, indem gezählt wird, wie oft die Fragmente des Testitens in den Lernitens vorkommen, und daraufhin dieser Wert durch die Gesamtanzahl der im Testitem enthaltenen Fragmente geteilt wird. Z. B. sind in dem grammatischen Testitem BXJ die drei Fragmente BX, XJ und BXJ enthalten. Das Bigramm BX kommt 12 mal, das Bigramm XJ und das Trigramm BXJ je 8 mal in den Lernitens vor, also beträgt die Fragmentstärke 9,33 ( $12 + 8 + 8 = 28$ ;  $28 : 3 = 9,33$ ).

Die Testitens können so nach ihren vier Eigenschaften in vier Gruppen unterteilt werden: grammatische Items mit hoher Fragmentstärke (g/hFs), grammatische Items mit niedriger Fragmentstärke (g/nFs), nichtgrammatische Items mit hoher Fragmentstärke (ng/hFs) und nichtgrammatische Items mit niedriger Fragmentstärke (ng/nFs).

g/hFs		g/nFs		ng/hFs		ng/nFs	
XXXXBT	9,22	BJTBT	5,57	XBXBJ	9	XXBBJJ	6
BX	12	XBTBJJ	6,22	XXB	12	BJJXBT	4,89
XBJ	9,33	BJTBJ	6,14	XBXX	9,6	JXBT	5,2
XXBX	11,2	BTB	5,67	XXBXXB	10,11	BXJTJ	4,71
XXXXBJ	9,44	BTBJJ	5,29	XXX	10,33	BXJJX	5,86
BXJ	9,33	BTBJJJ	5,33	XBXB	10	TBJ	6,33
XXXBXJ	10	XBJTBT	6,33	XBXJT	8	BTBX	6,2
XXXBJ	9,57	BJTXBT	5,44	BXBJ	8	XXJT	5,2
<b>Mw der Fs</b>	<b>10,01</b>	<b>Mw der Fs</b>	<b>5,75</b>	<b>Mw der Fs</b>	<b>9,63</b>	<b>Mw der Fs</b>	<b>5,55</b>

Tabelle 2: In dieser Tabelle sind die verwendeten Testitens nach ihren Eigenschaften (grammatisch / nichtgrammatisch und niedriger / hoher Fragmentstärke) unterteilt. Hinter jedem Item ist die jeweilige Fragmentstärke dieses Items angegeben. Mw = Mittelwert, g = Grammatisch, ng = Nichtgrammatisch, Fs = Fragmentstärke, h = hohe, n = niedrige

Typisch ist für die Testitems, die nach der Grammatik aufgebaut sind, dass sie eine höhere Fragmentstärke haben, als die Testitems, die die Regeln der Grammatik verletzen, weil die nichtgrammatischen Items teils aus Fragmenten bestehen, die nicht oder nur wenig mit den Fragmenten der Lernitems übereinstimmen. Um dies zu vermeiden, hatten Knowlton & Squire dem Experiment 1 (1996) die Testitems so konstruiert, dass die durchschnittliche Fragmentstärke der grammatischen und nichtgrammatischen Testitems in etwa gleich war, entweder hoch mit einer durchschnittliche Fragmentstärke von 8,6 für grammatische und nicht grammatische Items oder niedrig mit einer durchschnittliche Fragmentstärke von 5,8 für die grammatischen Items bzw. 5,4 für die nichtgrammatischen Items. Je acht der grammatischen und nichtgrammatischen Testitems hatten also eine hohe Fragmentstärke hFs, die restlichen acht eine niedrige Fragmentstärke nFs. Die Auswahl der Items für dieses Experiment war an diejenige von Knowlton & Squire (1996, Experiment 1) angelehnt, die Items wurden sogar so modifiziert, dass sich die einzelnen Items in ihrer Fragmentstärke noch stärker unterschieden. Die durchschnittliche Fragmentstärke hatte zwar bei Knowlton & Squire (1996, Experiment 1) eine Differenz von ungefähr 3, es kamen jedoch einzelne Items vor, die sich, obwohl sie in niedriger und hoher Fragmentstärke unterteilt wurden, im Vergleich kaum in ihrer Fragmentstärke unterschieden. z. B. war die Fragmentstärke eines grammatischen Items mit hoher Fragmentstärke 6,8, während die Fragmentstärke eines grammatischen Items mit niedriger Fragmentstärke 6,7 war. Um also den Effekt zwischen der Befolgung der grammatischen Regeln und der Fragmentstärke bei der Klassifizierung der Testitems besser unterscheiden zu können, war in diesem Experiment die Fragmentstärke der Items mit hoher Fragmentstärke mindestens 8 und die der Items mit niedriger Fragmentstärke höchstens 6,33.

Die durchschnittliche Fragmentstärke der acht g/hFs-Items (grammatisch mit hoher Fragmentstärke) betrug 10,01, die der acht g/nFs-Items (grammatisch mit niedriger Fragmentstärke) war 5,75, die durchschnittliche Fragmentstärke der acht ng/hFs-Items

(nicht grammatisch mit hoher Fragmentstärke) betrug 9,63 und die der acht ng/hFs-Items (nicht grammatisch mit niedriger Fragmentstärke) 5,55.

Des Weiteren wurden in der Originalgrammatik von Abrams & Reber (1989) und auch bei Knowlton & Squire (1996, Experiment 1) die Buchstaben X, V, J und T verwendet. In diesem Experiment wurde, wegen der zu großen Ähnlichkeit des Buchstaben V zum Buchstaben X, das im Original verwendete V gegen ein B ausgetauscht. Die Wahl fiel auf den Buchstaben B, da dieser eine andere äußere Form hat und direkt neben dem Buchstaben V auf der Tastatur liegt. Somit ist ein genügender Abstand zu den anderen Buchstaben X, J und T gewährleistet, um das Nachschreiben der Lernitems mit Hilfe einer Schablone zu ermöglichen.

Die folgende Tabelle zeigt die verwendeten Übungsitens der Übungsphasen (näher erläutert in 4.3.3) und die Lernitems.

Übungsitens		Lernitems	
<u>Übungslernphase:</u>	<u>Übungstestphase:</u>	XXXXBX	XBT
NKKN	KC	XXXBX	XBJTBJ
RC	CRC	XXXBTB	BXJJJ
RKNCRR	RKNCN	XXXBT	BXJJ
RKCR	KCNK	XXBXJJ	BTBJ
KCNC	CRNKCR	XXBXJ	BT
CRCN		XXBT	BJTXBJ
RKKNK		XXBJ	BJTBXJ
NCN		XBXJJ	BJTBX
KCNKCN		XBXJJ	BJTBTB
RCR		XBXJ	BJ
		XBX	

Tabelle 3: Übungsitens und Lernitems des Artificial Grammar Learning-Tests

### 4.3.3 Durchführung und Instruktionen

Der Artificial Grammar Learning-Test ist ursprünglich beschrieben als ein „paper-and-pencil“ Test. In dieser Studie wurden mit Hilfe eines Notebooks die Items den Probanden über ein Computerprogramm präsentiert. Das AGL-Experiment wurde immer von derselben Person (mir) vormittags, in einer ruhigen Umgebung ohne Störungen von außen, mit den Kindern einzeln unter Verwendung der gleichen

Instruktionen, also zusammengefasst unter konstanten Bedingungen durchgeführt. Der Test dauerte pro Kind ca. eine Schulstunde. Er bestand aus zwei Phasen, einer Lern- und einer Testphase. Zu diesen beiden Phasen wurde zuerst ein entsprechender Übungsdurchgang (Lernphase-Übungsdurchgang und Testphase-Übungsdurchgang) durchgeführt, der den Probanden den Ablauf der darauffolgenden relevanten Phase verdeutlichte. Somit erhielten die Kinder insgesamt vier Instruktionen: zu Anfang des Lernphase-Übungsdurchgangs, der Lernphase, des Testphase-Übungsdurchgangs und der Testphase.

#### **4.3.3.1 Lernphase**

##### **Aufbau und Durchführung:**

Der Aufbau der Lernphase und des entsprechenden Lernphase-Übungsdurchgangs ist gleich bis auf die Anzahl der präsentierten Items und ihre Wiederholungen. In dem Lernphase-Übungsdurchgang wurden nur zehn Übungsitens je einmal präsentiert, die unabhängig von einer Grammatik und aus anderen Buchstaben gebildet wurden (siehe Tabelle 3). In der Lernphase wurden dagegen 23 Lernitens verwendet, die ohne Fehler beim Nachschreiben je dreimal und mit (mindestens einem) Fehler je viermal in zufälliger Reihenfolge wiederholt gezeigt wurden.

Je eines der Items war für sechs Sekunden auf der Mitte des Bildschirms groß und deutlich zusehen. Die Probanden wurden instruiert jedes Wort, nachdem es verschwunden war und ein Fragezeichen in der Mitte auftauchte, auf der Tastatur nachzuschreiben. Dabei wurde die Tastatur während der gesamten Phase so von entsprechenden Schablone abgedeckt, dass nur die dafür nötigen Tasten benutzt werden konnten. Für das Nachschreiben des Items waren die Probanden an kein Zeitlimit gebunden und beendeten ihre Eingabe mit Hilfe der „Enter“-Taste. Auch eine Korrektur des Wortes mit Hilfe der „backspace“-Taste war möglich. Bevor das nächste Item aufleuchtete, wurde eine Pause von vier Sekunden eingehalten.

Zwischen dem Lernphase-Übungsdurchgang und der Lernphase fand eine kurze Pause statt, die für das Wechseln der Schablone und Einstellungen im Programm benötigt wurden.

## **Instruktionen:**

Lernphase-Übungsdurchgang:

„Du lernst gleich mit Hilfe des Notebooks eine Geheimsprache. Bevor es aber richtig losgeht, üben wir erstmal gemeinsam andere Wörter, die noch nichts mit der Geheimsprache zu tun haben, damit du es dann nachher alleine kannst.

Du siehst gleich auf dem Bildschirm ein Wort, was du dir merkst. Wenn es verschwunden ist, versuchst du es ohne Fehler nachzuschreiben. Falls du einen Fehler machst, kannst du mit Hilfe dieser Taste (zeige auf „backspace“-Taste) das Wort verbessern. Bist du mit dem Wort fertig, drückst du auf diese Taste (zeige auf „Enter“-Taste), damit das nächste Wort auftaucht. Hast du noch Fragen? Ok, dann fangen wir an.“

Lernphase:

„Schön, das hat sehr gut geklappt. Jetzt schreibst du die Wörter der richtigen Geheimsprache nach. Das sieht genauso aus wie gerade eben, nur dass die Wörter aus anderen Buchstaben bestehen. Damit du die Geheimsprache gut lernst, sind es sehr viele Wörter, die du gleich üben wirst. Es dauert, bis du damit fertig bist, aber es hört auf jeden Fall irgendwann auf. Also nicht zwischendrin aufgeben. Versuche die Wörter so gut wie möglich nachzuschreiben. Du schaffst das! Ist alles klar? Gut, dann siehst du gleich das erste Wort.“

### **4.3.3.2 Testphase**

**Aufbau und Durchführung:**

In dem Testphase-Übungsdurchgang wurden nacheinander fünf Übungssitems (siehe Tabelle 3) gezeigt, die aus den gleichen Buchstaben wie in dem Lernphase-Übungsdurchgang bestanden und auch unabhängig von einer Grammatik gebildet wurden. In der eigentlichen Testphase wurden die 32 Testitems nacheinander in zufälliger Reihenfolge präsentiert, und direkt danach in anderer zufälliger Reihenfolge noch einmal wiederholt, also insgesamt 64 mal gezeigt.

Die Items waren jeweils für vier Sekunden zu sehen. Danach erschien für drei Sekunden in der Mitte wieder ein Fragezeichen, auf der linken Seite ein Daumen, der nach unten zeigte, und auf der rechten Seite ein Daumen, der nach oben zeigte. Die Probanden wurden instruiert zu entscheiden, ob das Wort zu der Geheimsprache dazugehört oder nicht. Für den nach unten gerichteten Daumen, der bedeutet, dass das Wort nicht dazugehört, wurde die „A“-Taste betätigt, für den nach oben gerichteten Daumen die „Ä“-Taste. Diese beiden Tasten liegen auf der Tastatur am linken und rechten Rand und damit auf gleicher Höhe, wie die entsprechenden Daumen. Zwischen dem Betätigen der Tasten („A“ oder „Ä“) und dem Erscheinen des nächsten Items war eine Pause von einer Sekunde.

Auch in dieser Phase war die Tastatur so mit entsprechenden Schablonen abgedeckt, dass nur die nötigen Tasten benutzt werden konnten.

### **Instruktionen:**

Testphase-Übungsdurchgang:

„Super. Jetzt hast du schon viele Wörter der Geheimsprache geübt. Du siehst jetzt gleich weitere Wörter und sollst für jedes Wort entscheiden, ob es zu der Geheimsprache dazugehört oder nicht. Wir üben das wieder vorher zusammen, damit du weißt, was dich gleich erwartet.“

Hier siehst du gleich Wörter, die nichts mit der Geheimsprache zu tun haben. Wenn ein Wort aufgetaucht ist und wieder verschwindet, tauchen zwei Daumen auf. Auf der linken Seite zeigt der Daumen nach unten, das bedeutet, dass das Wort nicht dazugehört und du drückst die „A“-Taste, auf der rechten Seite zeigt er nach oben, das bedeutet, dass das Wort dazugehört und du drückst die „Ä“-Taste. Bei diesen Wörtern ist egal für welchen Daumen du dich entscheidest, da sie ja nichts mit der Geheimsprache zu tun haben. Gut, das schauen wir uns jetzt mal zusammen an.“

Testphase:

„Gut. Wichtig ist, dass du dir die Wörter, die jetzt auftauchen genau anschaust. Entscheide für jedes Wort, ob es zu der Geheimsprache dazugehört oder nicht. Hast du noch Fragen? Ok, dann geht es jetzt los.“

## 4.4 Datenerhebung der abhängigen Variable

Bei den Durchführungen der Tests werden von dem Computerprogramm der Testzeitpunkt (Datum), die Testdauer (Zeit der Anfänge und Enden der einzelnen Phasen) und die Nummer des Probanden gespeichert.

In der Lernphase wurde jedes vorgegebene Lernitem und der Versuch, dieses nachzuschreiben, erfasst und mit falsch oder richtig geschrieben bewertet. Dadurch sind die Anzahl der insgesamt gesehenen Wörter und die Anzahl der davon falsch geschriebenen Wörter auch in Abhängigkeit von der Länge der Items von jedem Kind nachzuvollziehen.

In der Testphase werden die Klassifikationen der einzelnen Kinder nach „ja, gehört dazu“, „nein, gehört nicht dazu“ und „keine Entscheidung“ (d.h. es wurde keine der beiden Tasten „A“ oder „Ä“ gedrückt) mit Beachtung der vier Eigenschaften der Testitems registriert. Zusätzlich wurden die Reaktionszeiten der Kinder – ab dem Einblenden des Items bis zur Klassifikationsentscheidung jedes Kindes in der Testphase – mit erfasst.

## 4.5 Auswertestrategie der Daten

Die Überprüfung der ersten drei Hypothesen, dass die nicht von einer Rechtschreibstörung betroffenen eine bessere Klassifikationsleistung als die betroffenen Kinder haben, dass Testitems mit hoher Fragmentstärke eher als zugehörig klassifiziert werden als Items mit niedriger Fragmentstärke und dass Testitems, die die Regeln der Grammatik befolgen (= grammatische Items), eher als zugehörig klassifiziert werden als Items, die die Regeln der Grammatik verletzen, erfolgt anhand Mittelwertbetrachtungen, während die weiteren Fragestellungen mit Hilfe von Korrelationen überprüft werden.

Um die ersten drei Hypothesen auf Signifikanz zu überprüfen, wird eine Varianzanalyse mit Messwiederholung zunächst mit den Effekten Grammatik, Fragmentstärke (Messwiederholungsfaktoren) und Rechtschreibstörung verwendet. Der p-Wert ist signifikant, wenn er einen Wert  $< 5\%$  erreicht.

Die Mittelwerte sind Angaben in Prozent der „ja, gehört dazu“-Antworten in der Testphase, wobei fehlende Antworten nicht berücksichtigt wurden. Die Mittelwerte werden aus dem Quotienten: Anzahl der „ja, gehört dazu“-Antworten geteilt durch die Summe der Anzahl aus „nein, gehört nicht dazu“- und „ja, gehört dazu“-Antworten, berechnet. Diese Mittelwerte werden mit den zugehörigen Standardabweichungen in Vierfeldertafeln mit den Feldern grammatische und nichtgrammatische Items, Items mit hoher Fragmentstärke und Items mit niedriger Fragmentstärke veranschaulicht.

Die vierte Hypothese, dass Kinder, die in höheren Klassenstufen sind, bessere Klassifikationsleistungen als Kinder in unteren Klassenstufen haben, wird zunächst auch anhand von Mittelwerten und Standardabweichungen in Vierfeldertafeln zu den einzelnen Klassenstufen überprüft. Daraufhin wird eine weitere Varianzanalyse mit Messwiederholung mit den Effekten Grammatik, Fragmentstärke, Rechtschreibstörung und zusätzlich Klassenstufe durchgeführt.

Um die weiteren Fragestellungen zu überprüfen werden explorative Analysen (Korrelationen und Effektstärken) verwendet, um die unangemessene mehrfache Anwendung von Signifikanztestüberprüfungen zu vermeiden. Zum Beispiel ist die Frage gestellt worden, ob die Leseleistung einen Einfluss auf die Klassifikationsleistung der Kinder hat. Falls dies zutrifft, könnte man vermuten, dass eine schwache Leseleistung mit einer schlechteren Klassifikationsleistung einhergeht. Solch ein Zusammenhang wird mit Hilfe des Korrelationskoeffizienten quantifiziert, der Werte zwischen -1 und 1 annehmen kann. Ein Wert von 0 besagt das kein Zusammenhang besteht. Positive Werte stehen für einen positiven Zusammenhang (beide Variablen werden in ihrer Ausprägung gemeinsam größer oder kleiner – wie in diesem Beispiel), negative Werte besagen einen negativen Zusammenhang, d. h. die eine Variable wird in ihrer Ausprägung kleiner, während die andere in ihrer Ausprägung größer wird.

Die Effektstärke wird zum Vergleich von Mittelwerten verwendet und ist definiert als standardisierte Mittelwertdifferenz (Mittelwert 1 minus Mittelwert 2, geteilt durch die gemeinsame Standardabweichung<sup>14</sup>). Die Effektstärke wird in dieser Arbeit für den

---

14 Die gemeinsame Standardabweichung ( $\sigma_{\text{pooled}}$ ) wird folgendermaßen berechnet:  $\sigma_{\text{pooled}} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$ . Die Formel für die Effektstärke (d) lautet:  $d = M_1 - M_2 / \sigma_{\text{pooled}}$ ; wobei M = Mittelwert.



Vergleich der beiden Gruppen (rechtschreibgestörte Kinder vs. Kontrollkinder) und für den Vergleich der Klassen untereinander benutzt. Wenn die Zielgröße des Gruppenvergleichs beispielsweise die Fehleranzahl in der Lernphase ist, wird die Differenz der beiden Mittelwerte gebildet, geteilt durch die gemeinsame Standardabweichung. Als Ergebnis erhält man also, um wie viel Standardabweichungen sich die Gruppen unterscheiden. Für die Bedeutung unterschiedlich großer Effektstärken sind von Cohen (1988) Richtwerte definiert worden, die sich allgemein etabliert haben.

Eine Effektstärke  $d$  von 0,2 wird nach Cohen (1988) als klein, eine von  $> 0,5$  als moderat bzw. mittelgroß und eine von  $> 0,8$  als groß bezeichnet. Eine bestimmte Effektstärke entspricht dabei genau einer bestimmten Korrelation, die als punkt-biseriale Korrelation berechnet wird – um einen Zusammenhang zwischen einer dichotomen und einer kontinuierlichen Variable zu bestimmen. Diese entspricht damit mathematisch genau der üblichen Korrelation nach Pearson. Einer Effektstärke von 0,2 entspricht eine Korrelation von 0,1, einer Effektstärke von 0,5 eine Korrelation von 0,24 und einer Effektstärke von 0,8 eine Korrelation von 0,37. Dieser Zusammenhang zeigt, dass die Effektstärke  $d$  mit zunehmender Korrelation etwas größer als das Doppelte der Korrelation ist.

## 5 Ergebnisse

### 5.1 Testphase

Zunächst wird die Klassifikationsleistung der Gesamtgruppe anhand einer Vierfeldertafel (Tabelle 4) veranschaulicht.

<b>Gesamtgruppe n = 54</b>	<b>Mittelwert <math>\pm</math> Standardabweichung</b>	
	<b>niedrige Fs</b>	<b>hohe Fs</b>
<b>nichtgrammatisch</b>	37,7 $\pm$ 20,9	45,1 $\pm$ 22,7
<b>grammatisch</b>	57,3 $\pm$ 18,3	60,4 $\pm$ 19,9

Tabelle 4: Klassifikationsleistung der Gesamtgruppe unterteilt in die verschiedenen Eigenschaften der Items, angegeben in Mittelwerten  $\pm$  Standardabweichung. Fs = Fragmentstärke

In der Gesamtgruppe wurden im Mittel 60,4 % der grammatischen Items mit hoher Fragmentstärke und nur 37,7 % der nichtgrammatischen mit niedriger Fragmentstärke als zugehörig klassifiziert. Die Items wurden also über bzw. unter dem Zufallslevel von 50 % klassifiziert, welches zu erwarten gewesen wäre, wenn die Probanden kein Wissen in der vorausgehenden Lernphase erworben haben oder dieses Wissen beim Klassifizieren der Items nicht anwenden konnten. Außerdem ist zu erkennen, dass die grammatischen Items mit niedriger Fragmentstärke mit 57,3% über dem Zufallslevel klassifiziert wurden, während die nichtgrammatischen Items mit hoher Fragmentstärke mit 45,1% unter dem Zufallslevel klassifiziert wurden. Der Einfluss der Grammatik auf die Klassifikationsleistung der Kinder ist anscheinend stärker als der der Fragmentstärke, d. h., dass grammatische Items eher als zugehörig klassifiziert werden als Items mit hoher Fragmentstärke. Diese Abweichungen der Ergebnisse vom

Zufallslevel von 50% sind in ihrer Größenordnung vergleichbar mit den Ergebnissen anderer AGL-Experimente.<sup>15</sup>

Die folgenden Vierfeldertafeln (Tabellen 5 und 6) zeigen die Klassifikationsleistungen der Kontrollkinder und der rechtschreibgestörten Kinder.

<b>Kontrollkinder n = 27</b>	<b>Mittelwert ± Standardabweichung</b>	
	<b>niedrige Fs</b>	<b>hohe Fs</b>
<b>nichtgrammatisch</b>	38,9 ± 23,4	45,9 ± 26,7
<b>grammatisch</b>	59,3 ± 18,9	62,3 ± 20,1

Tabelle 5: Klassifikationsleistung der Kontrollkinder unterteilt in die verschiedenen Eigenschaften der Items, angegeben in Mittelwerten ± Standardabweichung. Fs = Fragmentstärke

<b>Rechtschreibgestörte n = 27</b>	<b>Mittelwert ± Standardabweichung</b>	
	<b>niedrige Fs</b>	<b>hohe Fs</b>
<b>nichtgrammatisch</b>	36,4 ± 18,4	44,2 ± 18,4
<b>grammatisch</b>	55,3 ± 17,9	58,5 ± 19,8

Tabelle 6: Klassifikationsleistung der rechtschreibgestörten Kinder unterteilt in die verschiedenen Eigenschaften der Items, angegeben in Mittelwerten ± Standardabweichung. Fs = Fragmentstärke

Das Klassifikationsverhalten der rechtschreibgestörten Kinder über bzw. unter dem Zufallslevel nach den verschiedenen Itemeigenschaften entspricht dem der Kontrollkinder und damit auch dem der gesamten Gruppe. Anhand der Mittelwerte der Klassifikationsleistung ist eine minimale Tendenz der rechtschreibgestörten Kinder zu erkennen, alle Items eher abzulehnen. Sie ordnen die grammatischen Items (gr/hFs, gr/

<sup>15</sup> Zum Beispiel waren die Durchschnittswerte der Kontrollgruppe (nicht von einer Amnesie betroffene Erwachsene) im Experiment 1 von Knowlton & Squire (1996), in dem ein sehr ähnliches Testmaterial verwendet wurde, für grammatische Items hoher Fragmentstärke 64,6 %, grammatische Items niedriger Fragmentstärke 62,5 %, für nichtgrammatische Items hoher Fragmentstärke 54,9 % und für nichtgrammatische Items niedriger Fragmentstärke 28,1 %. Diese Ergebnisse wurden als aussagekräftig erachtet. Auch im Experiment 2 B von Knowlton & Squire (1994) waren die Ergebnisse aussagekräftig, mit den Durchschnittswerten der Kontrollgruppe für grammatische Items hoher Fragmentstärke von 61,5 %, für grammatische Items niedriger Fragmentstärke von 60,4 %, für nichtgrammatische Items hoher Fragmentstärke von 44,8 % und für nichtgrammatische Items niedriger Fragmentstärke von 39,6 %. In diesem Fall ist eine andere Grammatik verwendet worden.

nFs) um ca. 4 – 5 % geringer als die Kontrollkinder als „ja, gehört dazu“ ein, und die nichtgrammatischen Items (ngr/nFs, ngr/ hFs) lehnten sie um ca. 1 – 2 % eher ab.

Die Daten wurden im Folgenden mit einer Varianzanalyse mit Messwiederholung mit den Effekten Grammatik, Fragmentstärke und Rechtschreibstörung analysiert. Die Auswertung der Daten in der Varianzanalyse mit Messwiederholung mit allen Effekten ergibt folgendes.

Effekt	F-Wert	p-Wert
<b>Grammatik (gr)</b>	41,58	<b>&lt; 0,0001</b>
<b>Fragmentstärke (fs)</b>	4,29	<b>0,048</b>
<b>Rechtschreibstörung (rs)</b>	0,76	0,39
<b>Wechselwirkung (gr * fs)</b>	1,24	0,27
<b>Wechselwirkung (gr * rs)</b>	0,12	0,74
<b>Wechselwirkung (rs * fs)</b>	0,01	0,92
<b>Wechselwirkung (gr * fs * rs)</b>	0,01	0,94

Tabelle 7: Varianzanalyse mit Messwiederholung mit den Effekten Grammatik, Fragmentstärke und Rechtschreibstörung; fettgedruckte p-Werte sind < 5 % und damit signifikant

Es ist für die gesamte Gruppe ein hochsignifikanter Grammatikeffekt mit  $p < 0,0001$  und ein knapp signifikanter Fragmentstärkeneffekt mit  $p = 0,048$  vorhanden.

Die Rechtschreibstörung hat mit  $p = 0,39$  keinen signifikanten Effekt auf die Klassifikationsleistung. Es bestehen keine signifikanten Wechselwirkungen zwischen den einzelnen Messfaktoren (Grammatik, Fragmentstärke und schlechte Rechtschreibleistung).

Anhand dieser Ergebnisse kann die Hypothese 1, dass die nicht von einer Rechtschreibstörung betroffenen Kinder eine bessere Klassifikationsleistung haben als die betroffenen Kinder, nicht bestätigt werden. Es besteht kein Zusammenhang zwischen der Klassifikationsleistung der Kinder und der Rechtschreibstörung (im Sinne der Gruppenzugehörigkeit).

Die Ergebnisse dieser Analyse bestätigen die Hypothesen 2 und 3, dass Testitems mit hoher Fragmentstärke und grammatische Testitems eher als zugehörig klassifiziert werden als Items mit niedriger Fragmentstärke und nichtgrammatische Testitems.

Zusätzlich wurde die Hypothese formuliert, dass Kinder, die in höheren Klassenstufen sind, bessere Klassifikationsleistungen als Kinder in unteren Klassenstufen haben.

In Tabelle 8 sind die Mittelwerte und Standardabweichungen der einzelnen Klassen wiederum anhand des Vierfelderschemas gezeigt.

	Klasse 1 (n = 10)		Klasse 2 (n = 12)		Klasse 3 (n = 16)		Klasse 4 (n = 16)	
	nFs	hFs	nFs	hFs	nFs	hFs	nFs	hFs
<b>ngr</b>	38,1 ± 12,3	35,1 ± 17,9	48,9 ± 19,0	47,7 ± 19,6	34,2 ± 23,7	50,4 ± 19,3	32,4 ± 21,9	43,2 ± 29,6
<b>gr</b>	53,0 ± 21,5	46,5 ± 19,7	56,3 ± 18,7	56,2 ± 20,1	59,0 ± 18,6	68,4 ± 16,8	59,0 ± 16,9	64,2 ± 18,9

Tabelle 8: Klassifikationsleistung der einzelnen Klassenstufen unterteilt in die verschiedenen Eigenschaften der Items, angegeben in Mittelwerten ± Standardabweichung. ngr = nichtgrammatisch, gr = grammatisch, Fs = Fragmentstärke, n = niedrige, h = hohe

An den Durchschnittswerten des Antwortverhaltens der Kinder der einzelnen Klassenstufen ist zu erkennen, dass je höher die Klassenstufe, d. h. je älter die Kinder sind, desto besser klassifizieren sie die grammatischen Items mit hoher Fragmentstärke über bzw. die nichtgrammatischen Items mit niedriger Fragmentstärke unter dem Zufallslevel, bis zu einem in etwa gleich bleibendem Niveau der 3. und 4. Klasse.

Diese Erkenntnisse werden mit einer weiteren Varianzanalyse mit Messwiederholung mit dem zusätzlichen Faktor „Klasse“ überprüft. Dieser Faktor wurde in der ersten Varianzanalyse mit Messwiederholung nicht berücksichtigt, um die erste Varianzanalyse nicht mit Messfaktoren zu überfrachten.

In der folgenden Tabelle sind nur die Effekte gezeigt, die unter dem Aspekt des zusätzlichen Faktors „Klasse“ weitere Schlussfolgerungen ermöglichen. Nicht aufgelistete Effekte brachten keine weiteren Erkenntnisse.

	<b>F-Wert</b>	<b>p-Wert</b>
<b>Grammatikeffekt (gr)</b>	39,41	<b>&lt;0,0001</b>
<b>Fragmentstärkeneffekt (fs)</b>	2,84	0,105
<b>Wechselwirkung (gr * Klasse)</b>	1,96	0,15
<b>Wechselwirkung (fs * Klasse)</b>	2,92	0,056

Tabelle 9: Varianzanalyse mit Messwiederholung mit den Effekten Grammatik, Fragmentstärke, Klasse und Rechtschreibstörung; fettgedruckte p-Werte sind < 5 % und damit signifikant

Es gibt keine signifikante Wechselwirkung zwischen dem Klasseneffekt und dem Grammatik- oder Fragmentstärkeneffekt. Der Grammatikeffekt bleibt mit  $p < 0,0001$  erhalten, während sich der schon in der ersten Analyse nur knapp signifikante Fragmentstärkeneffekt aufteilt in einen Haupteffekt ( $p = 0,105$ ) und einen Wechselwirkungseffekt, welcher nur knapp über der Signifikanzgrenze ( $p = 0,056$ ) liegt.

Dieser tendenziell signifikante Wechselwirkungseffekt wird in einem Diagramm verdeutlicht. In dem folgenden Diagramm sind die Wechselwirkung Fragmentstärke und Klassenstufe dargestellt.

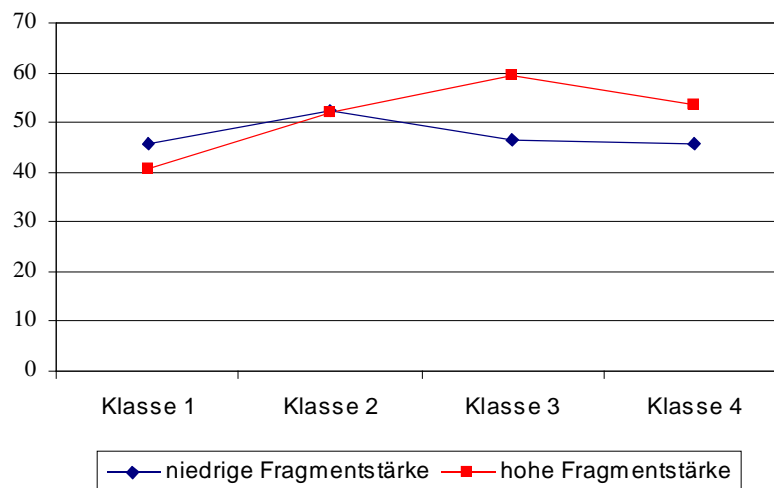


Abbildung 3: Wechselwirkung zwischen Fragmentstärke und Klassenstufe, Werte siehe Tabelle 10

	Klasse 1	Klasse 2	Klasse 3	Klasse 4
niedrige Fragmentstärke	45,57	52,59	46,59	45,7
hohe Fragmentstärke	40,82	51,95	59,41	53,71

Tabelle 10: Wechselwirkung zwischen Fragmentstärke und Klassenstufe

In diesem Diagramm ist zu erkennen, dass Kinder der ersten und zweiten Klasse Items mit niedriger Fragmentstärke genauso (oder eher besser) als zugehörig klassifizieren als Items mit hoher Fragmentstärke, während Kinder der dritten und vierten Klasse Items mit hoher Fragmentstärke eher als zugehörig klassifizieren als Items mit niedriger Fragmentstärke.

Anhand dieses Diagramms nehme ich an, dass Kinder aus unteren Klassen weniger Wissen über die Häufigkeitsverteilung von Fragmenten (so genanntes orthographisches Fragmentwissen) erwerben und anwenden können als Kinder aus oberen Klassen.

Grammatische Items werden dagegen von allen Kindern unabhängig von der Klassenstufe immer eher als zugehörig klassifiziert als nichtgrammatische Items. Zur Verdeutlichung zeigt das folgende Diagramm die nicht signifikante Wechselwirkung Grammatikalität und Klassenstufe.

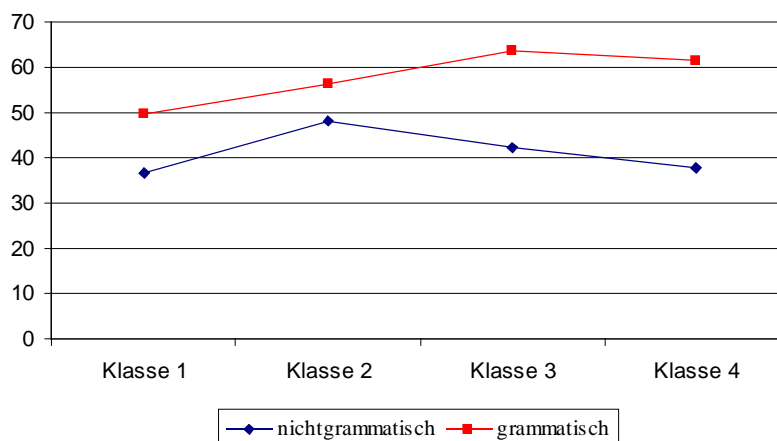


Abbildung 4: Wechselwirkung Grammatikalität und Klassenstufe, Werte siehe Tabelle 11

	<b>Klasse 1</b>	<b>Klasse 2</b>	<b>Klasse 3</b>	<b>Klasse 4</b>
<b>nichtgrammatisch</b>	36,63	48,3	42,29	37,81
<b>grammatisch</b>	49,75	56,24	63,66	61,6

Tabelle 11

Hier ist zu erkennen, dass nichtgrammatische Items in allen Klassenstufen immer eher als nicht zugehörig klassifiziert werden.

Zusammengefasst kann die vierte Hypothese, dass Kinder aus höheren Klassen besser implizit lernen als Kinder aus niedrigeren Klassen, bestätigt werden, da die Kinder der dritten und vierten Klasse mehr Wissen über die Fragmentstärke erwarben und besser beim Klassifizieren der Items anwenden konnten. Allerdings zeigt sich dieser Unterschied nur im Bereich des orthographischen Fragmentwissens, während Wissen über grammatische Eigenschaften in allen Klassenstufen gleich gelernt wird.

Des Weiteren wurde die Frage gestellt, ob die Leseleistung Einfluss auf die Klassifikationsleistung der Kinder hat. Bei Berücksichtigung der Leseleistung der Kinder als Kriterium für die Gruppenzugehörigkeit hätten keine in ihrer Größe geeigneten Gruppen gebildet werden können. Der Zusammenhang zwischen Leseleistung und Klassifikationsleistung der Kinder kann mittels Korrelation untersucht werden. Somit erfolgt die Auswertung der Variable Leseleistung mittels Korrelation mit den anderen Variablen – der Rechtschreibleistung und den vier verschiedenen Eigenschaften der Testitems.

Tabelle 12 zeigt die Korrelationen der Gesamtgruppe (n = 54) zwischen den Variablen Rechtschreibleistung (RST), Leseleistung (LES) und den vier Gruppen der Testitems, die unterteilt sind in grammatische Items mit hoher Fragmentstärke (gr/hFs), in nichtgrammatische Items mit hoher Fragmentstärke (ngr/hFs), in grammatische Items mit niedriger Fragmentstärke (gr/nFs) und in nichtgrammatische Items mit niedriger Fragmentstärke (ngr/nFs).



	<b>RST</b>	<b>LES</b>	<b>gr/hFs</b>	<b>ngr/hFs</b>	<b>gr/nFs</b>
<b>LES</b>	<b>0,66</b>				
<b>gr/hFs</b>	0,13	0,15			
<b>ngr/hFs</b>	0,05	0,19	<b>0,35</b>		
<b>gr/nFs</b>	0,19	0,12	0,15	0,23	
<b>ngr/nFs</b>	0,06	0,02	0,10	<b>0,54</b>	0,22

Tabelle 12: Pearson Korrelationskoeffizienten zwischen den Variablen Rechtschreibleistung (RST), Leseleistung (LES) und den vier Eigenschaften der Testitems; fett gedruckte r-Werte sind  $>0,30$  und haben einen korrelativ bedeutsamen Zusammenhang; RST = Rechtschreibleistung, LES = Leseleistung; gr = grammatisch, ngr = nichtgrammatisch, Fs = Fragmentstärke; h = hohe, n = niedrige.

Die Rechtschreibleistung der Kinder korreliert positiv ( $r = 0,66$ ) mit der Leseleistung. Dies ist auch von der Höhe des Koeffizienten erwartungsgemäß nach Schulte-Körne (2001, S. 155/156). Jedoch korreliert weder die Rechtschreibleistung noch die Leseleistung mit den vier Antwortkategorien der Testitems in der Testphase. Es bestehen keine korrelativen Zusammenhänge zwischen der Leseleistungen und dem Antwortverhalten der Kinder. Wie bereits in der Varianzanalyse gezeigt wurde, hat die Rechtschreibstörung mit  $p = 0,39$  keinen signifikanten Effekt auf die Klassifikationsleistung und wie anhand Tabelle 6 zu erkennen ist, gibt es auch keinen korrelativen Zusammenhang zwischen der Rechtschreibleistung und der Klassifikationsleistung.

Die restlichen Korrelationen beziehen sich auf Zusammenhänge der vier Gruppen der Items untereinander. Dabei sind vor allem die Items mit hoher Fragmentstärke und die nichtgrammatischen Items von Bedeutung. Bei der Klassifikation von Items mit hoher Fragmentstärke korreliert die Klassifikation von grammatischen Items positiv mit der Klassifikation von nichtgrammatischen Items ( $r = 0,35$ ). Das bedeutet, dass Kinder, die grammatische Items mit hoher Fragmentstärke als zugehörig („ja gehört dazu“) klassifizieren, insbesondere dann auch nichtgrammatische Items mit hoher Fragmentstärke als zugehörig klassifizieren. Bei der Klassifikation von nichtgrammatischen Items korreliert die Klassifikation von Items mit hoher Fragmentstärke positiv mit der Klassifikation von Items niedriger Fragmentstärke ( $r = 0,54$ ). Das bedeutet, dass Kinder, die nichtgrammatische Items mit hoher

Fragmentstärke als „nein, gehört nicht dazu“ klassifizieren auch nichtgrammatische Items mit niedriger Fragmentstärke eher als „nein, gehört nicht dazu“. Anscheinend haben die Kinder Schwierigkeiten Items mit hoher Fragmentstärke und nichtgrammatische Items in ihrem zweiten Merkmal (entweder grammatisch vs. nichtgrammatisch oder hohe vs. niedrige Fragmentstärke) zu unterscheiden, beziehungsweise die Merkmale, hohe Fragmentstärke oder nichtgrammatisch, ist für die Konstanz bei den Klassifikationsentscheidungen der Kinder ausschlaggebender. Es ist ein unplausibles Ergebnis für das keine naheliegende Erklärung gefunden werden kann.

## 5.2 Lernphase

In diesem Abschnitt 5.2 wird zunächst die Lernphase unter besonders interessanten Aspekten, die aus einer Vielzahl von möglichen Auswertungen ausgewählt wurden, analysiert. Desweiteren wird überprüft, ob diese Aspekte sich evtl. durch unterschiedliche Bedingungen in der Lernphase (z.B. Anzahl der insgesamt geschriebenen Items) auf das Klassifikationsverhalten der Kinder in der Testphase ausgewirkt haben. Auf diesen Zusammenhang zwischen der Lernphase und dem Klassifikationsverhalten in der Testphase wird in 5.3 eingegangen.

Die Items der Lernphase wurden von den Kindern auf der Tastatur nachgeschrieben. Für jedes Kind kann anhand der aufgezeichneten Daten nachvollzogen werden, welches Item mit welcher Buchstabenlänge wie oft falsch geschrieben wurde. Es wird anhand dieser Daten überprüft, wie viele Items mit wie vielen Buchstaben falsch geschrieben wurden und ob sich unter diesem Gesichtspunkt die beiden Gruppen (rechtschreibgestörte Kinder vs. Kontrollkinder) unterscheiden. Es werden also zunächst die in Abhängigkeit von ihrer Buchstabenlänge falsch geschriebenen Items im Gruppenvergleich überprüft (siehe 5.2.1).

Jedes Kind musste jedes der 23 Items in der Lernphase dreimal nachschreiben (insgesamt 69 Items). Wurde ein Item in der Lernphase einmal oder auch mehrmals falsch geschrieben, so wurde es noch ein viertes Mal präsentiert. Maximal müssten so 92 Items nachgeschrieben werden. Je mehr Fehler gemacht werden, desto mehr Items werden also gesehen. Somit werden die Daten der Lernphase unter dem Gesichtspunkt

der Anzahl der gesehenen Items wiederum im Gruppenvergleich ausgewertet (siehe 5.2.2).

Zusätzlich wird die Anzahl der bei allen vier Malen falsch geschriebenen Items im Gruppen- und Klassenvergleich betrachtet (5.2.3). Hier werden nur die Items berücksichtigt, die jedes Mal falsch geschrieben wurden. Es wird überprüft, ob sich die rechtschreibgestörten Kinder von den Kontrollkindern und ob sich die einzelnen Klassenstufen untereinander unterscheiden.

Die Ergebnisse dieser Betrachtungen fasse ich unter 5.2.4 kurz zusammen.

### 5.2.1 In Abhängigkeit von ihrer Buchstabenlänge falsch geschriebene Items im Gruppenvergleich

In der Lernphase wurden die Lernitems von den Kindern nachgeschrieben. In der folgenden Tabelle (Tabelle 13) ist dargestellt, wie viele Fehler die rechtschreibgestörten Kinder und die Kontrollkinder im Mittel bei dem Nachschreiben der Lernitems in Abhängigkeit von deren Länge machen.

<b>Falsch geschriebene Items mit:</b>	<b>Rechtschreibgestörte Mw ± Std.</b>	<b>Kontrollkinder Mw ± Std.</b>	<b>Effektstärke d</b>
<b>2 Buchstaben</b>	0,00 ± 0,00	0,04 ± 0,19	0,27
<b>3 Buchstaben</b>	0,56 ± 1,05	0,22 ± 0,58	0,39
<b>4 Buchstaben</b>	1,96 ± 2,70	0,96 ± 1,60	0,45
<b>5 Buchstaben</b>	5,22 ± 3,87	3,00 ± 3,35	0,61
<b>6 Buchstaben</b>	17,89 ± 6,57	10,93 ± 8,52	0,92

Tabelle 13: Anzahl der falsch geschriebenen Items in Abhängigkeit ihrer Länge im Gruppenvergleich, angeben in Mittelwert ± Standardabweichung und Effektstärke. Mw. = Mittelwert, Std. = Standardabweichung

Je länger die Lernitems sind, desto öfter werden sie von den Kindern falsch geschrieben. Dabei steigen die Falschreibungen mit der Buchstabenanzahl (Länge) der Items deutlich an. Dieser Anstieg ist bei den rechtschreibgestörten Kindern deutlich größer als bei den Kontrollkindern. Ab eine Länge der Items von 5 Buchstaben ist die Effektstärke der Fehleranzahl der rechtschreibgestörten Kinder im Vergleich zu der der Kontrollkindern mittelgroß ( $d = 0,61$ ) bei einer Itemlänge von 5 Buchstaben und groß ( $d$

= 0,92) bei einer Itemlänge von 6 Buchstaben. Kinder mit einer Rechtschreibstörung schreiben längere Items deutlich öfter falsch als die Kontrollkinder. Der Zusammenhang zwischen der Rechtschreibstörung und der Anzahl der Fehler beim Nachschreiben von längeren Items ist groß.

### **5.2.2 Anzahl der gesehenen Items im Gruppenvergleich**

Jedes der 23 Items wurde in der Lernphase drei Mal präsentiert (insgesamt 69 Items). Wer ein Item falsch schreibt, sieht es ein viertes Mal, egal wie oft es falsch geschrieben wurde (max. 92 Items). Da die rechtschreibgestörten Kinder mehr Items falsch schreiben, sehen sie im Mittel auch mehr Items – 79,67 Items ( $\pm$  3,94 Standardabweichung) – als die Kontrollkinder, die im Mittel nur 76,04 Items ( $\pm$  4,39 Standardabweichung) sehen. Die Effektstärke ist hier mit  $d = 0,87$  groß.

### **5.2.3 Anzahl der bei allen vier Malen falsch geschriebenen Items im Gruppen- und Klassenvergleich**

Am Ende der Lernphase konnten von den 54 Kindern 17 Kinder alle Items mindestens einmal richtig schreiben, 13 Kinder schrieben nur ein Item (von 23) alle vier Male falsch.

Tabelle 14 stellt dar, wie viele Items im Mittel in der Lernphase alle vier Mal falsch geschrieben wurden. Die Kinder sind dabei nach Klassen und Rechtschreibleistung aufgeteilt. Desweiteren sind die Effektstärken zum Gruppenvergleich und die Mittelwerte der Gesamtgruppe eingetragen worden. Tabelle 15 veranschaulicht die Effektstärken zwischen den einzelnen Klassen der Gesamtgruppe.

	<b>Anzahl der bei allen vier Malen falsch geschriebenen Items Mittelwert <math>\pm</math> Standardabweichung</b>			
	<b>Rechtschreib- gestörte</b>	<b>Kontrollkinder</b>	<b>Effektstärke d</b>	<b>gesamt (n=54)</b>
<b>Klasse 1</b>	5,0 $\pm$ 2,1	2,6 $\pm$ 1,8	1,22	3,8 $\pm$ 2,3
<b>Klasse 2</b>	3,8 $\pm$ 1,3	2,3 $\pm$ 1,6	1,01	3,1 $\pm$ 1,6
<b>Klasse 3</b>	2,1 $\pm$ 1,6	0,5 $\pm$ 0,5	1,34	1,3 $\pm$ 1,5
<b>Klasse 4</b>	0,8 $\pm$ 1,0	0,0 $\pm$ 0,0	1,02	0,4 $\pm$ 0,8

Tabelle 14: Anzahl der bei allen vier Mal falsch geschriebenen Items nach Klassenstufen im Gruppenvergleich und der Gesamtgruppe, angegeben in Mittelwert  $\pm$  Standardabweichung

	<b>gesamt (n=54) Effektstärke d</b>		
	<b>Kl. 1</b>	<b>Kl. 2</b>	<b>Kl. 3</b>
<b>Kl. 2</b>	0,37		
<b>Kl. 3</b>	1,32	1,15	
<b>Kl. 4</b>	2,02	2,11	0,79

Tabelle 15: Effektstärken zwischen den Klassenstufen bezüglich der Anzahl der bei allen vier Mal falsch geschriebenen Items. Kl. = Klasse

Anhand der Tabellen fällt nicht nur auf, dass die rechtschreibgestörten Kinder im Vergleich zu den Kontrollkinder mehr Items bei allen vier Versuchen falsch schreiben – die Effektstärke liegt hier bei etwa  $d = 1,3$  –, sondern auch, dass es deutliche Unterschiede zwischen den Klassenstufen gibt. Die Effektstärke der Gesamtgruppe beispielsweise zwischen der zweiten und dritten Klasse beträgt  $d = 1,15$ . Je höher die Klassenstufe, desto mehr Items werden mindestens einmal richtig geschrieben. Die 1. und 2. Klasse schrieben im Mittel noch 3,8 bzw. 3,1 Items bei allen vier Versuchen falsch, während die 3. und 4. Klasse nur 1,3 bzw. 0,4 Items falsch schrieben. Es fällt also auf, dass nicht nur die Rechtschreibleistung, sondern auch die Klassenstufe (bzw. das Alter) das Schreiben der Items beeinflusst.

## 5.2.4 Zusammenfassung

Zusammengefasst wird anhand dieser Ergebnisse deutlich, dass rechtschreibgestörte Kinder mehr Fehler vor allem beim Nachschreiben von langen Wörtern machen als die nicht von der Rechtschreibstörung betroffenen Kinder. Die rechtschreibgestörten Kinder sehen dadurch auch mehr Items, da jedes Item, welches mindestens einmal falsch geschrieben wurde, viermal statt dreimal gezeigt wird. Es schreiben neben den rechtschreibgestörten Kindern auch jüngere Kinder (1. und 2. Klasse) mehr Items alle vier Mal falsch.

Welche Auswirkungen diese Unterschiede in der Lernphase auf die Klassifikationsleistungen der Kinder haben, wird in 5.3 überprüft, um dadurch alle Fragen, die zur Lernphase gestellt wurden, beantworten zu können.

## 5.3 Zusammenhang zwischen der Lernphase und der Klassifikationsleistung in der Testphase

Tabelle 16 zeigt die Korrelation zwischen den Variablen Rechtschreibleistung (RST), Anzahl aller falsch geschriebenen Lernitems (Anz F), Anzahl der gesehenen Items (Anz I) und die Klassifikation der grammatischen Items mit hoher Fragmentstärke (gr/hFs). In der Tabelle sind nur die Korrelationen für grammatische Items mit hoher Fragmentstärke dargestellt, da sich bei den restlichen drei Gruppen von Testitems keine korrelativen Zusammenhänge zeigten.

	<b>RST</b>	<b>Anz F</b>	<b>Anz I</b>
<b>Anz F</b>	<b>- 0,50</b>		
<b>Anz I</b>	<b>- 0,49</b>	<b>0,96</b>	
<b>gr/hFs</b>	0,12	<b>- 0,33</b>	<b>- 0,35</b>

Tabelle 16: Pearson Korrelationskoeffizienten zwischen den Variablen Rechtschreibleistung (RST), Anzahl aller falsch geschriebenen Lernitems (Anz F), Anzahl der gesehenen Items (Anz I) und die Klassifikation der grammatischen Items mit hoher Fragmentstärke (gr/hFs).

Die Anzahl der gesehenen Items (Anz I) korreliert sehr hoch ( $r = 0,96$ ) mit der Anzahl der falsch geschriebenen Items (Anz F). Der Grund hierfür ist, dass die Anzahl der

gesehenen Items nur eine Dichotomisierung der Anzahl der falsch geschriebenen Items ist, da ein oder mehr Fehler pro Wort bedeutet, dass dieses in der Lernphase einmal mehr gezeigt wird. Die Korrelation ist so hoch, dass man sich in der Auswertung auf eine der beiden Variable beschränken kann.

Die Anzahl der in der Lernphase falsch geschriebenen Items korreliert negativ mit der Rechtschreibleistung ( $r = -0,50$ ) und negativ mit der Klassifikation der grammatischen Items mit hoher Fragmentstärke ( $r = -0,33$ ). Die Rechtschreibleistung korreliert allerdings nicht mit der Klassifikation der grammatischen Items mit hoher Fragmentstärke ( $r = 0,12$ ).

Inhaltlicher formuliert bedeutet das, dass Kinder, die schlechte Rechtschreiber sind, viele Fehler beim Nachschreiben der Items in der Lernphase machen, und Kinder, die viele Fehler in der Lernphase machen, grammatische Items mit hoher Fragmentstärke schlechter klassifizieren. Man kann allerdings nicht daraus schließen, dass Kinder mit einer schlechten Rechtschreibleistung grammatische Items mit hoher Fragmentstärke schlechter klassifizieren ( $r = 0,12$ ).

Beide negativen korrelativen Zusammenhänge, die Anzahl der falsch geschriebenen Items korreliert mit der Rechtschreibleistung ( $r = -0,50$ ) negativ und mit der Klassifikation der grammatischen Items mit hoher Fragmentstärke ( $r = -0,33$ ) negativ, sind mit  $r = -0,50$  und  $r = -0,33$  nur mittelgroß. Es ist hierdurch statistisch möglich, dass ein aufgrund dieser beiden Korrelationen zu erwartender Zusammenhang – hier hätte man einen negativen korrelativen Zusammenhang der Rechtschreibleistung mit der Klassifikation der grammatischen Items mit hoher Fragmentstärke erwartet – korrelativ nicht mehr bedeutsam ist, da bereits die beiden anderen Korrelationen schon in keinen perfekten Zusammenhang stehen.

In der folgenden Tabelle (Tabelle 17) ist nun die Anzahl der gesehenen Items (Anz I) herauspartialisiert, um zu überprüfen, ob diese Variable einen Gruppeneffekt zwischen den rechtschreibgestörten Kindern und den Kontrollkindern maskiert.

	<b>RST</b>	<b>Anz F</b>
<b>Anz F</b>	- 0,12	
<b>gr/hFs</b>	-0,06	-0,01

Tabelle 17: Pearson Korrelationskoeffizienten zwischen Rechtschreibleistung (RST), Anzahl aller falsch geschriebenen Items (Anz F) und Klassifikation der grammatischen Items mit hoher Fragmentstärke (gr/hFs).

Wird die Anzahl der gesehenen Items kontrolliert, sind alle Korrelationen unbedeutend, das heißt, dass diese Variable keinen Gruppeneffekt maskiert. Es gibt keinen Zusammenhang zwischen der Rechtschreibleistung und der Klassifikation der grammatischen Items mit hoher Fragmentstärke ( $r = -0,06$ ). Die rechtschreibgestörten Kinder haben also kein eventuell vorhandenes Defizit in ihrer Klassifikationsleistung durch das Sehen einer größeren Anzahl von Items in der Lernphase ausgeglichen.

Zusammengefasst bedeutet das, dass zwar ein korrelativer Zusammenhang zwischen der Anzahl aller falsch geschriebener Items und der Klassifikation der grammatischen Items mit hoher Fragmentstärke existiert, jedoch die Rechtschreibleistung, obwohl rechtschreibgestörte Kinder mehr Items falsch schreiben, in keinem korrelativen Zusammenhang zur Klassifikation der Testitems steht und auch keinen signifikanten Effekt in der Varianzanalyse mit Messwiederholung auf die Klassifikation hat.

## 5.4 Weitere Aspekte in der Auswertung der Testphase

In der Testphase wurden die 32 Testitems in einem Block in zufälliger Reihenfolge nacheinander gezeigt und klassifiziert und danach direkt im Anschluss noch mal alle 32 Testitems in einem zweiten Block in anderer zufälliger Reihenfolge nacheinander gezeigt und klassifiziert.

In Tabelle 18 sind die Mittelwerte  $\pm$  Standardabweichung der Klassifikationen der Testitems der ersten und zweiten Abrufphase in Vierfeldertafeln dargestellt.



	Abrufphase 1		Abrufphase 2	
	niedrige Fs	hohe Fs	niedrige Fs	hohe Fs
<b>nichtgrammatisch</b>	39,1 ± 22,1	45,7 ± 23,0	36,2 ± 23,9	44,5 ± 25,2
<b>grammatisch</b>	61,2 ± 21,3	61,7 ± 22,4	53,4 ± 22,1	59,0 ± 24,4

Tabelle 18: Klassifikationen der Testitems in der ersten und zweiten Abrufphase, angegeben in Mittelwerte ± Standardabweichung

Im Vergleich der jeweiligen Mittelwerte der beiden Phasen wurden in der zweiten Abrufphase alle Items tendenziell eher als „nein, gehört nicht dazu“ klassifiziert, wobei die einzelnen Differenzen minimal sind. Bei der größten Differenz von ca. 7,8 % der Mittelwerte der grammatischen Items mit niedriger Fragmentstärke liegt nur eine kleine Effektstärke von ca.  $d = 0,35$  vor, d. h., dass sich die Klassifikationsleistungen in den beiden Abrufphasen sich nicht voneinander unterscheiden.

In Tabelle 19 sind die Mittelwerte ± Standardabweichungen der Reaktionszeit der Kinder, die Testitems nach ihren Eigenschaften zu klassifizieren, aufgelistet.

	Mw ± Std der Reaktionszeiten in msec	
	Kontrollkinder	Rechtschreibgestörte
<b>gr/hFs</b>	2983,94 ± 1063,71	3065,65 ± 680,78
<b>ngr/hFs</b>	2965,52 ± 1126,69	3135,52 ± 708,22
<b>gr/nFs</b>	3216,02 ± 1043,16	3162,70 ± 630,43
<b>ngr/nFs</b>	3089,83 ± 1092,11	3197,61 ± 696,92

Tabelle 19 Reaktionszeiten der Kinder in der Testphase, angegeben in msec

Die Reaktionszeiten der Kinder werden ab der Einblendung des Items in der Testphase gemessen. Damit sind Zeiten bis 7 Sekunden möglich. Die Reaktionszeiten liegen im Mittel bei ca. 3 Sekunden. Die Durchschnittswerte der rechtschreibgestörten Kinder unterscheiden sich nicht von denen der Kontrollkinder. Die Streuung ist bei den Kontrollkindern jedoch deutlich größer. Die Standardabweichungen der rechtschreibgestörten Kinder streuen um ± 680, während die Streuung der Kontrollkinder um ± 1080 liegt. Dies ist ein Ergebnis, für das keine plausible Erklärung gefunden werden kann.

## 6 Diskussion

Hauptgegenstand dieser Arbeit ist, die Untersuchung des impliziten Lernens von Kindern mit und ohne Rechtschreibstörung im Umgang mit schriftsprachlichem Material und die Anwendung dieses implizit erworbenen Wissens.

Es folgt zunächst eine kurze Zusammenfassung der Ergebnisse dieser Arbeit anhand der aufgestellten Hypothesen.

Die Haupthypothese dieser Arbeit, dass nicht von einer Rechtschreibstörung betroffene Kinder besser klassifizieren als von einer Rechtschreibstörung betroffene Kinder, kann ich nicht bestätigen. Die Rechtschreibstörung hatte keinen signifikanten Einfluss auf die Klassifikationsleistung der Kinder.

Die vorliegende Arbeit zeigt, dass die Kinder in dem AGL-Experiment Wissen erworben und beim Klassifizieren der Testitems angewendet haben. Die Kinder erbrachten eine Klassifikationsleistung, die von dem Zufallslevel von 50 % abweicht. Ein Zufallslevel von 50 % wäre zu erwarten gewesen, wenn die Klassifikation der Testitems ohne vorheriges in der Lernphase erworbenes Wissen stattgefunden hätte. Die Ergebnisse der Kinder im AGL-Experiment sind von der Größenordnung ähnlich im Vergleich zu Ergebnissen anderer AGL-Experimente, die als aussagekräftig gelten (siehe Fußnote 15).

Desweiteren liegen in der Klassifikationsleistung der Gesamtgruppe ein hochsignifikanter Grammatikeffekt ( $p < 0,0001$ ) und ein knapp signifikanter Fragmentstärkeneffekt ( $p = 0,048$ ) vor. Dies bedeutet, dass die Kinder insbesondere Testitems, die den Regeln der Grammatik folgen und auch Testitems mit hoher Fragmentstärke als „ja, gehört dazu“ klassifizieren. Damit können die Hypothesen, dass grammatische Testitems und Testitems mit hoher Fragmentstärke eher als zugehörig klassifiziert werden, bestätigt werden.

Es wurde auch die Hypothese aufgestellt, dass Kinder aus niedrigeren Klassenstufen im Vergleich zu Kinder aus höheren Klassenstufen schlechter klassifizieren. Die Ergebnisse dieser Arbeit zeigen, dass Kinder aus höheren Klassenstufen (3. und 4.

Klasse) im Vergleich zu Kindern aus niedrigeren Klassenstufen (1. und 2. Klasse) insbesondere Items nach ihrer Fragmentstärke tendenziell besser klassifizieren: In der Varianzanalyse mit Messwiederholung ist die Wechselwirkung ( $p = 0,056$ ) zwischen den Variablen Fragmentstärke und Klasse fast signifikant. Dies lässt vermuten, dass Kinder aus höheren Klassenstufen besser implizites orthographisches Fragmentwissen erwerben und anwenden können.

Zusätzlich wurde den Fragen nachgegangen, ob die Leseleistung und Unterschiede in Leistungen und Alter in der *Lernphase* Einfluss auf das Klassifikationsverhalten haben. Die Ergebnisse zu dieser Thematik wurden unter mehreren Aspekten ermittelt, und werden deswegen erst später in dem sich darauf beziehenden Kontext erläutert.

Zunächst gehe ich jedoch darauf ein, inwiefern ein impliziter Lernprozess bei den Kindern angenommen werden kann.

Das Ergebnis des in dieser Untersuchung durchgeführten AGL-Experimentes ähnelt den Ergebnissen der AGL-Experimente von Knowlton & Squire (1996) und der Replikation von Kinder & Assmann (2000), die die gleiche Grammatik verwendet hatten. Beide Experimente wurden allerdings im Gegensatz zu dieser Untersuchung mit Erwachsenen durchgeführt. Ein AGL-Experiment mit Kindern durchzuführen ist eine besondere Herausforderung, da insbesondere durch die anstrengende Lernphase – dem Nachschreiben von sehr vielen Items – die Geduld und Konzentration der Kinder schnell überstrapaziert werden kann. Deswegen wurde in der Lernphase das Nachschreiben der Lernitems auf maximal vier Mal pro Item beschränkt, im Gegensatz zu der Anforderung in der Lernphase im Experiment von Kinder & Assmann (2000), in der die Lernitems von den Probanden solange nachgeschrieben wurden, bis alle Items mindestens einmal richtig geschrieben worden sind. Es wurden also alle Items so lange gezeigt bis sie alle korrekt wiedergegeben werden konnten. Je intensiver Lernphasen sind, desto eher besteht die Gefahr, dass Wissen über die Items explizit erworben wird (Gomez 1997). Da in dieser Untersuchung die Lernphase im Vergleich zu anderen Experimenten (z. B. Kinder 2000, Kinder & Assmann 2000, Kinder & Shanks et al. 2003, Gomez, Gerken & Schvaneveldt 2000), in denen implizite Lernvorgänge angenommen wurden, weniger intensiv war, kann auch in diesem Experiment implizites

Lernen der Kinder angenommen werden. Voraussetzung hierfür ist auch die Instruktion am Anfang des Experimentes (Reber 1967, 1989), die ermöglichen sollte, dass sich die Lernitems ohne das Wissen, dass die Items nach einem Regelsystem konstruiert wurden, eingeprägt werden. In diesem Experiment wurden die Kinder darüber informiert, dass sie eine Geheimsprache erlernen werden und dafür die Wörter der Geheimsprache nachschreiben. Zu Anfang der Testphase wurde ihnen gesagt, dass sie nun für jedes einzelne Wort entscheiden sollen, ob es zur Geheimsprache gehört oder nicht. Ich nehme somit zusammengefasst an, dass die Kinder in dem vorliegenden AGL-Experiment implizit gelernt haben.

Dieses implizit erworbene Wissen wird von den Kindern für die Klassifikation der Testitems angewendet. Signifikanten Einfluss auf die Klassifikationsentscheidungen der Kinder hatten von der Grammatik vorgegebene Merkmale der Items ( $p < 0,0001$ ) und zu einem geringeren Anteil die Fragmentstärke der Items ( $p = 0,048$ ). Die Hypothesen, dass grammatische Eigenschaften und die Fragmentstärke der Items die Klassifikationsleistung der Kinder beeinflussen, können bestätigt werden.

Die Items, die die Regeln der Grammatik befolgen, wurden signifikant mehr als zugehörig klassifiziert als die Items, die die Regeln der Grammatik verletzen. Es kann hieraus allerdings nicht sicher geschlossen werden, dass die Kinder tatsächlich implizit Regeln der Grammatik abstrahiert haben, also abstraktes Regelwissen erwarben. Wie bereits im theoretischen Teil dieser Arbeit erläutert, gibt es Unstimmigkeiten, welche Art von Wissen in einem AGL-Experiment gelernt wird und zu den Klassifikationen der Probanden über bzw. unter dem Zufallslevel führen. In den regelbasierenden Annahmen mit den Befürwortern des abstrakten Regelwissens (z. B., Reber, 1967,1989) und des kontinuierlichen Regelwissens (Mathews et al., 1989; Mathews & Roussel, 1997) wird angenommen, dass implizites Regelwissen einerseits aus Wissen über ein abstraktes komplexes Regelsystem der Grammatik und andererseits aus Wissen über selbst aufgestellte Regeln, die sich je nach ihrer Häufigkeit in ihrer Stärke beeinflussen, besteht. Die ähnlichkeitsbasierenden Annahmen befürworten dagegen entweder Wissen über spezifische, komplette Items (z. B.: Brooks, 1978; Brooks & Vokey, 1991,1992) oder Fragmentwissen (z. B. Perruchet & Pacteau, 1990; Redington & Charter, 1996;

Servan-Schreiber & Anderson, 1990). Durch die Erkenntnisse von Aaron, Wilczynski & Keetay (1998) spielt das Lernen kompletter Wörter im Schriftspracherwerb – wie auch im AGL-Experiment, da hierfür die Lernphase in der Regel nicht intensiv genug ist – nur eine untergeordnete Rolle. Fragmentwissen ist dahingegen, wie bereits erläutert, für den Schriftspracherwerb (Aaron, Wilczynski & Keetay, 1998) und die Klassifikationsleistung im AGL-Experiment (Pacton et al., 2001) von größerer Bedeutung. Vermutungen, dass beides, Regelwissen und Fragmentwissen (Knowlton & Squire, 1994,1996; Meulemanns & Van der Linden, 1997), die Klassifikationsentscheidungen beeinflusst, werden in der vorliegenden Untersuchung bestätigt. Es ist prinzipiell möglich, dass Merkmale der Items, wie z. B. Anfangsbuchstaben, Buchstabenwiederholungen und die Häufigkeit von einzelnen Buchstaben die Klassifikation mit beeinflussen, da sie ja durch die Grammatik vorgegeben werden. Allerdings ist es nicht möglich, die verwendeten Items der Lern- und Testphase im Voraus nach allen möglichen Eigenschaften zu kontrollieren. Wenn nun mit Hilfe einer Multiplen Regressionsmethode („multiple regression methode“ von Lorch & Myers, 1990) die Ergebnisse der Klassifikationen im Nachhinein mit ausreichend vielen dieser einzelnen Merkmale analysiert werden, kann der Grammatikeffekt, wie es Kinder & Assmann (2000) letztendlich gelungen ist, seine statistische Signifikanz verlieren. Diese Multiple Regressionsmethode wurde erstmals von Johnstone & Shanks (1999) und ein weiteres Mal von Kinder & Assmann (2000) auf AGL-Ergebnisse angewendet, wobei sie, soweit mir bekannt ist, in keinem späteren AGL-Experiment Verwendung für die Auswertung fand. Diese Methode hat sich in der Analyse von Daten aus AGL-Experimenten nicht etabliert. Es ist nicht Zielstellung dieser Arbeit sich mit den theoretischen Aspekten des möglichen Wissenserwerbes zu beschäftigen, sondern es steht im Vordergrund herauszufinden, ob sich rechtschreibgestörte Kinder von nicht rechtschreibgestörten Kindern in ihrer Klassifikationsleistung unterscheiden. Da die Ergebnisse dieser Arbeit zeigen, dass sie sich darin nicht voneinander unterscheiden, kann daraus geschlossen werden, dass bei den Kindern der gleiche Wissenserwerb stattgefunden hat. Es wurde deswegen, und da sich diese Methode bei späteren AGL-Experimenten nicht etabliert hat, zusammenfassend entschieden, die Daten dieses AGL-Experimentes nicht mit der

Multiplen Regressionsmethode auszuwerten.

Neben den von der Grammatik vorgegebenen Merkmalen hat auch die Fragmentstärke der Items die Klassifikationsleistungen der Kinder signifikant beeinflusst. Orthographisches Fragmentwissen basiert auf Wissen von Buchstabeneinheiten in Bi- und Trigrammen, welches von der Häufigkeit ihres Vorkommens in der Schriftsprache bedingt wird (Aaron, Wilczynski & Keetay, 1998). Da die Fragmentstärke auch ein Maß für die Häufigkeitsverteilung der in der Lernphase implizit erlernten Bi- und Trigramme ist und diese Eigenschaft einen signifikanten Einfluss auf die Klassifikation der Testitems bei den Kindern hatte, vermute ich, dass die Kinder in dem AGL-Experiment implizites orthographisches Fragmentwissen erworben haben. Übertragen auf den Schriftspracherwerb bestätigt dies die Annahme von Aaron, Wilczynski & Keetay (1998), dass orthographisches Fragmentwissen durch Auseinandersetzung mit schriftsprachlichem Material implizit erworben wird.

Von einer Rechtschreibstörung betroffene Kinder unterscheiden sich laut den Ergebnissen dieser Untersuchung beim impliziten Erlernen und Anwenden dieses orthographischen Fragmentwissens nicht von nicht betroffenen Kindern.

Anhand der Ergebnisse kann allerdings vermutet werden, dass Kinder in höheren Klassenstufen (3. und 4. Klasse) besser implizites orthographisches Fragmentwissen erwerben und anwenden können als Kinder in niedrigeren Klassenstufen (1. und 2. Klasse). In der Varianzanalyse mit Messwiederholung ist die Wechselwirkung ( $p = 0,056$ ) zwischen den Variablen Fragmentstärke und Klasse tendenziell signifikant.

In der vorliegenden Arbeit wurde der Frage nachgegangen, ob die Leseleistung diesen impliziten Lernprozess beeinflusst. Bei Berücksichtigung der Leseleistung der Kinder als Kriterium für die Gruppenzugehörigkeit hätten keine in ihrer Größe geeigneten Gruppen gebildet werden können. Der Zusammenhang zwischen Leseleistung und Klassifikationsleistung der Kinder kann allerdings mittels Korrelation untersucht werden. Die Leseleistung hatte keinen korrelativen Zusammenhang mit der Klassifikationsleistung der Testitems. Somit konnte die Vermutung, dass eine Lesedefizit das implizite Lernen der Kinder beeinträchtigen könnte, in dieser Arbeit nicht bestätigt werden:

In der vorliegenden Untersuchung stellte ich mir auch die Frage, inwiefern unterschiedliche Leistungen in der Lernphase einen Einfluss auf die Klassifikationsleistung haben. Es ist davon auszugehen, dass Kinder mehr Schwierigkeiten beim Nachschreiben der Items haben als Erwachsene. In den mir bekannten mit Erwachsenen durchgeführten AGL-Experimenten wird auf mögliche Unterschiede in den Leistungen der Probanden in der Lernphase und deren Auswirkung auf die Klassifikationsleistung nicht eingegangen. Rechtschreibgestörte Kinder schreiben mehr Items in der Lernphase falsch nach (Effektstärke von  $d = 1,3$ ) und sehen dadurch auch mehr Items, da jedes Item, welches einmal oder mehrere Male falsch geschrieben wurde, statt dreimal viermal präsentiert wurde. Es gibt allerdings keinen korrelativen Zusammenhang zwischen der Rechtschreibleistung und der Klassifikationsleistung. Auch gibt es keinen signifikanten Effekt der Rechtschreibleistung in der Varianzanalyse mit Messwiederholung. Es besteht jedoch ein negativer korrelativer Zusammenhang ( $r = -0,33$ ) zwischen der Klassifikation der grammatischen Items mit hoher Fragmentstärke (gr/hFs) nach „ja, gehört dazu“ und den Kindern, die in der Lernphase viele Items falsch schreiben und damit auch viele Items gesehen haben. Wird die Anzahl der gesehenen Items herauspartialisiert, ist auch kein korrelativer Zusammenhang zwischen der Rechtschreibleistung und der Klassifikationsleistung festzustellen. Ein eventuell vorhandenes Defizit der rechtschreibgestörten Kinder, die Items in der Testphase zu klassifizieren, wurde also nicht durch das Sehen von mehr Items in der Lernphase ausgeglichen. Auch Kinder aus niedrigeren Klassenstufen (Klasse 1 und 2) schreiben in der Lernphase mehr Items falsch als Kinder aus höheren Klassenstufen (Klasse 3 und 4). Anhand dem Diagramm (Abbildung 3, 5.1) zur tendenziell signifikanten Wechselwirkung ( $p = 0,056$ ) zwischen Klassenstufe und Fragmentstärke kann ein unterschiedliches Antwortverhalten der Klassenstufen vermutet werden (siehe oben).

Im folgenden werde ich auf leichte Einschränkungen in der Interpretierbarkeit der Ergebnisse eingehen, die sich durch Größe der einzelnen nach Klassenstufe und Rechtschreibleistung unterteilten Gruppen ergeben. Durch die Auswahl einer möglichst breiten Altersgruppe wurden die einzelnen Gruppen der Klassen von der Anzahl der teilnehmenden Kinder sehr klein. Es ist sehr unwahrscheinlich, dass die auf die Kinder

einwirkenden Umweltbedingungen, vor allem in Form von expliziten und impliziten Kontakt mit der Schriftsprache, für die Kinder aller Klassenstufen gleich sind. Durch die beinahe signifikante Wechselwirkung zwischen Fragmentstärke und Klassenstufe ( $p = 0,056$ ) kann vermutet werden, dass die Kinder aus niedrigeren Klassenstufen sich schlechter implizites orthographisches Fragmentwissen aneignen als Kinder aus höheren Klassenstufen. Es ist vorstellbar, dass sich eventuell ein Effekt nur bei der 3. und 4. Klasse zwischen der mittleren Klassifikationsleistung von einer Rechtschreibstörung betroffene Kinder und der mittleren Klassifikationsleistung von nicht betroffenen Kindern ergibt, also die schlechtere Klassifikationsleistung der niedrigeren Klassen (1. und 2. Klasse) in dieser Analyse nicht eingeschlossen wird. Um dies zu überprüfen, werden die Mittelwerte der Klassifikation der grammatischen Items mit hoher Fragmentstärke (bei diesen Items mit diesen Merkmalen ist der höchste Effekt zu erwarten) nach „ja, gehört dazu“ der 3. und 4. Klasse im Gruppenvergleich mit der zugehörigen Effektstärke berechnet. Die nicht betroffenen Kinder der dritten Klasse klassifizieren 70 % dieser Items, während die betroffenen Kinder der dritten Klasse nur 67 % als zugehörig klassifizieren. Zwischen diesen mittleren Klassifikationsleistungen ergibt einen kleiner Effekt von 0,22. In der 4. Klasse klassifizieren allerdings die betroffenen Kinder 67 % der grammatischen Items mit hoher Fragmentstärke als zugehörig, und haben damit eine bessere Klassifikationsleistung als die nicht betroffenen Kinder, die nur 61 % dieser Items als zugehörig klassifizieren. In der 4. Klasse ist demnach sogar ein negativer Effekt von -0,31 zwischen den mittleren Klassifikationsleistungen. Zusammenfassend ist hieran zu erkennen, dass sich die rechtschreibgestörten Kinder auch nicht von den nicht rechtschreibgestörten Kindern in ihren Klassifikationsleistungen unterscheiden, wenn nur Kinder der 3. und 4. Klasse betrachtet werden.

Als Frage offen, ob die Entwicklung des impliziten Lernens der Kinder (unabhängig von der Rechtschreibstörung) in einem AGL-Experiment von der Variable „Alter“ mit Hilfe einer Längsschnittuntersuchung deutlicher beeinflusst werden kann. Zu berücksichtigen ist bei Längsschnittuntersuchungen allerdings die Tatsache, dass durch die mehrmalige Durchführung eines AGL-Experimentes an ein und derselben Person, und damit dem potentiellen Wissen dieser Person, dass eine Testphase folgen wird, die



auf die Erkenntnisse in der Lernphase aufbaut, sich beispielsweise durch explizites Bemühen in der Lernphase die Lernitems zu erlernen, die Klassifikationsleistung beeinflusst wird (Turner & Fischler, 1993) und damit die Vergleichbarkeit der Ergebnisse nicht möglich ist. Allerdings testeten beispielsweise Knowlton & Squire in ihren Experimenten von 1996 die Versuchspersonen mehrmals ohne dies in Frage zu stellen.

Zusammenfassend haben die Kinder in dem AGL-Experiment implizites orthographisches Fragmentwissen und von der Grammatik vorgegebene Merkmale der Lernitems gelernt und konnten dieses implizite Wissen in der Testphase zur Klassifikation der Testitems anwenden. Dies steht auch im Einklang zu den AGL-Experimenten von Knowlton & Squire (1996), Meulemann & Van der Linden (1997) und Kinder & Assmann (2000).

Insgesamt bestätigt die vorliegende Untersuchung den Erwerb und die Anwendung von implizitem Wissen bei schriftsprachlichem Material. Sie steht damit im Einklang mit den Untersuchungen von Aaron, Wilczynski & Keetay (1998) und Pacton et al. (2001).

Der vermutete Einfluss einer Rechtschreibstörung auf den Erwerb und die Anwendung von implizitem Wissen auf schriftsprachliches Material konnte ich nicht bestätigen.

Als Ausblick auf künftige Forschung stellen sich mir die folgenden, im Rahmen der Diskussion bereits entwickelten, weiterführenden Fragestellungen:

1. Ist es möglich, AGL-Experimente in einer Längsschnittstudie durchzuführen um die Entwicklung von implizitem Lernen bei Kindern zu erforschen?
2. Lassen sich die Ergebnisse dieser Untersuchung auch mit AGL-Experimenten, in denen andere künstlichen Grammatiken verwendet werden, erheben?

## 7 Zusammenfassung

Die vorliegende Arbeit untersucht implizites Lernen von orthographischen Fragmentwissen von Kindern der 1. - 4. Klasse im Umgang mit schriftsprachlichem Material und die Bedeutung dieses impliziten Wissenserwerbes bezüglich der Rechtschreibstörung.

Die Fähigkeit, implizit zu lernen und implizit erworbenes Wissen anzuwenden, wurde anhand eines Artificial Grammar Learning-Experimentes (AGL) untersucht. Hierfür werden Buchstabenreihen (Items) mit Hilfe einer künstlichen Grammatik gebildet. Ein Teil dieser Items (Lernitems) wird in einer Lernphase den Probanden präsentiert. Die Probanden haben die Aufgabe, diese Lernitems nachzuschreiben und sich einzuprägen, allerdings ohne dass die Probanden wissen, dass die Items einem bestimmten Regelsystem folgen und, dass dieses Wissen in einer sich anschließenden Testphase des Experimentes auf weitere Items angewendet werden muss. Die Items (Testitems), die für die folgende Testphase verwendet werden, entsprechen zur Hälfte den Regeln der Grammatik, die andere Hälfte verstößt gegen die Regeln der Grammatik. In dem vorliegenden AGL-Experiment wurden die Items der Testphase zusätzlich nach ihrer Fragmentstärke kontrolliert. Die Fragmentstärke ist ein Maß für die Häufigkeit von Buchstaben-Bigrammen und -Trigrammen in den Lernitems. In dieser Testphase werden die Probanden instruiert, zu entscheiden, ob die Testitems der Grammatik entsprechen oder nicht. Wenn die Probanden in der Lernphase implizites Wissen über die Regeln der Grammatik bzw. über Eigenschaften der Items, die durch diese Regeln vorgeschrieben werden, erwerben, werden die Items der Testphase, die den Regeln der Grammatik folgen, über dem Zufallslevel von 50 % als „ja gehört dazu“ klassifiziert im Gegensatz zu denen, die die Regeln der Grammatik verletzen.

Mit Hilfe dieser Klassifikationsleistung der Probanden in der Testphase des AGL-Experimentes können Rückschlüsse gezogen werden, ob implizit Wissen erworben und angewendet wurde und welcher Art dieser Wissenserwerb ist.

Die 54 teilnehmenden Kinder wurden unter Berücksichtigung der Rechtschreibleistung in Extremgruppen eingeteilt und unter Berücksichtigung der Intelligenz und

Klassenstufe parallelisiert. Darüber hinaus wurde die Leseleistung erfasst, so dass in dieser Untersuchung den Hypothesen,

- dass nicht von einer Rechtschreibstörung betroffene Kinder besser klassifizieren als davon betroffene Kinder,
- dass Testitems mit hoher Fragmentstärke eher als zugehörig klassifiziert werden,
- dass grammatische Testitems eher als zugehörig klassifiziert werden,
- dass Kinder in niedrigeren Klassenstufen im Vergleich zu Kinder in höheren Klassenstufen schlechter klassifizieren

und den Fragen, ob die Leseleistung und Unterschiede in Leistungen und Alter in der Lernphase Einfluss auf das Klassifikationsverhalten haben, nachgegangen wird.

Die Ergebnisse dieser Arbeit lassen vermuten, dass Kinder aus höheren Klassenstufen (3. und 4. Klasse) implizites orthographisches Fragmentwissen besser erlernen und anwenden können als Kinder aus niedrigeren Klassenstufen (1. und 2. Klasse). Unterschiedliche Voraussetzungen der Kinder in Klassenstufe, Lese- und Rechtschreibleistung und damit verbundene unterschiedliche Leistungen in der Lernphase hatten dabei keinen Einfluss auf die Klassifikationsleistung der Kinder.

Die Ergebnisse des AGL-Experimentes zeigen, dass die Fragmentstärke und auch die durch die Grammatik vorgegebenen Eigenschaften der Testitems (Regeln) auf das Klassifikationsverhalten der Kinder einen signifikanten Einfluss haben. Die Kinder haben in der Auseinandersetzung mit schriftsprachlichen Material unter Laborbedingungen implizit gelernt und konnten dieses erworbene Wissen, als sogenanntes Fragmentwissen und Regelwissen auf schriftsprachliches Material anwenden.

Nach den Ergebnissen dieses AGL-Experimentes unterscheiden sich die Klassifikationsleistungen der rechtschreibgestörten Kinder nicht signifikant von der Klassifikationsleistungen der Kinder ohne Rechtschreibstörung. Anhand der vorliegenden Untersuchung wird somit angenommen, dass rechtschreibgestörte Kinder im Vergleich zu Kindern ohne Rechtschreibstörung nicht in ihrem impliziten Lernen beeinträchtigt sind. Insbesondere unterscheiden sich nicht von einer

Rechtschreibstörung betroffene Kinder nicht von betroffenen Kindern bezüglich des Erwerbs von impliziten orthographischen Fragmentwissens, welches nach Stand der Literatur auch eine bedeutende Rolle beim Schriftspracherwerb spielt.

## 8 Anhang

Klasse	Paar- nummer	Rechtschreibgestörte			Kontrollkinder		
		RST	LES	CFT	RST	LES	CFT
<b>1. Klasse</b>	1	37	40	48	52	40	48
	2	39	36	51	52	59	52
	3	39	48	52	55	57	54
	4	39	39	59	52	48	59
	5	40	36	62	52	54	63
<b>2. Klasse</b>	6	39	25	35	52	52	37
	7	39	44	56	68	58	58
	8	29	50	47	54	29	47
	9	27	38	50	54	52	50
	10	37	30	49	64	59	49
	11	38	36	54	50	43	52
<b>3. Klasse</b>	12	34	47	39	53	49	41
	13	27	27	47	57	53	47
	14	39	29	47	51	50	52
	15	37	45	48	56	59	53
	16	38	42	49	52	61	47
	17	38	45	52	51	56	54
	18	40	47	54	58	57	55
	19	37	45	54	63	57	59
<b>4. Klasse</b>	20	40	41	52	52	67	54
	21	38	53	47	54	52	48
	22	40	41	49	69	54	47
	23	38	39	51	69	63	51
	24	30	34	55	52	59	55
	25	30	43	56	50	41	55
	26	36	34	57	57	66	58
	27	29	47	56	57	46	57

Tabelle 20: **Individuelle Ergebnisse der teilnehmenden Kinder:** RST = Rechtschreibtest, CFT = Culture-Fair-Intelligenz-Test, LES = Lesetest; in diesen Spalten sind T-Werte angegeben

	Anzahl	Rechtschreibgestörte			Anzahl	Kontrollkinder		
		Ø RST	Ø LES	Ø CFT		Ø RST	Ø LES	Ø CFT
<b>Klasse 1</b>	5	38,8	39,8	54,4	5	52,6	51,6	55,2
<b>Klasse 2</b>	6	34,83	37,17	48,5	6	57	48,83	48,83
<b>Klasse 3</b>	8	36,25	40,86	48,75	8	55,13	55,25	51
<b>Klasse 4</b>	8	35,13	41,5	52,88	8	57,5	56	53,13
<b>insgesamt</b>	27	36,07	40,04	50,96	27	55,78	53,37	51,93

Tabelle 21: **Durchschnittliche Leistungen der Kinder mit und ohne Rechtschreibstörung unterteilt in Klassenstufen in Rechtschreib-, Lese- und Intelligenztests.** RST = Rechtschreibtest, CFT = Culture-Fair-Intelligenz-Test, LES = Lesetest; in diesen Spalten sind durchschnittliche T-Werte angegeben

## 9 Literatur

### Tests:

Birkel, Peter: Weingartener Grundwortschatz Rechtschreib-Test für erste und zweite Klassen (WRT 1+). Hogrefe-Verlag für Psychologie GmbH & Co. KG, Göttingen 1995

Birkel, Peter: Weingartener Grundwortschatz Rechtschreib-Test für zweite und dritte Klassen (WRT 2+). Hogrefe-Verlag für Psychologie GmbH & Co. KG, Göttingen 1995

Küspert, P. & Schneider W.: Würzburger-Leise-Leseprobe (WLLP). © Hogrefe-Verlag für Psychologie GmbH & Co., Göttingen 1998

May, Peter; Vieluf, Ulrich & Malitzky, Volkmar: Diagnose orthographischer Kompetenz, zur Erfassung der grundlegenden Rechtschreibstrategien mit der Hamburger-Schreib-Probe (HSP). vpm (Verlag für pädagogische Medien Hamburg), 5. unveränderte Auflage 2000

Mayringer, Heinz & Wimmer, Heinz: Salzburger Lese Screening für die Klassenstufen 1-4 (SLS). Verlag Hans Huber 2003

Müller, Rudolf: Diagnostischer Rechtschreibtest (DRT 3+), Leistungstest für 3. Klassen. In: Ingenkamp, Karlheinz (Hsg.): Deutsche Schultests. Beltz Verlag, Weinheim und Basel, 3. Auflage 1997 © Beltz Test GmbH, Göttingen

Weiß, Rudolf H. & Osterland, Jürgen: Grundintelligenztest Skala 1 (CFT 1). Hogrefe-Verlag für Psychologie GmbH & Co. KG, Göttingen 1997

Weiß, Rudolf H.: Grundintelligenztest Skala 2 (CFT 20). Hogrefe-Verlag für Psychologie GmbH & Co. KG, Göttingen 1987

### Artikel:

Aaron, P. G. (1997): The impending demise of the discrepancy formula. Review of Educational Research 67 (4), 461-502; In: Deimel, W. (2002): Diagnostik der Lese-Rechtschreibstörung. In: Legasthenie: Zum aktuellen Stand der

Ursachenforschung, der diagnostischen Methoden und der Förderkonzepte. Hsg. Schulte-Körne, Gerd, ISBN 3-89911-000-5, Verlag Dr. Dieter Winkler, Bochum 2002

- Aaron, P. G.; Wilczynski, Susan & Keetay, Victorio (1998): The anatomy of word-specific memory. In: Hulme, C. & Joshi, R. M. (Eds.): Reading and spelling: Development and disorders. Mahwah, NJ 1998 : Erlbaum, 405-419
- Abrams, M. & Reber, A. S. (1989): Implicit learning: Robustness in the face of psychiatric disorders. *Journal of Psycholinguistic Research* 1989, 17, 425–439
- Altmann, Gerry T. M.; Dienes, Zoltán & Goode, Alastair (1995): Modality Independence of Implicit Learned Grammatical Knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition* 1995, Vol. 21, No. 4, 899-912
- Anderson, J. R. (1976): Language, memory, and thought. Hillsdale, NJ: Lawrence Erlbaum
- AWMF-Leitlinien der Deutschen Gesellschaft für Kinder- und Jugendpsychiatrie und -psychotherapie; <http://www.uni-duesseldorf.de/WWW/AWMF/II/028-017.htm>
- Brooks, L. R. (1978): Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 170–211); Hillsdale, NJ 1978: Erlbaum
- Brooks, L.R. & Vokey, J.R. (1991): Abstract analogies and abstracted grammars: Comments on Reber (1989) and Matthews et al (1989). *Journal of Experimental Psychology: General* 1991, 120, 316-323
- Cassar, Marie & Treiman, Rebecca (1997): The beginning of orthographic knowledge: Children's knowledge of double letters in word. *Journal of Educational Psychology* 1997, Vol. 89, No. 4, 631-644
- Cleeremans, Axel; Destrebecqz, Arnaud & Boyer, Maud: Implicit learning: news from the front, *Trends in Cognitive Sciences*, Vol. 2, No. 10, 1998, 406-416



- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2<sup>nd</sup> ed.). Hillsdale NJ: Erlbaum
- Coltheart, M. & Rastle, K. (1994): Serial Processing in Reading Aloud: Evidence for Dual-Route Models of Reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20(6), 1197 – 1211
- Deimel, W. (2002): Diagnostik der Lese-Rechtschreibstörung. In: *Legasthenie: Zum aktuellen Stand der Ursachenforschung, der diagnostischen Methoden und der Förderkonzepte*. Hsg. Schulte-Körne, Gerd, ISBN 3-89911-000-5, Verlag Dr. Dieter Winkler, Bochum 2002
- Dienes, Z. & Altmann, G. (1997): Transfer of implicit knowledge across domains: How implicit and how abstract?. In: *How implicit is implicit learning?.* D. Berry (Ed.). Oxford University Press 1997
- Dienes, Altmann, Kwan & Goode (1995): Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1995, 21, 1322–1338
- Dulany, D. E.; Carlson, R. A. & Dewey, G. I. (1984): A case of syntactical learning and judgment: How conscious and how abstract?. *Journal of Experimental Psychology: General* 1984, 113, 541–555
- Dulany, D. E.; Carlson, R. A. & Dewey, G. I. (1985): On consciousness in syntactic learning and judgment: A reply to Reber, Allen, and Regan. *Journal of Experimental Psychology: General* 1985, 114, 33–49
- Ehri, L. C. (1986): Sources of difficulty in learning to spell and read. In M. L. Wolraich & D. Routh (Eds.), *Advances in developmental and behavioral pediatrics* (Vol. 7, pp. 121–195). Greenwich 1986, CT: JAI Press
- Esser, G. & Schmidt, M. (1993): Die langfristige Entwicklung von Kindern mit Lese-Rechtschreibschwäche. *Zeitschrift für Klinische Psychologie* 1993, 22, 100-116
- Facoetti, A. & Molteni, M. (2001): The gradient of visual attention in developmental dyslexia. *Neuropsychologia* 2001, 39, 352–357

- Fawcett, A. J., Nicolson RI, Maclagan F (2001). Cerebellar tests differentiate between groups of poor readers with and without IQ discrepancy. *Journal of Learning Disabilities* 34 (2), 119-135, In: Deimel, W. (2002): Diagnostik der Lese-Rechtschreibstörung. In: Legasthenie: Zum aktuellen Stand der Ursachenforschung, der diagnostischen Methoden und der Förderkonzepte. Hsg. Schulte-Körne, Gerd, ISBN 3-89911-000-5, Verlag Dr. Dieter Winkler, Bochum 2002
- Gomez, Rebecca L. (1997): Transfer and Complexity in Artificial Grammar Learning. *Cognitive Psychology* 1997, 33, 154-207
- Gomez, Rebecca L.; Gerken, Louann (1999): Artificial grammar learning by 1-year-old leads to specific and abstract knowledge. *Cognition* 1999, 70, 109-135
- Gomez, Rebecca L.; Gerken, Louann & Schvaneveldt, Roger W. (2000): The basis of transfer in artificial grammar learning. *Memory & Cognition* 2000, 28 (2), 253-263
- Gomez, Rebecca L.& Lakusta, Laura (2004): A first step in form-based category abstraction by 12-month-old infants. *Developmental Science* 7:5 (2004), pp 567–580
- Gomez, R. & Schvaneveldt, R. W. (1994): What is learned from artificial grammars? Transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1994, 20, 396-410
- Habib, M. (2000): The neurological basis of development dyslexia. An overview and working hypothesis. *Brain* (2000), 123, 2373- 2399
- Haffner, J.; Zehran-Hartung, C.; Pfüller, U.; Parzer, P.; Strehlow, U. & Resch, F. (1998): Auswirkungen und Bedeutung spezifischer Rechtschreibprobleme bei jungen Erwachsenen – empirische Befunde einer epidemiologischen Stichprobe. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 1998, 26, 124-135
- Heiervang E.; Stevenson J.; Lund A. & Hugdahl K. (2001): Behaviour problems in children with dyslexia. *Nordic Journal of Psychiatry* 2001, Volume 55, Number 4, 1 August 2001, pp. 251-256(6)

- Johnstone, T. & Shanks, D. R. (1999): Two mechanisms in implicit grammar learning? Comment on Meulemans and Van der Linden (1997); *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1999, 25, 524–531
- Johnstone, Theresa & Shanks, David R.(2001): Abstractionist and Processing accounts of Implicit Learning. *Cognitive Psychology* 2001, 42, 61-112
- Kinder, Annette (2000): The knowledge acquired during artificial grammar learning: Testing the prediction of two connectionist models. *Psychological Research* 2000, 63, 95-105
- Kinder, Annette & Assmann, Anja (2000): Learning artificial grammars: No evidence for the acquisition of rules. *Memory & Cognition* 2000, 28 (8), 1321-1332
- Kinder, Annette; Shanks, David R.; Cock, Josephine & Tunney, Richard J. (2003): Recollection, Fluency, and the Explicit/ Implicit Distinction in Artificial Grammar Learning. *Journal of Experimental Psychology: General* 2003, Volume 132 (4), p 551-565
- Knowlton, B. J.; Ramus, S. & Squire, L. R. (1992): Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science* 1992, 3, 172–179
- Knowlton, Barbara J. & Squire, Larry R. (1994): The Information acquired during Artificial Grammar Learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 1994, Vol. 20, No. 1, 79-91
- Knowlton, Barbara J. & Squire, Larry R.(1996): Artificial Grammar Learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory and Cognition* 1996, Vol. 22, No. 1,169-181
- Lennox, C. & Siegel, S. L. (1996): The development of phonological rules and visual strategies in average and poor spellers. *Journal of Experimental Child Psychology* 1996, 62, 60–83

- Lorch, R. F., & Myers, J. L. (1990): Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 1990, 16, 149-157
- Manza, L. & Reber, A.S. (1997): Representing artificial grammars: Transfer across stimulus forms and modalities. In D. C. Berry (Ed.): *How implicit is implicit learning?* (pp. 73–106), Oxford: Oxford University Press 1997
- Mathews, R. C.; Buss, R. R.; Stanley, W. B.; Blanchard-Fields, F.; Cho, J. R. & Druhan, B. (1989): Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory and Cognition* 1989, 15, 1083–1100
- Mathews, R. C. & Roussel, L. G. (1997): Abstractness of implicit knowledge: A cognitive evolutionary perspective. In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 162-194). Oxford: Oxford University Press 1997
- Meulemans, T. & Van der Linden, M. (1997): Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 1997, 23, 1007-1028
- Meyer, M. S. (2000): The ability-achievement discrepancy: Does it contribute to an understanding of learning disability? *Educational Psychology Review* 12 (3), 315-337; In: Deimel, W. (2002): *Diagnostik der Lese-Rechtschreibstörung. In: Legasthenie: Zum aktuellen Stand der Ursachenforschung, der diagnostischen Methoden und der Förderkonzepte.* Hsg. Schulte-Körne, Gerd, ISBN 3-89911-000-5, Verlag Dr. Dieter Winkler, Bochum 2002
- Murphy, Kristina; McKone Elinor & Slee, Judith (2003): Dissociation between implicit and explicit memory in children: The role of strategic processing and the knowledge base. *Journal of Experimental Child Psychology* 2003, 84, 124-165
- Naito, Mika (1990): Repetition Priming in Children and Adults: Age-Related Dissociation between Implicit and Explicit Memory. *Journal of Experimental Child Psychology* 1990, Vol. 50, No. 3, 462-84

- Pacton, Sébastien; Perruchet, Pierre; Fayol, Michel & Cleeremans, Axel (2001): Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General* 2001, Vol. 130, No. 3, 401-426
- Perruchet, P. (1994): Defining the knowledge units of a synthetic language: Commentary on Vokey and Brooks (1992). *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1994, 20, 223–228.
- Perruchet, P.; Gallego, J. & Pacteau, C. (1992): A reinterpretation of some earlier results for abstractiveness of implicitly acquired knowledge. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 1992, 44A, 193–210
- Perruchet & Pacteau (1990): Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge?. *Journal of Experimental Psychology: General* 1990, 119, 264–275
- Pothos, Emmanuel M. & Kirk, Jane (2004): Investigating Learning Deficits associated with Dyslexia. © 2004 John Wiley & Sons, Ltd. *DYSLEXIA* 10: 61–7 (2004), Published online in Wiley InterScience ([www.interscience.wiley.com](http://www.interscience.wiley.com)). DOI: 10.1002/dys.266
- Rack, John P. (1985): Orthographic and phonetic coding in developmental dyslexia. *British Journal of Psychology* 1985, vol. 76, issue 3, p 325
- Reber, Arthur S. (1967): Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour* 1967, 6, 855-119
- Reber, Arthur S. (1989): Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology: General* 1989, Vol. 118 (3), 219-235
- Reber, Arthur S. (1993): *Implicit Learning and Tacit Knowledge. An Essay on the Cognitive Unconscious*. Oxford University Press
- Reber, Arthur S.; Walkenfeld, Faye F. & Hernstadt, Ruth (1991): Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1991 Sep Vol 17(5) 888-896

- Redington, Martin & Charter, Nick (1996): Transfer in Artificial Grammar Learning: A Reevaluation. *Journal of Experimental Psychology: General* 1996, Vol. 125, No. 2, 123-138
- Schulte-Körne, G.; Deimel W. & Remschmidt H. (1997): Die Bedeutung von phonologischer Dekodierfähigkeit und orthographischem Wissen für die Rechtschreibfähigkeit Erwachsener. *Zeitschrift für Klinische Psychologie* 1997 26, 210-217
- Schulte-Körne, Gerd (2001): *Lese-Rechtschreibstörung und Sprachwahrnehmung*. ISBN 3-89325-790-X, Waxmann Verlag GmbH, 2001, Hsg. D.H. Rost
- Servan-Schreiber, E. & Anderson, J. R. (1990): Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 1990, 16, 592-608.
- Shanks, D. R.; Johnstone, T. & Staggs, L. (1997): Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology* 1997, 50A, 216-252
- Shaywitz, S. E., Escobar, M. D., Shaywitz B. A., Fletcher J. M., Makuch R. (1992): Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine* 326 (3), 145-150; In: Deimel, W. (2002): Diagnostik der Lese-Rechtschreibstörung. In: *Legasthenie: Zum aktuellen Stand der Ursachenforschung, der diagnostischen Methoden und der Förderkonzepte*. Hsg. Schulte-Körne, Gerd, ISBN 3-89911-000-5, Verlag Dr. Dieter Winkler, Bochum 2002
- Spoehr, K. T. & Smith, E. E. (1975): The role of orthographic and phonotactic rules in perceiving letter patterns. *Journal of Experimental Psychology: Human Perception and Performance* 1975, 104, 21-34
- Steffler, Dorothy J. (2001): Implicit cognition and spelling development. *Developmental Review* 2001, 21, 168-204

- Turner, C. W.; Fischler, I. S. (1993): Speeded tests of implicit knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition* 1993, 19, 1165-1177
- Vicari, Stefano; Marotta, Luigi; Menghini, Deny; Molinari, Marco & Petrosini, Laura (2003): Implicit learning deficit in children with developmental dyslexia. *Neuropsychologica* 2003, 41, 108-114
- Vokey, J. R. & Brooks, L. R. (1992): The salience of item knowledge in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 1992, 18, 328-344.
- Waber, Deborah P.; Marbus, David J.; Forbes, Peter W.; Bellinger, David C.; Weiler, Michael D.; Sorensen Lisa G. & Curran, Tim (2003): Motor sequence learning and reading ability: Is poor reading associated with sequencing deficits?. *Journal of Experimental Child Psychology* 2003, 84, 338-354
- Whittlesea, B. W. A. (1997): Production, evaluation, and preservation of experiences: Constructive processing in remembering and performance tasks. In G. H. Bower (Ed.): *The psychology of learning and motivation* (Vol. 37, pp. 211–264). New York: Academic Press.
- Whittlesea, B. W. A. & Dorken, M. D. (1993): Incidentally, things in general seem to be particularly determined: An episodic processing account of implicit learning. *Journal of Experimental Psychology: General* 1993, 122, 227-248.
- Whittlesea, B. W. A. & Wright, R. L. (1997): Implicit (and explicit) learning: Acting adaptively without knowing the consequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 1997, 23, 181-200.
- Witt K., Nühsman A., Deuschl G. (2002): Intact artificial grammar learning in patients with cerebellar degeneration and advanced Parkinson's disease. *Neuropsychologica* 2002, 40, 1534-1540

# 10 Appendix

## 10.1 Verzeichnis der akademischen Lehrer

Meine akademischen Lehrer waren:

Herren/Frauen in Marburg:

Adamkiewicz	Lang
Basler	Lenz
Baum	Lill
Becker	Maier
Boudriot	Maisch
Christiansen	Mandrek
Eilers	Moll
Fuchs	Mueller
Gerdes	Mutters
Görg	Neubauer
Gotzen	Oertel
Griss	Radsak
Grundmann	Renschmidt
Grzeschik	Renz
Gudermann	Richter
Herrmann-Lingen	Schäfer
Jungclas	Schmidt
Koolman	Schrader
Krause	Schulte-Körne
Krieg	Seitz
Kroll	Vogelmeier



Wagner

Westermann

Weiler

Wulff

Werner

## 10.2 Danksagung

Mein besonderer Dank gilt Herrn Prof. Dr. G. Schulte-Körne für die Überlassung dieses Themas, die Betreuung als Doktorvater und für seine freundliche und engagierte Bereitschaft mich jederzeit bei der Anfertigung dieser Arbeit zu unterstützen.

Herrn Deimel danke ich insbesondere für die gute Zusammenarbeit während der Test- und Auswertungsphase und die konstruktiven Hilfestellungen in der Durchsicht dieser Arbeit.

Bei Herrn Bartling bedanke ich mich sehr für das Programmieren des Artificial Grammar Tests und freundliche Betreuung von informatischer Seite.

Insbesondere danke ich den Lehrern und Schülern der Grundschule Breiter Hagen in Bad Wildungen, die mir mit ihrer lebenswerten Kooperation und der zahlreichen Teilnahme an den Tests die Datenerhebung ermöglichten und damit wesentlich zu dieser Arbeit beitrugen.

Herrn Benjamin Streeck danke ich für seine wertvolle und lebenswerte Hilfe und Geduld beim Erstellen der Grafiken und Formatieren dieser Arbeit, sowie beim Korrekturlesen.

Meinen Eltern danke ich für ihre lebenswerte Unterstützung, die mir das Studium und die Doktorarbeit ermöglicht hat.