



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Entwicklung eines Verfahrens zur Curriculum-basierten
Messung des Leseverständnisses

verfasst von

Hannes Wittek

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2014

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Ao. Univ.-Prof. Mag. Dr. Alfred Schabmann

Danksagung

Allen voran möchte ich meinem Diplomarbeitsbetreuer Ao. Univ.-Prof. Mag. Dr. Alfred Schabmann für die kritischen Diskussionen in der Anfangsphase dieses Projekts, seinen pointierten fachlichen Rat bei der Durchführung und Auswertung, sowie die großen Freiräume, die er mir bei der gesamten Umsetzung einräumte danken. Die Zusammenarbeit hätte auf sachlicher wie auch menschlicher Ebene nicht besser sein können!

Großer Dank gebührt auch meinen Eltern, die mich in meiner gesamten Bildungskarriere in meinen Interessen bekräftigt und in der Erreichung meiner Ziele unterstützt haben.

Weiters danke ich meiner Freundin Sonja, die für mich immer ein offenes Ohr hatte und mich zur passenden Zeit mit sanfter Hand zur Weiterarbeit animierte. Außerdem bedanke ich mich bei meinem Bruder und all meinen Freunden, ohne die meine Studienzeit nicht zu solch einer aufregenden und wunderbaren Lebensphase geworden wäre!

Zuletzt möchte ich mich auch noch bei den Direktorinnen und Lehrerinnen der teilnehmenden Volksschulen in Wiener Neustadt und Umgebung für ihr Interesse und die Mitarbeit an diesem Projekt bedanken. Ohne ihr Engagement und das ihrer fleißigen Schülerinnen und Schüler hätte dieses Projekt nicht realisiert werden können.

Anmerkung zur Gender- gerechten Formulierung:

Als Autor dieser Diplomarbeit ist es mir ein wichtiges Anliegen, die optimale Lesbarkeit des Textes zu gewährleisten und für inhaltliche Klarheit zu sorgen. Die strikte Anwendung geschlechtsneutraler Formulierungen, wie „der/die SchülerIn“ oder „den LehrerInnen“, würde dies jedoch maßgeblich erschweren. Bezeichnungen für Gruppen von Personen, wie „Schüler“, „Lehrer“, usw. beziehen sich daher stets gleichermaßen auf männliche und weibliche Vertreter dieser Gruppen.

Inhaltsverzeichnis

1	Einleitung	8
2	Theoretische Grundlagen - Leseverständnis	10
2.1	Begriffsbestimmung	10
2.2	Komponenten des Leseverständnisses	11
2.2.1	Wortverständnis	11
2.2.2	Satzverständnis	12
2.2.3	Textverständnis	13
2.2.3.1	Kohäsionsmittel und Inferenzen	13
2.2.3.2	Textstrukturen und Diskursformen	14
2.3	Hierarchische Ordnung der Teilprozesse	15
2.3.1	Hierarchieniedere Prozesse	15
2.3.2	Hierarchiehohe Prozesse	15
2.3.3	Allgemeine kognitive Voraussetzungen	16
2.4	Problemlokationen des Leseverständnisses	16
2.4.1	Probleme auf der Wortebene	16
2.4.2	Probleme auf der Satzebene	17
2.4.3	Probleme auf der Textebene	17
2.5	Verlauf von Leseverständnisschwierigkeiten	17
3	Kriterien für den neuen Lesetest	18
3.1	Curriculum-Based Measurement	19
3.1.1	Charakteristiken des CBM	19
3.1.2	Schwierigkeiten in der Implementierung von CBM	20
3.2	Die MAZE Technik	21
3.3	Messung der Lesbarkeit	23
3.3.1	Expertenrating	23
3.3.2	Lesbarkeitsindex	23
4	Forschungsfragen und Hypothesen	25
5	Empirischer Teil	29
5.1	Auswahl der Stichprobe	29
5.2	Stichprobenbeschreibung	29
5.3	Testmaterial	30
5.4	Versuchsplan	33
5.5	Datenerfassung	35
6	Auswertung	35
6.1	Ergebnisse	35
6.1.1	Leistungsverlauf	36
6.1.2	Paralleltestreliabilität des MAZE Tests	39
6.1.3	Retestrelabilität des MAZE Tests	40
6.1.4	Leistungsveränderungen und Gruppenvergleiche	41
6.1.5	Itemanalyse	45
6.1.6	Validität	48
7	Diskussion	49
7.1	Etablierung im Unterricht	49
7.2	Auffälligkeiten im Leistungsverlauf	50
7.2.1	Qualität der MAZE Texte	50
7.2.2	Einfluss von Geschlecht und Sprachkenntnis	52

7.3	Ausblick	53
8	Literaturverzeichnis	55
9	Anhang	58
9.1	Zusammenfassung (deutsch)	58
9.2	Abstract (english)	59
9.3	Fragenkatalog für Schüler	60
9.4	MAZE – Instruktion	61
9.5	Probandencode und Übungs-MAZE Text „Nilpferd“	63
9.6	MAZE Text „Hase“	64
9.7	MAZE Text „Wolf“	66
9.8	MAZE Text „Haushuhn“	68
9.9	MAZE Text „Elefant“	70
9.10	Lesbarkeitsindex für Texte aus Schulbüchern	72
9.11	Lesbarkeitsindex für MAZE Texte	72
9.12	LEBENS LAUF	73

Abbildungsverzeichnis

Abb. 1: Die Zergliederung eines Satzes in einzelne Phrasen (aus Lenhard, 2013, S. 19)	12
Abb. 2: Mittelwerte der MAZE Texte von T1 bis T4	36
Abb. 3: Mittelwerte für das oberste und unterste Quartil	37
Abb. 4.1: MW für Mädchen, getrennt nach Sprachkenntnis	44
Abb. 4.2: MW für Buben getrennt nach Sprachkenntnis	44

Tabellenverzeichnis

Tab. A: Untersuchungsplan	34
Tab. 1: Mittelwerte der MAZE Texte von T1 bis T4	36
Tab. 2: Mittelwerte für das oberste und unterste Quartil	37
Tab. 3: Test auf Normalverteilung für das unterste (1) und oberste (3) Quartil	38
Tab. 4: Mauchly Test auf Sphärizität für das unterste (1) und oberste (3) Quartil	38
Tab. 5: Pearson-Korrelationen der MAZE Texte (bei Erstvorgabe)	39
Tab. 6: Pearson-Korrelationen der MAZE Wiederholungsmessungen	40
Tab. 7: Retestrelabilität der einzelnen MAZE Texte	40
Tab. 8: Test auf NV mit den Faktoren Geschlecht und Mehrsprachigkeit	42
Tab. 9: Mauchly-Test auf Sphärizität	42
Tab. 10: Mittelwerte der Gesamtstichprobe (N=105)	43
Tab. 11.1: MW nach Sprachkenntnis	44
Tab. 11.2: MW nach Geschlecht	44
Tab. 12.1: MW für Mädchen, getrennt nach Sprachkenntnis	45
Tab. 12.2: MW für Buben, getrennt nach Sprachkenntnis	45
Tab. 13.1: Item-Lösungswahrscheinlichkeiten für MAZE Text "Hase"	46
Tab. 13.2: Item-Lösungswahrscheinlichkeiten für MAZE Text "Wolf"	46
Tab. 13.3: Item-Lösungswahrscheinlichkeiten für MAZE Text "Haushuhn"	47
Tab. 13.4: Item-Lösungswahrscheinlichkeiten für MAZE Text "Elefant"	47
Tab. 14: durchschnittliche Lösungswahrscheinlichkeit der Items pro Text; Werte nach Ausschluss der Ausreißer befinden sich in Klammern	48

1 Einleitung

Lesen ist eine der wohl wichtigsten Kulturfähigkeiten unserer Gesellschaft und so ist es auch nicht weiter verwunderlich, dass sie vom Tag des ersten Schulbesuchs an erlernt, geübt und mit der Zeit in fast allen Lebensbereichen angewendet wird. Kompetentes Lesen beschreibt jedoch nicht nur die einfache Umwandlung von Schriftzeichen in Lautäußerungen. Es bedeutet vielmehr einem Text seinen Inhalt, seine Bedeutung zu entnehmen und mit dem bisherigen Wissen zu verknüpfen, sowie daraus Schlussfolgerungen zu ziehen um konkrete Probleme zu lösen. Verschiedene internationale Schulleistungsuntersuchungen, allen voran die PISA-Studien (*Programme for International Student Assessment*) und PIRLS (*Progress in International Reading Literacy Study*) der OECD, haben Österreich in den letzten Jahren im Bereich der Lesekompetenz ein eher verbesserungswürdiges Zeugnis ausgestellt. Bergmüller und Wintersteller (2012) verglichen die Lesefähigkeiten österreichischer Schüler der 4.Schulstufe mit 13 anderen Nationen, die Österreich geographisch oder ökonomisch nahe stehen. In diesem Vergleich fällt Österreich besonders durch seinen hohen Anteil (20%) von leseschwachen Schülern, der Lese-„Kompetenzstufe 1“ und darunter, auf (S. 15). Auf dieser Stufe beherrscht ein Schüler lediglich die basale Fähigkeit leicht auffindbare Informationen im Text herauszusuchen. Zum Vergleich: In Finnland und Russland beträgt der Anteil leseschwacher Schüler nur 8%. Noch dazu weist Österreich den niedrigsten Lese-Mittelwert und einen geringen Anteil extrem guter Leser (5%) auf. Darüber hinaus scheinen die im Durchschnitt wenig zufriedenstellenden Leseleistungen österreichischer Schüler bei PISA 2009 im Wesentlichen durch die niedrigen Leistungen der Schüler in der unteren Hälfte der Leistungsverteilung bedingt zu sein (Schabmann, Landerl, Bruneforth & Schmidt, 2012, S. 19). Beim Lesen handelt es sich aber glücklicherweise um einen stark förderungsabhängigen Leistungsbereich und so stellen sich vor allem die Fragen: WER soll WIE lange und auf WELCHE Weise unterrichtet oder gefördert werden? Im Optimalfall sollten also für alle Schüler kontinuierlich und objektiv Indikatoren der Lesekompetenz erhoben werden, um den Lernverlauf zu kontrollieren, individuelle Probleme frühzeitig zu identifizieren und gegebenenfalls maßgeschneidert zu intervenieren. Den hohen Standards des CBM (*Curriculum Based Measurement*; vgl. Deno, 2003) für derartige Fragestellungen

folgend, wurde für dieses Diplomprojekt ein neuer Lesetest nach der MAZE-Technik entwickelt, der die Lesefähigkeit in regelmäßigen Abständen präzise und objektiv erfassen soll. Anschließend an die nun folgenden theoretischen Grundlagen der Arbeit, werden die genauen Konstruktionsprinzipien des Tests erklärt, die Erfüllung der Testgütekriterien anhand empirischer Daten überprüft und weitere auffällige Ergebnisse aus der Datenerhebung diskutiert.

2 Theoretische Grundlagen - Leseverständnis

Bei der Auseinandersetzung mit Literatur zum Thema Lesefertigkeiten, wie auch mit den Manualen gängiger Lesetests (z.B.: ELFE 1-6 (Lenhard & Schneider, 2006), LEVE (Proyer, Wagner-Menghin & Grafinger, 2006), SLS 1-4 (Mayringer & Wimmer, 2003),...) ¹ offenbart sich eine gewisse Vielfalt an Begriffsdefinitionen des Wortes „Leseverständnis“. Dies beginnt bereits bei der uneinheitlichen Abgrenzung von anderen Begriffen wie Lesekompetenz oder Leseverstehen, und setzt sich bei der Definition differenzierbarer Teilfähigkeiten des Lesens fort.

2.1 Begriffsbestimmung

Um eine inhaltliche Zusammenfassung, sowie eine Abgrenzung der zentralen Begriffe in der Erforschung des Leseverständnisses bemühte sich Lenhard (2013). Er zeigt explizit die Unterschiede und Gemeinsamkeiten zwischen Leseverständnis, Lesekompetenz und Literalität auf.

Das *Leseverständnis* beschreibt er als die allgemeine Fähigkeit bzw. den Prozess, den Inhalt eines gelesenen Textes zu rekonstruieren, wobei mehrere Teilfähigkeiten auf unterschiedlichen Ebenen ablaufen (siehe Kapitel 2.2). Gleichzeitig betont Lenhard, dass es sich dabei nicht unbedingt um eine tatsächlich erbrachte Leistung handeln muss, sondern eher um das grundlegende Potential zu einer solchen Leistung. Um diese auch tatsächlich zu erbringen, muss zusätzlich die Bereitschaft zur aktiven Verarbeitung des Textes gegeben sein.

Der Begriff *Lesekompetenz* wird bereits etwas weiter gefasst, und beinhaltet neben den kognitiven Prozessen auch emotionale und motivationale Faktoren. Betrachtet man den Kompetenzbegriff aus der Perspektive der Expertiseforschung, so ergeben sich hohe Leistungen aus dem wechselseitigen Zusammenspiel von Fähigkeiten, Vorwissen und der Motivation. Im Gegensatz zum Leseverständnis wird hier also das Hauptaugenmerk auf tatsächlich erbrachte Leistungen gelegt, womit auch eine gewisse Aufgaben- und Kontextabhängigkeit einhergeht. In der Kulturwissenschaft werde der Begriff sogar noch weiter gefasst, indem der soziale Kontext mitintegriert wird. Lenhard zieht den Schluss, dass die Lesekompetenz nicht explizit genug

¹ „Ein Leseverständnistest für Erst- bis Sechstklässler“, „Leseverständnistest“, „Salzburger Lese-Screening“

definiert ist um einen Test zu erstellen, von dem eindeutig auf die zugrundeliegende Kompetenz geschlossen werden kann.

Literalität (abgeleitet vom angloamerikanischen „literacy“) bezieht auch den kulturellen Kontext einer Person mit ein. Im Fokus stehen jene schriftbezogenen Tätigkeiten, die es dem Individuum ermöglichen, sein Leben in einem gegebenen kulturellen Umfeld zu bewältigen. Darunter fällt auch die Fähigkeit einem Text den Sinn zu entnehmen.

Aufgrund der eindeutigen Ausrichtung dieser Diplomstudie auf den kognitiven (Fähigkeits-)Aspekt des Leseprozesses, im Gegensatz zu möglichen emotionalen, motivationalen oder kulturellen Faktoren, entspricht der Begriff „Leseverständnis“ am ehesten den hier angestrebten Forschungszielen.

2.2 Komponenten des Leseverständnisses

Zwar sind sich Forscher laut Klicpera und Gasteiger-Klicpera (1993) bislang noch nicht gänzlich über die Anzahl und Beschaffenheit der Komponenten des Leseverständnisses einig, jedoch lassen sich viele Prozesse der Wort-, Satz- oder Textebene zuordnen (Lenhard, 2013). Klicpera und Gasteiger-Klicpera (1995) liefern hierzu einen guten Überblick:

2.2.1 Wortverständnis

Im Zuge des Leselernprozesses werden Wörter und deren Bedeutung in einem mentalen Lexikon abgespeichert. Beim Erkennen eines bereits bekannten bzw. gespeicherten Wortes im Text werden dessen kritische (Bedeutungs-)Merkmale abgerufen, wodurch es von anderen Objekten oder Handlungen unterschieden werden kann. Das Wortwissen über sämtliche Wörter eines Texts allein reicht jedoch nicht aus, um dessen Inhalt richtig zu erfassen, schließlich kann die Bedeutung eines Wortes durch den Kontext im Satz oder Textabschnitt modifiziert werden, zum Beispiel durch Analogien oder Metaphern. Aufgrund der kontextgebundenen Bedeutungsvielfalt beschreiben Klicpera und Gasteiger-Klicpera (1995) das Wortwissen als „komplexes, nur teilweise geordnetes Wissen“ (S. 134).

2.2.2 Satzverständnis

Zunächst muss zwischen zwei Arten der Satzstruktur unterschieden werden. Zum einen der *Oberflächenstruktur*, also der gegebenen Reihenfolge der Wörter eines Satzes, und zum anderen der *syntaktischen- oder Tiefenstruktur*. Um letztere zu erkennen, muss ein Satz in seine elementaren Strukturelemente, sog. Phrasen, aufgegliedert werden. Dabei wird zwischen Nominalphrasen und Verbalphrasen unterschieden. Verbalphrasen können wiederum selbst in eine Nominalphrase und ein Verb aufgeteilt werden. In der folgenden Grafik wird beispielhaft ein Satz in seine einzelnen Phrasen zerlegt:

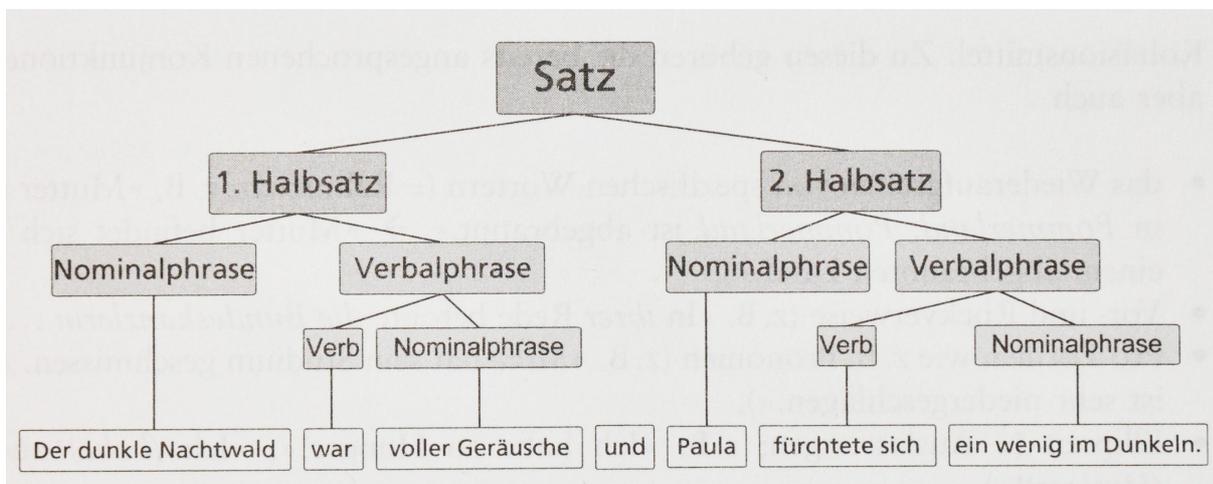


Abb. 1: Die Zergliederung eines Satzes in einzelne Phrasen (aus Lenhard, 2013, S. 19)

Die relevanten Bedeutungseinheiten sind folglich nicht einzelne Wörter, sondern Gruppen von Wörtern, die zueinander in Beziehung stehen, wodurch sich auch deren Sinn ändern kann. Die grammatikalische Analyse wird als „Parsing“ bezeichnet, läuft automatisch und bereits während des Lesens des aktuellen Satzes ab. Es wird jedem gelesenen Wort sofort eine bestimmte Position in der Struktur des Satzes zugewiesen. Selbst wenn ein Satz noch nicht zu Ende gelesen wurde, wird auf Grund der bisherigen Informationen die derzeit am wahrscheinlichsten zutreffende Satzstruktur angenommen und gegebenenfalls bei der Aufnahme weiterer Informationen revidiert.

Erst durch die Kombination von Wortbedeutungen und Tiefenstruktur ergibt sich dann für den Leser die „propositionale Struktur“ (Lenhard, 2013, S. 18), also die gedankliche Repräsentation von Bedeutungsinhalten eines Satzes.

Ob die syntaktische Analyse nun eher auf strukturellen Informationen oder semantischem Gehalt basiert, ist noch nicht gänzlich geklärt. Klicpera und Gasteiger-Klicpera (1995) gehen davon aus, dass bei Erwachsenen beide Faktoren miteinander interagieren, jedoch eine schnelle syntaktische Analyse der Verarbeitung semantischer Informationen voraus geht. Bei Kindern, die sich noch in der Aneignungsphase der Sprache befinden, laufe die syntaktische Analyse hingegen weitgehend autonom ab und sei weiters ein maßgebender Faktor für das Leseverständnis.

2.2.3 Textverständnis

Nach Klicpera und Gasteiger-Klicpera (1995) beruht das Leseverständnis „auf zwei teilweise unabhängigen Prozessen“ (S. 136), nämlich erstens der schrittweisen Bearbeitung beim Lesen aufgenommener Information, was der Schaffung einer Textbasis dient. In diesem Prozess werden vorerst eher kleine Informationseinheiten integriert und dann schrittweise erweitert, sodass eine zunehmend komplexere Struktur entsteht, die eine interne Gliederung auf höherem Niveau notwendig macht. Dadurch entsteht eine Makrostruktur, die verschiedene Themenbereiche beinhalten kann. Der zweite wesentliche Prozess ist die Aktivierung von Vorwissen, welches im Bezug zum Inhalt des Textes steht und dessen Verständnis und Interpretation mitbeeinflusst.

Bei Kintsch (1998) sind in die Aufnahme von Informationen aus dem Text, sowie die Aktivierung von Vorerfahrungen und Vorwissen Teilaspekte der Konstruktionsphase im „Construction-Integration Model“. Die daraus entstehenden propositionalen Netzwerke können jedoch nicht vollständig für die Lesedauer eines längeren Textes im Arbeitsgedächtnis behalten werden. Daher durchlaufen sie in der anschließenden Integrationsphase eine Art Reinigungsprozess, bei dem unwichtige Details und logische Widersprüche eliminiert werden, sodass die letztendlich wichtig erscheinenden Informationen in kohärenter Form im Langzeitgedächtnis gespeichert werden können.

2.2.3.1 Kohäsionsmittel und Inferenzen

Eine wichtige Fähigkeit für den Aufbau kohärenter propositionaler Netzwerke ist das Verständnis für die Bedeutung von *Kohäsionsmitteln*. Ihr Sinn besteht darin, Sätze

bzw. die darin vorkommenden Referenten miteinander in Beziehung zu setzen. Dazu zählen unter anderem Possessivpronomina und rückbezügliche Fürwörter. Werden sie falsch verstanden, können Zusammenhänge zwischen den Sätzen verloren gehen.

Des Weiteren sollen *Inferenzen* (= Schlussfolgerungen) aus dem Text abgeleitet werden. Man unterscheidet zwischen *notwendigen* und *weiterführenden* Inferenzen. Notwendige Inferenzen werden bereits während des Lesens des Textes gebildet, da sie für dessen Verständnis notwendig sind. Klicpera und Gasteiger-Klicpera (1995) nennen als Beispiel die sog. Instantiation, bei der für eine bereits konkret genannte Person in einem späteren Satz ein allgemeiner Ausdruck wie „der Mann“ gewählt wird. Weiterführende Inferenzen gehen über den konkreten Textinhalt hinaus, ergänzen ihn, führen ihn weiter oder engen ihn ein. Umgangssprachlich wird hierfür oft vom „zwischen den Zeilen lesen“ gesprochen.

2.2.3.2 Textstrukturen und Diskursformen

Je nachdem mit welcher Textform wir uns auseinandersetzen, haben wir unterschiedliche Erwartungen über den Aufbau des Textes oder den Ablauf der Handlung. Diese ergeben sich aus unseren bisherigen Erfahrungen mit verschiedenen Textformen und helfen uns die Aufmerksamkeit gezielt auf bestimmte inhaltliche Aspekte zu lenken, die für eine spätere Wiedergabe des Inhalts wichtig sind. Dabei kann das Wissen um verschiedene Schemata die erfolgreiche Wiedergabe zusätzlich positiv beeinflussen. Bei Geschichten trägt beispielsweise das Wissen über Handlungsschemata dazu bei, die Handlung vom initialen Handlungsmotiv bis zum Versuch der Zielerreichung und dessen Ergebnis korrekt zu rekonstruieren.

Bei Sachtexten kommen hingegen Schemata zur Anwendung, die sich auf die Erklärung allgemeiner Prinzipien, Definitionen, das Verhältnis von Gegenständen zueinander oder den Aufbau von Objekten beziehen.

Dabei ist es bei jedweder Textform wichtig, das eigene Verständnis des Textes auch zu überwachen. Markman (1977, 1979, zitiert nach Klicpera & Gasteiger-Klicpera, 1995, S. 141) meint, dass gerade jüngere Kinder Widersprüche im Text wenig berühren und Informationslücken kaum bemerken.

2.3 Hierarchische Ordnung der Teilprozesse

Ausgehend von einem Text entsteht unter Berücksichtigung der genannten Teilprozesse und Komponenten des Leseverständnisses eine mentale Zusammenfassung des Inhalts im Gedächtnis des Lesers. Man spricht hierbei auch von einem sogenannten *Situationsmodell*. Nach Lenhard (2013) lassen sich die beteiligten Prozesse einer hohen und einer niederen Hierarchieebene zuordnen. Diese Differenzierung ist jedoch nicht wertend im Sinne der Wichtigkeit aufzufassen. Während die hierarchieniederen Prozesse eher allgemeine Aspekte, wie die Entschlüsselung von Wörtern und der Syntax umfassen, beziehen sich die hierarchiehohen Prozesse auf komplexeres Ausgangsmaterial, sind strategisch-zielorientiert und durch ihre leichtere bewusste Zugänglichkeit auch teilweise gezielt steuerbar. Er nimmt an, dass viele der Prozesse parallel ablaufen und während des Lesens keine strenge Abfolge besteht.

2.3.1 Hierarchieniedere Prozesse

Auf der niederen Hierarchieebene finden wir die *Re- und Dekodierprozesse* der geschriebenen Sprache. Dabei steht geübten Lesern entsprechend dem *Zwei-Wege Modell des Lesens* (Coltheart & Rastle, 1994) eine lexikale und nicht-lexikale Route zum korrekten (lauten) Lesen von Wörtern zur Verfügung. Das nicht-lexikale (Vor-) Lesen erfolgt über die in der Schule erlernte Zuordnung von Graphemen (Buchstaben) zu Phonemen (Laute) eines Wortes. Diese Strategie wird gewählt, wenn sich zu dem Wort noch kein Eintrag im orthographischen Lexikon des Individuums befindet oder dieser Eintrag nicht aktiviert wurde. Erkennt der Leser hingegen das zu lesende Wort, so erfolgt die Verarbeitung über die lexikale Route und es kann aus dem Gedächtnis (phonologisches Lexikon) heraus richtig ausgesprochen werden. Auf diesem Weg ist dem Leser zumeist auch der Sinn des Wortes schnell verfügbar.

Weiters zählt Lenhard (2013) die bereits dargestellte *Propositionsbildung* und das Erkennen von *Kohäsionsmitteln* zu den hierarchieniederen Prozessen.

2.3.2 Hierarchiehohe Prozesse

Hierzu zählen zum einen die bereits beschriebene *globale Kohärenzbildung* und das Ziehen von *Schlussfolgerungen* - vor allem jener, die über den eigentlichen

Textinhalt hinausführen. Zum anderen die Aktivierung von *Vorwissen* bzw. *Textformatwissen* und deren Einflüsse auf die Interpretation des Inhaltes, sowie die Fähigkeit zur *Selbstregulation* im Sinne der gezielten Anwendung von Lesestrategien und Kontrolle des eigenen Textverständnisses.

2.3.3 Allgemeine kognitive Voraussetzungen

Als weitere relevante Einflussfaktoren des Leseverständnisses werden Aspekte wie die Arbeitsgedächtniskapazität, bereichsspezifisches Vorwissen, Sprachverständnis, schlussfolgerndes Denken und weitere kognitive Fertigkeiten angeführt (siehe Lenhard, 2013, S. 15).

Die Einteilung der Komponenten des Leseverständnisses nach Wort-, Satz- und Textebene sowie auf Hierarchieebenen eines Situationsmodells schließen einander nicht aus. Vielmehr sollten diese beiden Betrachtungsweisen als einander ergänzend gesehen werden und auf die erstaunliche Komplexität sinnerfassenden Lesens aufmerksam machen. Diese Komplexität führt in der Praxis auch zur Frage, auf welche Teilkomponente(n) unterdurchschnittliche Leistungen im Leseverständnis zurückzuführen sind.

2.4 Problemlokationen des Leseverständnisses

Grundsätzlich können Schwierigkeiten in jeder Komponente des Leseverständnisses auftreten. In Anlehnung an Klicpera und Gasteiger-Klicpera (1995) werden exemplarisch einige Beispiele gestörter Teilprozesse besprochen.

2.4.1 Probleme auf der Wortebene

Da sich das laute Lesen von Wortlisten als verlässlicher Indikator für verschiedene Aspekte der Lesefähigkeit etabliert hat, nehmen Probleme beim Wortverständnis eine wichtige Rolle ein. Beim Vorlesen unbekannter Wörter können Kinder zum Teil noch durch die Verarbeitung auf der nicht-lexikalen Route zu einer richtigen Lösung gelangen. Wenn das Wort jedoch Teil eines Satzes ist, dessen Inhalt verstanden und wiedergegeben werden soll, dann entgeht dem Kind mit dieser Strategie womöglich ein wesentlicher Bedeutungsaspekt. Es zeigte sich beispielsweise, dass leseschwache Kinder dazu neigen, bisher unbekannte Wörter im Satz zu ignorieren, und ihre Interpretation allein auf den Kontext stützen (McKeown, 1985, zitiert nach

Klicpera & Gasteiger-Klicpera, 1995). Unauffällige Leistungen bei Lesetests zum „lauten Lesen“ sind folglich kein Garant für ein gutes Leseverständnis.

2.4.2 Probleme auf der Satzebene

Ein weiteres Bindeglied zwischen dem lauten Lesen und dem Leseverständnis stellt die Fähigkeit zur korrekten Intonation eines Satzes dar. Leseschwache Kinder tendieren dazu, sämtliche Wörter eines Satzes gleich stark zu betonen, wodurch der Satzzusammenhang zerreißt. Die Satzanalyse wird somit erschwert, was jedoch mit zunehmender Geläufigkeit im mündlichen Vorlesen kompensiert werden kann. (Hinchley & Levy, 1988, zitiert nach Klicpera & Gasteiger-Klicpera, 1995).

2.4.3 Probleme auf der Textebene

Auf der satzübergreifenden Ebene können Schwierigkeiten im Verständnis für rückbezügliche Ausdrücke zum Tragen kommen. Das äußert sich beispielsweise in Problemen bei der korrekten Zuordnung oder Auswahl eines rückbezüglichen Fürworts zur entsprechenden Person. Daraus ergeben sich nicht nur inhaltliche Missverständnisse sondern auch ein größerer Zeitaufwand beim Lesen. Wie Murray und Kennedy (1988, zitiert nach Klicpera & Gasteiger-Klicpera, 1995) mittels Registrierung von Augenbewegungen herausfanden, suchen schwache Leser einen größeren Textabschnitt nach dem passenden Referenten des Fürworts ab, anstatt den Blick direkt auf die entsprechende Satzstelle zu richten, was den zeitlichen Mehraufwand erklären könnte.

Schwache Leser haben zumeist auch Probleme beim Beantworten von textbezogenen Fragen, die ein gewisses Maß an Informationsintegration und schlussfolgerndem Denken erfordern. Darüber hinaus tendieren sie dazu, bereits beim Lesen inhaltliche Inkonsistenzen des Textes seltener wahrzunehmen oder diese nur unzureichend erklären zu können.

2.5 Verlauf von Leseverständnisschwierigkeiten

Klicpera und Gasteiger-Klicpera (1993) betonen, dass Kinder mit Leseschwierigkeiten ihren Leistungsrückstand gegenüber durchschnittlichen sowie sehr guten Lesern in der Pflichtschulzeit im Regelfall nicht aufholen. Dies gilt sowohl für Rückstände in der Lesesicherheit und Lesegeschwindigkeit beim lauten Lesen, wie auch für verschiedene Teilaspekte des Leseverständnisses.

3 Kriterien für den neuen Lesetest

Grundsätzlich stehen im deutschen Sprachraum einige sehr verlässliche Tests zur Erfassung verschiedenster Teilaspekte der Lesefähigkeit zur Verfügung. Das gilt sowohl für die Lesegeschwindigkeit (z.B. mit dem SLS 1-4 von Mayringer und Wimmer (2003)) wie auch für das Leseverständnis (LEVE (Proyer, Wagner-Menghin & Grafinger, 2006), ELFE 1-6 (Lenhard & Schneider, 2006),...). Die etablierten Tests zeichnen sich zumeist durch eine hohe Reliabilität und Validität für schulisch relevante Fähigkeiten aus und eignen sich somit für klinisch psychologische Einzeldiagnostik (z.B. SLRT-II (Moll & Landerl, 2010)), groß angelegte Lernstandserhebungen oder zum Teil auch als Screeningverfahren zur Identifizierung von förderungsbedürftigen Schülern. Auch der subjektive Eindruck der Lehrperson ist ein wichtiger Indikator für die Lesefähigkeiten eines Kindes, jedoch fehlt es hierbei an Objektivität, da jeder Lehrer ein anderes Maß zur Leistungsbeurteilung heranziehen kann.

Im Kapitel 2.4 wurde deutlich gemacht, dass Leseschwierigkeiten sehr vielfältig sein können und der dadurch verursachte Leistungsrückstand ohne spezifische Intervention im Regelfall nicht aufgeholt wird. Mit den etablierten Tests lassen sich zwar durchaus leseschwache Schüler identifizieren, allerdings erweist sich die Evaluierung etwaiger Interventionen oftmals als schwierig, da (wenn überhaupt) zumeist nur eine einzige inhaltlich unterschiedliche, aber gleichwertige Parallelversion zur Verfügung steht. Des Weiteren ist eine Frequenz von ein bis zwei Testterminen pro Schuljahr nicht engmaschig genug, um förderungsbedürftige Schüler zeitnah zu identifizieren. Außerdem beanspruchen einige Tests aufgrund ihres Formats oder der Vorgabemodalitäten sehr viel Zeit, was deren Akzeptanz beim Lehrpersonal negativ beeinflussen kann.

Um diesen Problemen gerecht zu werden, sollte ein neuer Test zur Erfassung des Leseverständnisses folgende Kriterien erfüllen:

- *Gruppenverfahren*
- *Kosten-, Nutzen- und Zeiteffizienz* bei Erstellung, Durchführung und Auswertung
- *Hochfrequente Messwiederholungen* → wichtig für die Lernverlaufsdagnostik
- *Änderungssensibilität*

- *Einfache Anwendung*

Kaum ein Leseverständnistest im deutschsprachigen Raum erfüllt sämtliche dieser Kriterien, und das obwohl diese Forderungen eigentlich nicht neu sind. Sie finden sich bereits beim sogenannten Curriculum-Based Measurement (CBM), das auf die 70er Jahre des 20. Jahrhunderts zurückgeht (siehe Deno & Mirkin, 1977, zitiert nach Deno, 2003). Die Kapitel 3.1 bis 3.3 befassen sich daher mit den Grundlagen des CBM und weiteren bedeutenden Richtlinien für die Erstellung eines neuen Leseverständnistests.

3.1 Curriculum-Based Measurement

Deno (2003) verweist auf die Ursprünge des CBM als Instrument zur kontinuierlichen Evaluation der Auswirkungen von Unterrichtsentscheidungen von Lehrern auf die Schulleistung von Kindern mit Lernstörungen oder sonderpädagogischem Förderbedarf. Das Ziel war damals wie heute die Feststellung des individuellen Leistungswachstums der Schüler in schulrelevanten Fertigkeiten über einen längeren Zeitraum. In der weiteren Entwicklung wurde CBM aber auch in regulären Schulklassen eingesetzt, unter anderem als Screeningverfahren zur Identifikation von Schülern, bei denen die Gefahr schulischen Misserfolgs besteht.

Die Bezeichnung „Curriculum-based“ stammt noch aus den Anfängen des CBM, da das zugrunde liegende Stimulusmaterial vieler solcher Testverfahren zum Teil tatsächlich aus den Unterrichtsmaterialien des jeweils gültigen Curriculums stammte. Schon bald wurde jedoch festgestellt, dass die Verwendung von standardisiertem Stimulusmaterial – welches also nicht aus einem Schulbuch o.ä. stammte – zu keinerlei Verlusten bei der Relevanz für unterrichtsbezogene Entscheidungen führte.

3.1.1 Charakteristiken des CBM

Um den Leistungszuwachs in einer schulrelevanten Fähigkeit feststellen zu können, müssen wiederholt Messungen mit einem bestimmten Aufgabentyp durchgeführt werden. Das setzt vor allem die Erstellung zahlreicher äquivalenter Parallelförmungen des jeweiligen Tests voraus. Sie müssen sich aber trotzdem inhaltlich soweit voneinander unterscheiden, dass keine Erinnerungseffekte die Testergebnisse mitbeeinflussen können. Leistungsänderungen sollen – abgesehen vom

testspezifischen Messfehler – ausschließlich von den Fähigkeiten der Probanden abhängen.

Der übliche Weg zur Erstellung und Evaluierung eines CBM Testverfahrens liegt zunächst in der Auswahl von Testaufgaben, die an relevanten Außenkriterien validiert werden können und sich für Wiederholungsmessungen eignen. Anschließend kann die Nützlichkeit des Testverfahrens anhand der Auswirkungen der daran orientierten Entscheidungen im Unterricht evaluiert werden.

Während CBM Testverfahren in der Praxis grundsätzlich den gleichen Gütekriterien genügen müssen, wie alle psychologischen Tests (vgl. dazu Kubinger, 2006; Bortz & Döring, 2009), gelten der Reliabilität und Ökonomie besonderes Augenmerk.

Die Äquivalenz der oft zahlreichen Parallelförmigkeiten gilt es anhand der Paralleltestreliabilität zu überprüfen. Werden hier keine zufriedenstellenden Werte erreicht, so sind individuelle Leistungsveränderungen nicht mehr eindeutig auf tatsächliche Änderungen in der zugrundeliegenden Fertigkeit rückführbar.

Den Aspekt der Ökonomie bzw. Effizienz gilt es sowohl bei der Testdurchführung, wie auch bereits bei dessen Erstellung, zu beachten. Die Parallelförmigkeiten sollen einfach zu generieren, und die dafür notwendigen Testmaterialien kostengünstig in der Beschaffung oder Herstellung sein. Da das primäre Anwendungsfeld der CBM Verfahren Schulklassen sind, müssen sie für Lehrer und Pädagogen leicht zu erlernen und durchführbar sein. Als letzter wichtiger Punkt gilt noch die Zeiteffizienz, schließlich soll das Testverfahren trotz hochfrequenter Messungen den Unterricht nicht unnötig lange unterbrechen.

3.1.2 Schwierigkeiten in der Implementierung von CBM

Laut Deno (2003) scheitert die Einführung von CBM an Schulen oft an den Bedenken des Lehrpersonals bzgl. eines vermeintlich hohen Zeitaufwands. Dabei ist jedoch anzumerken, dass die Verfahren oft mit Zeitbegrenzungen von nur wenigen Minuten vorgegeben werden.

Vor allem Verfahren zur Erfassung des Leseverständnisses stoßen laut Yell, Deno und Marston (1992, zitiert nach Deno, 2003) aufgrund ihrer zum Teil geringen Augenscheinvalidität oft auf Inakzeptanz. Die direkte Messung des Leseverständnisses durch Vorgabe eines Textes und anschließender inhaltlicher

Fragen hätte zwar eine hohe Augenscheinvalidität, entspricht laut Deno aufgrund des hohen Zeitaufwandes in der Durchführung, sowie erheblicher Entwicklungskosten der Parallelversionen und Schwierigkeiten in der Auswertung nicht den Voraussetzungen des CBM. Hingegen lässt sich beispielsweise das „Laute Lesen“ von kurzen Textabschnitten sehr einfach und - zumindest in der Einzeldiagnostik - schnell durchführen, und erwies sich bereits als guter Indikator für das Leseverstehen und andere Teilaspekte der Lesefähigkeit (vgl. Walter, 2008).

3.2 Die MAZE Technik

Das Vorlesen von Wortlisten oder Texten kann jedoch nur mit Einzelpersonen durchgeführt werden, was bei größeren Stichproben der Forderung nach möglichst hoher Zeiteffizienz zuwiderläuft. Im Gegensatz dazu können diverse „Papier und Bleistift“ - Tests auch zeitsparend in der Gruppe vorgegeben werden. Im englischen Sprachraum haben sich Verfahren, die mit der MAZE Technik erstellt wurden, mehrfach als zuverlässige Indikatoren verschiedener Teilaspekte der Lesefähigkeit, u.a. des Leseverständnisses (siehe beispielsweise Wiley & Deno, 2005), erwiesen. Dabei bekommen die Probanden einen altersgemäßen Text vorgelegt, der an bestimmten Stellen Textlücken mit einer Auswahl von drei Wörtern aufweist, aus denen die einzige inhaltlich und grammatikalisch richtige Alternative im Multiple-Choice Format zu wählen ist. Hier ein Beispielsatz aus dem Übungstext für die Probanden dieses Diplomprojekts:

Das Nilpferd ist eines der größten [am - zum - Kamm] Land lebenden Säugetiere.

Neben dem richtigen (Ziel-)Wort verfügt jede Klammer über zwei inkorrekte Wörter (Distraktoren), deren Zusammenstellung je nach Untersuchungsdesign leicht variiert. Die Distraktoren können beispielsweise eine phonemische Ähnlichkeit zum Zielwort aufweisen, oder zur gleichen Wortart gehören (Nomen, Verben, usw.), dürfen den Satz jedoch nicht sinnvoll und grammatikalisch korrekt vervollständigen, wie das Zielwort. Meist wird die Bearbeitungszeit für diese Art von Tests auf ein bis drei Minuten begrenzt und der Testwert ergibt sich aus der Anzahl richtig gewählter Wörter. Wiley und Deno (2005) konnten zeigen, dass die MAZE Testwerte amerikanischer Schüler der 3. und 5.Schulstufe mit $r = .73$ (S. 211) stark mit einem

standardisierten staatlichen Testverfahren korrelieren, das mehrere Facetten des Leseverständnisses erfasst. Es muss jedoch einschränkend vermerkt werden, dass dieser Zusammenhang für Kinder mit anderer Muttersprache niedriger ausfällt. Weitere positive Belege für die Eignung der MAZE-Prozedur als Indikator des Leseverständnisses fand Walter (2009). Für Grundschüler der Schulstufen 2 bis 4 ergaben sich Korrelationen zwischen $r = .59$ und $.71$ (S. 69) zwischen dem MAZE Test und dem Test ELFE, der das Wort-, Satz- und Textverständnis differenziert erfasst. In einer späteren Längsschnittstudie verlängerte Walter (2011) die Bearbeitungszeit, verwendete längere und zugleich sprachlich geglättete Texte in einer Papierversion und PC-Version, wobei gleich mehrere Messzeitpunkte mit den standardisierten Lesetests verglichen werden konnten. Dabei ergaben sich erneut bedeutsame Korrelationen vor allem mit dem Satzverständnis ($r = .65$ bis $.80$) und dem Textverständnis ($r = .54$ bis $.68$) des ELFE (S. 115). Auch zum Lesegeschwindigkeitstest SLS 1-4 ergaben sich hohe Korrelationen zwischen $r = .68$ und $.80$ für Grundschüler (S. 114). Die für diese Verfahrensart ebenso bedeutsame Paralleltestreliabilität (der Papierversion) lag bei dieser Längsschnittstudie im Durchschnitt bei $r = .79$ (S. 113) für Messzeitabstände von ein bis drei Wochen, für die 4.Schulstufe. In einer Studie von Shin, Deno und Espin (2000) ergab sich unter Verwendung einer PC-Version des MAZE Verfahrens eine Paralleltestreliabilität von durchschnittlich $r = .81$ bei monatlicher Testung, über einen Zeitraum von neun Monaten. Bradley, Ackerson und Ames (1978) stellten fest, dass die Reliabilitätswerte auch dann noch zufriedenstellend ausfallen, wenn die MAZE Texte nicht von fachkundigen Forschern, sondern von den Lehrern selbst erstellt wurden, vorausgesetzt die Richtlinien zur Auswahl geeigneter Distraktoren wurden genau eingehalten.

Die aufgeführten Ergebnisse untermauern exemplarisch den potentiellen Nutzen des MAZE Verfahrens in der Leseforschung, wie auch in der schulischen Praxis. Um eine hohe Paralleltestreliabilität zu gewährleisten, müssen die Texte trotz unterschiedlichen Inhalts in bestimmten Merkmalen möglichst äquivalent zueinander sein. Eine Möglichkeit um diese Forderung in Bezug auf die Lesbarkeit eines Textes zu erfüllen, bieten sogenannte „Lesbarkeitsindizes“.

3.3 Messung der Lesbarkeit

Gerade bei Schulbüchern oder allgemein Büchern für jüngere Leser, sollten die Texte dem Alter und den Fähigkeiten der Leserschaft angepasst sein. Um sicherzustellen, dass die Textbasis des neuen MAZE Tests ebenfalls dieses formale Kriterium erfüllt, sollen die gängigen Methoden zur Lesbarkeitsbestimmung im Folgenden diskutiert werden.

3.3.1 Expertenrating

Eine geläufige Methode zur Feststellung der Adäquatheit eines Textes für bestimmte Altersgruppen ist das Expertenrating. Personen, die aufgrund ihres Berufes (z.B. Lehrer, Schriftsteller, Bibliothekare) über die Ansprüche und Fähigkeiten verschiedener Lesergruppen Bescheid wissen, sollen die vorgelegten Texte nach persönlicher Einschätzung Altersgruppen oder Schulstufen zuordnen. Bamberger und Vanecek (1984) kritisieren an dieser Methode, dass die subjektiven Einschätzungen äußerst stark variieren können, mitunter ergeben sich bei der Zuordnung zu Altersgruppen selbst im Schulbereich Unterschiede von bis zu 5 Jahren (S. 65). Eine ausreichende Verlässlichkeit sei erst durch die mathematische Mittelung der Einschätzungen mehrerer Experten möglich.

3.3.2 Lesbarkeitsindex

Die ersten Versuche, die Lesbarkeit oder den Schwierigkeitsgrad eines Textes zu quantifizieren, gehen bereits auf die 1920er Jahre in den USA zurück. In den Anfängen der Lesbarkeitsmessung versuchte man mit Hilfe von Häufigkeitslisten für Wörter die Vokabellast im Text festzustellen (Bamberger & Vanecek, 1984). Je häufiger die verwendeten Wörter im allgemeinen Sprachgebrauch vorkamen, desto leichter sei der daraus bestehende Text verständlich.

In der nächsten Entwicklungsphase versuchte man die Lesbarkeit mit Formeln zu errechnen, die eine große Zahl sprachlicher Faktoren beinhalteten, welche sich aus Regressionsgleichungen ergeben hatten. Man verwendete weiterhin Wortlisten, diese dienten aber eher der Feststellung, ob bestimmte Wörter zur Kategorie „seltene Wörter“ gezählt werden müssen. Weiters wurden beispielsweise die durchschnittliche Satzlänge, Anzahl oder Prozentsatz von Präpositionen und persönlicher Fürwörter und ähnliches verwendet. Der hohe Zeitaufwand für die

Erhebung der zahlreichen Variablen führte aber bald dazu, dass man die Anzahl sprachlicher Faktoren in den Formeln deutlich reduzierte. Aufgrund ihrer zeitökonomischen Anwendbarkeit und verlässlichen Ergebnisse hat sich im englischen Sprachraum vor allem die Formel von Flesch unter der Bezeichnung „Reading Ease Index“ (Flesch, 1948) durchgesetzt. Sie enthält nur noch zwei Variablen, nämlich die Anzahl der Silben pro 100 Wörter (word length, *wl*), sowie die durchschnittliche Satzlänge in Wörtern (sentence length, *sl*):

$$x = 206.835 - 0.846 * wl - 1.015 * sl$$

Nach Einsetzen der relevanten Größen resultiert eine Zahl zwischen 0 und 100, wobei hohe Werte für einen einfach zu lesenden Text und einen geringen Schwierigkeitsgrad stehen. Der augenscheinlich einleuchtenden Kritik, dass die Lesbarkeit eines Textes mit derart wenigen Variablen nicht verlässlich erfassbar sei, kann entgegengehalten werden, dass zum einen die Wortlänge in starkem Zusammenhang mit anderen Faktoren, wie der Bedeutungsvielfalt und der Häufigkeit im Sprachgebrauch steht (Best, 2006). Zum anderen gilt die Satzlänge indirekt als Maß für Satzkomplexität (Flesch, 1948). Die Formel erfasst somit auf indirektem Weg mehr, als die Variablenbezeichnungen vermuten lassen.

Die Formel von Flesch kann grundsätzlich auch auf deutschsprachige Texte angewendet werden. Die Ergebnisse wären jedoch systematisch verzerrt, da Wort- und Satzlänge in den Sprachen Deutsch und Englisch unterschiedlich gewichtet werden müssen (vgl. Amstad, 1978; Bamberger & Vanecek, 1984). Im Englischen erfolgt die Silbentrennung nach der Aussprache, wodurch sehr viele Wörter des täglichen Gebrauchs als einsilbig (z.B. give, take, house,...) gelten. Im Deutschen hingegen sind zahlreiche Wörter der einfachen Sprache bereits mehrsilbig (z.B. Fenster, gehen, Kindergarten), aufgrund ihrer Häufigkeit sind jedoch kaum negative Auswirkungen auf die Textschwierigkeit zu befürchten. Amstad (1978) bemühte sich daher um eine Neugewichtung der Faktoren in der Flesch-Formel für die Deutsche Sprache, welcher von Bamberger und Vanecek (1984) aber trotzdem eine reduzierte Zuverlässigkeit in der Lesbarkeitsbestimmung deutschsprachiger Literatur nachgesagt wird. In weiterer Folge entwickelten sie mehrere Lesbarkeitsformeln für

bestimmte Literaturformen. Erstens die „neuen Wiener Literaturformeln“ für literarische Jugendbücher, und zweitens die „neuen Wiener Sachtextformeln“ für Sachbücher. In ihrer dritten und zugleich ökonomischsten Form müssen für die Wiener Sachtextformel nur mehr die durchschnittliche Satzlänge und die Anzahl von „Mehrsilbern“ (Wörter mit drei oder mehr Silben) im Text bestimmt werden. Bei den Analysen von Sachtexten und daraus resultierenden Regressionsgleichungen stellte sich heraus, dass mit den Mehrsilbern die gleichfalls bedeutsamen „seltenen Wörter“ zum Großteil bereits mitefassen werden, was sie zu einem sehr verlässlichen Indikator der Textschwierigkeit macht. Als Indikator der Satzkomplexität dient in der Formel wiederum die Satzlänge.

Als grundsätzliche Kritik an den Lesbarkeitsformeln wird die Vernachlässigung von Aspekten wie Inhalt bzw. der Leserbezug zum Thema, Aufbau und Druckbild gesehen. Die Schwierigkeit des Inhalts hängt jedoch stark vom subjektiven Vorwissen ab und auch der Leserbezug kann oder sollte nicht mathematisch für eine Altersgruppe gemittelt werden. Für einen Lesetest empfehlen sich somit Inhalte, die kein allzu fachspezifisches Wissen vom Probanden erfordern. Das Druckbild (Schriftart, Schriftgröße) lässt sich für sämtliche Parallelversionen einfach konstant halten und auch der Textaufbau kann bei geeigneter Themenwahl nach einem einheitlichen Schema erfolgen.

4 Forschungsfragen und Hypothesen

Mit der vorliegenden Untersuchung soll ein weiterer Beitrag zur Überprüfung und Etablierung der MAZE Technik im deutschsprachigen Schulbereich geleistet werden. Durch die Anwendung mehrerer gleichwertiger MAZE Texte sollte es möglich sein, die Entwicklung des Leseverständnisses über einen Zeitraum von mehreren Monaten zu erfassen. Dabei soll unter anderem die Frage geklärt werden, ob die gewählten Kriterien (siehe Kap. 5.3) zur Erstellung der MAZE Texte zu einer ausreichend hohen Reliabilität für längsschnittliche Leistungsbeobachtungen führen. Darüber hinaus werden ausgewählte personenbezogene Merkmale, wie Geschlecht und Mehrsprachigkeit, auf deren Auswirkungen auf das Leseverständnis hin untersucht.

(1a) Eine der herausragendsten Eigenschaften von CBM Verfahren ist die Möglichkeit zur Darstellung kontinuierlich und objektiv erfasster Fähigkeitsmaße von Schülern. Daher ist es von großer Bedeutung, dass das hierfür eingesetzte Messinstrument selbst recht kleine Leistungsveränderungen, wie sie sich in Zeitintervallen von nur 4 Wochen ergeben, erfassen kann. Wurde die Zeitgrenze für die Bearbeitung passend gewählt, und sind die Items in angemessener Zahl und Schwierigkeit vorhanden, so sollten sich im Leistungsdiagramm aufgrund der kontinuierlichen Beschulung bestimmte Trends – in diesem Fall eine Leistungssteigerung – erkennen lassen.

(1b) Entsprechend der bisherigen Erkenntnisse zum Entwicklungsverlauf von schriftsprachlichen Fertigkeiten, sollten die Leistungsabstände zwischen Pbn am oberen und unteren Ende des Leistungsspektrums über den Untersuchungszeitraum hinweg tendenziell konstant bleiben oder sogar geringfügig zunehmen.

(2a) Der Einsatz des MAZE Verfahrens zur längsschnittlichen Leistungsmessung stellt hohe Ansprüche an dessen Reliabilität. Laut Bortz und Döring (2006) sollte ein Test, „der nicht nur zu explorativen Zwecken verwendet wird“ (S. 199), eine Reliabilität von mindestens 0,80 aufweisen. Das für diese Arbeit gewählte Versuchsdesign erlaubt sowohl die Erfassung der Retestreliabilität (Stabilität) wie auch der Paralleltestreliabilität (Äquivalenz) der einzelnen MAZE Texte zueinander. Bei Walter (2011) zeigten sich für die 4.Schulstufe bei zeitlich benachbarten Messungen (eine Woche Abstand) Reliabilitätswerte zwischen 0,79 und 0,84. Mit zunehmendem zeitlichen Abstand der Messungen sanken die Werte jedoch auf bis zu 0,70 (3 Wochen Abstand). Die bei der Textkonstruktion realisierte Einhaltung enger Grenzen eines Lesbarkeitsindex für die deutsche Sprache, sowie die Wahl einer Zeitgrenze, die selbst bei deutlichen Leistungsveränderungen zwischen den Messzeitpunkten das Auftreten von Ceiling-Effekten (vgl. Bortz & Döring, 2006) unwahrscheinlich macht, sollten trotz der verhältnismäßig großen Messintervalle zu akzeptablen Reliabilitäten führen. Zumindest für zeitlich benachbarte Messungen werden Werte von 0,80 und höher erwartet.

(2b) Die Retestreliabilitäten der einzelnen MAZE Texte sollen Aufschluss über das Ausmaß von Fehlereinflüssen auf die Messwerte geben. Die Darbietungsdauer von nur 2 ½ Minuten und das Zeitintervall von ca. 3-4 Wochen zwischen erster und zweiter Darbietung des jeweiligen Textes sollten Erinnerungseffekte gering halten, gänzlich ausschließen kann man sie jedoch nicht. Auch hier sollte für sämtliche MAZE Texte ein Reliabilitätswert von 0,80 nicht unterschritten werden.

(3) Bei Walter (2009) konnten bei hochfrequenter Vorgabe von MAZE Texten über den Zeitraum von einem Schuljahr signifikante Veränderungen in den Lesefertigkeiten bei Kindern der 2. bis 4.Schulstufe festgestellt werden. Da Walter (2009) eine über die Zeit hinweg kontinuierliche Zunahme der durchschnittlichen Testleistung proklamiert, sollten sich auch bei dieser Untersuchung, trotz der auf ein Semester reduzierten Versuchslaufzeit, signifikante Effekte der fortwährenden Beschulung feststellen lassen.

H0: Über die gesamte Studiendauer kommt es im MAZE Test zu keinen signifikanten Leistungsveränderungen.

H1: Zwischen mindestens zwei Testterminen können signifikante Leistungsveränderungen im MAZE Test festgestellt werden.

(4a) Beim SLS 1-4 (Mayringer & Wimmer, 2003) lässt sich anhand der Normtabellen erkennen, dass für Buben und Mädchen derselben Schulstufe unterschiedliche Werte in den Lesefähigkeiten zu erwarten sind. Da geschlechtsspezifische Unterschiede auch beim MAZE Test nicht ausgeschlossen werden können, sollen diese in der Auswertung überprüft werden:

H0: Das Geschlecht der Pbn hat keine signifikanten Auswirkungen auf die Testleistung.

H1: Die durchschnittlichen Rohscores von Buben und Mädchen unterscheiden sich signifikant voneinander.

(4b) In den österreichischen Medien wird regelmäßig über die schulischen Leistungen und Bildungsaussichten von Kindern debattiert, die nicht mit Deutsch als einziger Muttersprache aufwachsen. Dabei wird Mehrsprachigkeit eher als

Risikofaktor, denn als Ressource in der Bildungskarriere gesehen. In Hinblick auf diese anhaltenden Diskussionen, sollen in dieser Arbeit die Testleistungen zwischen exklusiv deutschsprachig (bis auf weiteres als „einsprachig“ bezeichnet) aufwachsenden Kindern und Kindern, die zumindest eine weitere Sprache im familiären Umfeld sprechen, verglichen werden.

H0: Die Mehrsprachigkeit hat keine signifikanten Auswirkungen auf die Testleistungen.

H1: Die durchschnittlichen Rohscores von Kindern mit exklusiv deutscher Muttersprache unterscheiden sich signifikant von jenen der mehrsprachig aufwachsenden Kinder.

(4c) Es soll auch untersucht werden, ob sich die Mehrsprachigkeit, wenn überhaupt, auf die Leseleistung beider Geschlechter gleichermaßen auswirkt oder ob es hierbei Unterschiede gibt.

H0: Es bestehen keine signifikanten Wechselwirkungen zwischen Geschlecht x Mehrsprachigkeit in Bezug auf die Testleistungen.

H1: Es bestehen signifikante Wechselwirkungen zwischen Geschlecht x Mehrsprachigkeit in Bezug auf die Testleistungen.

(5) MAZE Verfahren haben sich bereits mehrfach als verlässliche Indikatoren verschiedener Teilfertigkeiten des Lesens, und dabei im speziellen des Leseverständnisses erwiesen. In dieser Studie wird die Validität des MAZE Tests anhand einer Gegenüberstellung mit dem Lesetest SLS 1-4 (Mayringer und Wimmer, 2003) untersucht. Beim SLS 1-4 handelt es sich um ein schriftliches Verfahren zur Messung der Lesegeschwindigkeit als Teil der basalen Lesefertigkeiten. Da dabei verschieden lange Sätze auf deren inhaltliche Richtigkeit bewertet werden müssen, kann davon ausgegangen werden, dass das Leseverständnis ebenfalls indirekt miterfasst wird. Daher werden hohe Korrelationen von $r = .60$ und darüber, zwischen MAZE Test und SLS 1-4 erwartet. Weiters sollen die Leistungsunterschiede, die sich vom Anfang bis zum Ende des Untersuchungszeitraumes in beiden Tests ergeben haben miteinander verglichen werden.

5 Empirischer Teil

5.1 Auswahl der Stichprobe

Mit Genehmigung des Niederösterreichischen Landesschulrats, wurden insgesamt vier öffentliche Volksschulen aus den Bezirken Wiener Neustadt Stadt und Wiener Neustadt Land für dieses Projekt gewonnen. Die Teilnahme erfolgte auf freiwilliger Basis der Direktorinnen und Klassenlehrerinnen. Pro Schule nahmen je zwei bis drei vierte Klassen teil. Für die Eltern wurde ein Informationsschreiben verfasst und um die Erlaubnis zur Teilnahme ihres Kindes an dieser Studie gebeten. Darin wurde den Pbn Anonymität zugesichert und klargestellt, dass die Ergebnisse keine Auswirkungen auf die Schulnoten haben werden. Es standen letztendlich 149 Schüler aus 9 Klassen von 4 verschiedenen Volksschulen zur Verfügung.

5.2 Stichprobenbeschreibung

Insgesamt nahmen N = 149 Kinder der vierten Schulstufe an dieser Untersuchung teil. Davon waren 73 männlich (49%) und 76 weiblich (51%).

Weiters wurden die Kinder danach unterteilt, ob sie ausschließlich mit Deutsch als Muttersprache aufwachsen (einsprachig), oder noch weitere Sprachen im Alltag verwenden (mehrsprachig). Wenn ein Kind auf dem Datenblatt angab, selbst eine weitere Sprache außer Deutsch zu sprechen und/oder mit zumindest einem Elternteil in einer anderen Sprache zu kommunizieren, so wurde es für die weitere Auswertung als „mehrsprachig“ gekennzeichnet. Die Kinder wurden angewiesen, jene Fremdsprachen die sie erst in der Schule erlernten, nicht auf dem Datenblatt zu notieren.

Entsprechend dieser Einteilung wachsen 54 Kinder (36,2%) mehrsprachig auf, während 95 Kinder (63,8%) ausschließlich Deutsch als Alltagssprache angaben. Hier nun eine Auflistung der Sprachen, inklusive des Prozentanteils jener Pbn an der Gesamtstichprobe, die diese beherrschen:

Albanisch (5,4%), Amahrish (0,7%), Arabisch (2,0%), Bangla (0,7%), Bengalisch (1,3%), Bosnisch (2,0%), Französisch (0,7%), Georgisch (0,7%), Kroatisch (0,7%), Pakistanisch (0,7%), Rumänisch (4,0%), Serbisch (0,7 %), Spanisch (0,7%), Türkisch (11,4%), Tschechisch (1,3%), Ungarisch (2,7%).

5.3 Testmaterial

Das Testverfahren besteht aus vier MAZE Texten, welche den Probanden (Pbn) ein Mal monatlich mit begrenzter Bearbeitungszeit vorgegeben wurden. Um mögliche intraindividuelle Veränderungen in den Lesefähigkeiten innerhalb des Untersuchungszeitraums verlässlich zu erfassen, mussten die Texte eine Reihe von Anforderungen erfüllen:

1. Konstantes „*Lesbarkeits-Niveau*“, abgestimmt auf die Schulstufe
2. *Textinhalte* stammen aus demselben Themengebiet und entsprechen dem Wissensstand der Pbn
3. Die *Textlücken* (=Items) lassen nur eine einzige inhaltlich korrekte Lösung zu

Übergeordnetes Ziel dieser drei Kriterien ist das Erreichen einer möglichst hohen Paralleltest-Reliabilität. Je größer sie ist, desto verlässlicher können mögliche Veränderungen in den Testwerten zwischen den Messzeitpunkten auf reale Veränderungen in der Lesefähigkeit zurückgeführt werden.

1.) Das Lesbarkeits-Niveau:

Der Erstellung der MAZE Texte ging eine Analyse altersspezifischer Schulliteratur voraus. Hierfür wurden jeweils fünf zufällig ausgewählte Texte aus Lesebüchern der dritten und vierten Schulstufe auf deren Lesbarkeit untersucht. Die Texte stammten aus dem „LASSO Lesebuch 3“ (Bartnitzky & Bunk, 2007a) und „LASSO Lesebuch 4“ (Bartnitzky & Bunk, 2007b). In Anlehnung an Walter (2009) sollte ursprünglich eine der Flesch-Formeln zur Berechnung der Lesbarkeit verwendet werden. Wie im Theorieteil dieser Arbeit bereits erklärt wurde, sind diese jedoch nicht ohne Weiteres für die Deutsche Sprache geeignet. Daher wurde die Lesbarkeit mit der von Amstad (1978) für das Deutsche angepassten Formel (=Verständlichkeitsindex), und der 3.nWS (3. neue Wiener Sachtextformel) von Bamberger und Vanecek (1984) berechnet. Entscheidungsgrundlage für die Festlegung des Lesbarkeits-Niveaus der MAZE Texte war jedoch einzig die 3.nWS. Der Verständlichkeitsindex von Amstad (1978) wurde hier lediglich zu Vergleichszwecken mit bisherigen Forschungsarbeiten mit der MAZE Technik herangezogen. Denn auch an der Amstad-Formel wird analog zu den Flesch-Formeln kritisiert, dass „die durchschnittliche Silbenzahl von Wörtern

in der einfachen Sprache nicht so ausschlaggebend“ (Bamberger & Vanecek, 1984, S. 56) sei wie im Englischen.

Die für die Formeln relevanten Statistiken: durchschnittliche Satzlänge in Wörtern (SL), durchschnittliche Wortlänge in Silben (WL), sowie die Anzahl von Wörtern mit drei oder mehr Silben (MS) wurden für jeden Text ausgezählt und berechnet. Eine tabellarische Übersicht zur Textanalyse findet sich im Anhang. Die Formeln in mathematischer Schreibweise:

Verständlichkeitsindex:

$$x = 180 - SL - 58,5 WL$$

3.nWS:

$$x = 0,2963 MS + 0,1905 SL - 1,1144$$

Es zeigte sich, dass die Lesbarkeitswerte der Texte im Buch der 3.Schulstufe sehr stark variierten. Aufgrund dieser Schwankungsbreite hätte auch die Wahl des Mittelwerts der Lesbarkeit für die zu erstellenden MAZE Texte die Gefahr mit sich gebracht, einen großen Teil der Schüler zu über- oder unterfordern. Daher wurde auf die Erstellung von MAZE Texten für die 3.Schulstufe zu Gunsten einer größeren Stichprobe von Viertklässlern verzichtet.

Mit einer Ausnahme, wiesen die Texte für die 4.Schulstufe nämlich deutlich homogenere Werte für die Lesbarkeit auf. Vier der fünf analysierten Texte erreichen Werte zwischen 5,74 und 6,07 (3.nWS). Um auch dem Vorhandensein leistungsschwächerer Schüler Rechnung zu tragen, wurden bei der Erstellung der MAZE Texte für die 4.Schulstufe etwas niedrigere Lesbarkeitswerte angestrebt. Die Werte der erstellten MAZE Texte finden sich ebenfalls in der oben genannten Tabelle (siehe Anhang). Sie bewegen sich zwischen 5,21 und 5,64.

2.) Der Textinhalt:

Wie bereits von Walter (2009, 2011) erprobt, handeln die Texte vom Aussehen und der Lebensweise von Tieren, die den Kindern aus dem Alltags- und Schulwissen bekannt sein sollten. Damit soll sichergestellt werden, dass das Testverfahren

weitgehend unabhängig vom interindividuellen Wissensstand der Pbn, zuverlässige Ergebnisse liefert. In der Überschrift eines jeden Textes findet sich die Bezeichnung des jeweils behandelten Tieres. Dadurch werde laut Walter (2009) diesbezügliches Vorwissen aktiviert. Folgende Tiere kamen in den vier Texten vor: der Hase, das Haushuhn, der Elefant und der Wolf. Als Informationsquelle für die Texte diente die deutschsprachige Version der online Enzyklopädie <http://de.wikipedia.org>². Jeder MAZE Text besteht aus 20 bis 23 Sätzen, bzw. 251 bis 253 Wörtern (siehe Anhang). Zusätzlich wurde ein weiterer, deutlich kürzerer MAZE Text über das Nilpferd erstellt (= Übungstext), den die Pbn in der Übungsphase bekamen, um das neue Testverfahren kennenzulernen.

3.) Die Textlücken:

Bei den vier erstellten Texten (bzw. dem Übungstext) wurde jedes siebente Wort entfernt, und an dessen Stelle eine eckige Klammer gesetzt. In jeder Klammer stehen 3 Wörter nebeneinander, nämlich das Zielwort und zwei Distraktoren. Dabei ähnelt immer ein Distraktor dem Zielwort phonologisch/phonemisch und der andere Distraktor ist von derselben Wortart wie das Zielwort. Letzteres konnte jedoch nur bei Verben, Nomen und Adjektiven problemlos umgesetzt werden. Bei bestimmten Wortarten war es stellenweise nicht möglich, einen Distraktor aus der derselben Wortart zu wählen, da dieser sonst ebenfalls grammatikalisch sowie inhaltlich den Satz sinnvoll ergänzt hätte, oder wie im Sonderfall der additiven Konjunktion „und“, zu der es im Deutschen keine direkte Alternative gibt. In solchen Fällen wurde ein Distraktor einer zumindest artverwandten Wortart gewählt. Der phonologisch/phonemisch ähnliche Distraktor wies naturgemäß zumeist die gleiche Anzahl von Silben auf, wie das Zielwort. Die Reihenfolge von Zielwort und Distraktoren innerhalb der Klammer wurde per Zufall variiert.

Nach Fertigstellung der MAZE Texte musste eine sinnvolle Grenze für die Bearbeitungszeit festgelegt werden. Im Idealfall sollten nur jene Pbn mit außergewöhnlich hohen Lesefähigkeiten zur Bearbeitung sämtlicher Items eines MAZE Texts innerhalb der Testzeit kommen (zumindest bei der erstmaligen

² Die web-links und Zugriffszeitpunkte zu den jeweiligen Subseiten finden sich im Literaturverzeichnis

Vorgabe). Dabei sollte die Zeit aber derart begrenzt werden, dass zu späteren Testzeitpunkten noch Leistungsverbesserungen für die Probanden messbar sind. Weiters sollte der Großteil der Pbn durch instruktionsgemäße Bearbeitung des Tests zumindest einen höheren Rohscore (=Anzahl Richtige) erreichen, als jene Pbn, die den Test nicht laut Instruktion, sondern mittels anderer Lösungsstrategien, wie z.B. „Raten“ bearbeiten.

Da ein Vortest an Personen der Zielpopulation aus organisatorischen Gründen nicht möglich war, wurden drei Studenten unterschiedlicher Studienrichtungen (Psychologie, Informatik, Rechtswissenschaften), mit Deutsch als einziger Muttersprache, für dieses Testformat mit dem Übungstext instruiert. Anschließend wurden ihnen die MAZE Texte vorgegeben und die Bearbeitungszeit mitgestoppt. Die Studenten benötigten für die einzelnen Texte jeweils zwischen 118 und 132 Sekunden und fanden bei sämtlichen Items auf Anhieb die richtige Lösung. Da davon auszugehen war, dass Schüler der 4.Schulstufe im Durchschnitt deutlich mehr Arbeitszeit für einen *kompletten* MAZE Text benötigen würden als Studenten, wurde die Bearbeitungszeit für den Test auf 150 Sekunden festgesetzt. Dies scheint auf den ersten Blick noch immer eine recht strenge Zeitgrenze zu sein, wie sich aber im Kapitel zur Auswertung zeigen wird, gab es dennoch ein paar wenige äußerst leistungsstarke Pbn, von denen einzelne sogar den maximalen Rohscore erreichen konnten.

5.4 Versuchsplan

Für dieses Diplomprojekt kamen folgende Materialien zum Einsatz (siehe Anhang):

- 1 Übungs- MAZE Text
- 4 MAZE Texte
- 1 Datenblatt
- „Salzburger Lese-Screening“ (SLS 1-4) von Mayringer und Wimmer (2003)

Damit sowohl die Paralleltestreliabilität, wie auch die Restestrelabilität erfasst werden können, musste jeder der vier MAZE Texte von den Pbn zwei mal bearbeitet werden. Die Wiederholungsmessungen zu den MAZE Texten fanden jeweils zum darauffolgenden Testtermin statt. Zwischen Test und Retest jedes Textes lagen

somit jeweils drei bis vier Wochen. Der Ablauf der Testvorgabe wird in der folgenden Tabelle veranschaulicht.

Testzeitpkt.	Kalenderwoche	Testvorgabe
T1	KW 9	MAZE-"Hase" und SLS 1-4
T2	KW 12	MAZE-"Wolf" und Messwiederholung MAZE-"Hase"
T3	KW 16	MAZE-"Haushuhn" und Messwiederholung MAZE-"Wolf"
T4	KW 20	MAZE-"Elefant" und Messwiederholung MAZE-"Haushuhn"
T5	KW 24	Messwiederholung MAZE-"Elefant" und Messwiederholung SLS 1-4

Tab. A: Untersuchungsplan

Das Projekt startete Ende Februar 2013, Kalenderwoche 9, und der letzte Testtermin fand in der Kalenderwoche 24 statt. Da es sich sowohl beim MAZE Test, wie auch beim SLS 1-4 um Testverfahren handelt, die in der Gruppe durchgeführt werden können, wurden die Pbn klassenweise eingeschult und getestet. Die Testungen fanden stets während der Unterrichtszeit in den ersten drei Schulstunden statt. Mit Ausnahme des ersten Zeitintervalls (T1 bis T2) lagen jeweils rund vier Kalenderwochen zwischen den Testterminen. Aus schulinternen Gründen konnten die Testungen nicht fortwährend an den gleichen Wochentagen durchgeführt werden, sodass das Zeitintervall nicht konstant 28 Tage betrug, sondern um bis zu vier Tage innerhalb der jeweiligen Woche variierte.

Zum ersten Termin wurde den Pbn mündlich und schriftlich anhand eines Beispiels die Erstellung ihres persönlichen 6-stelligen Probandencodes erklärt. Anschließend folgte eine standardisierte mündliche Instruktion (siehe Anhang) zum MAZE Verfahren anhand des Übungstextes, wobei die Pbn unter anderem über die Bearbeitungszeit und Korrekturmöglichkeiten informiert wurden. Nach der Vorgabe vom MAZE Text „Hase“ sollten die Kinder Angaben zu Alter, Geschlecht und Sprachkenntnissen machen. Zum Abschluss des ersten Termins wurde den Pbn der SLS 1-4 vorgegeben. Jene Pbn, die zum ersten Termin nicht anwesend waren, bekamen die Einschulung und das Datenblatt unmittelbar vor dem 2.Termin; der SLS 1-4 sowie der MAZE Text „Hase“ wurden jedoch nicht nachgeholt.

Der SLS 1-4 wurde eingesetzt, um mögliche Veränderungen in den Testleistungen beim MAZE Test – der vorwiegend das Leseverständnis messen soll –

Veränderungen in der Lesegeschwindigkeit gegenüberzustellen, und diente zudem als Validitätskriterium.

5.5 Datenerfassung

Für jeden Probanden wurden die Informationen aus dem Datenblatt sowie die Testwerte des SLS 1-4 und den MAZE Texten in eine Tabelle des Programms „PASW Statistics 18“ eingetragen. Für die SLS 1-4 Bögen wurde das Testergebnis, also die Anzahl richtig gelöster Items notiert. Bei den MAZE-Texten wurde zwecks Durchführung einer Item-Analyse für jedes einzelne Item eingetragen, ob dieses korrekt/falsch gelöst oder gar nicht bearbeitet wurde.

6 Auswertung

6.1 Ergebnisse

Vier Untersuchungsteilnehmer wurden entweder nach einem Sonderschullehrplan beurteilt oder aufgrund mangelnder Sprachkenntnis als „außerordentliche“ Schüler geführt und bleiben daher für die Hypothesenprüfung unberücksichtigt.

Bei fünf weiteren Teilnehmern wurde ein zunehmend auffälliges Testverhalten festgestellt: Zu den Testterminen t1 und t2 bearbeiteten sie bis zu ca. 60% der Items, hatten keine Auslassungen und keine bis sehr wenige Fehler. Bei den späteren Testterminen wurden hingegen überdurchschnittlich viele Items bearbeitet, bis hin zu einer vollständigen Bearbeitung; jedoch mit bis zu 50% falsch gelösten oder nicht bearbeiteten Textlücken. Da eine derart sprunghafte Veränderung in den Lesefähigkeiten äußerst unwahrscheinlich erscheint, legen diese Ergebnisse nahe, dass der Text nicht eingehend gelesen wurde, sondern andere Lösungsstrategien, wie z.B. „Raten“ zur Anwendung kamen. Die Datensätze der fünf betreffenden Schüler wurden somit ebenfalls in der Hypothesenprüfung nicht berücksichtigt.

Der für die Auswertung verwendete Datensatz enthält somit noch N = 140 Kinder, davon 35 % mehrsprachig und 48,6 % männlich.

Zur Ermittlung der Reliabilitäten, sowie für die Gruppenvergleiche, werden für die Probanden Rohscores für jeden MAZE Text berechnet. Diese ergeben sich jeweils aus der Anzahl korrekt gelöster Items. Für inkorrekt gelöste oder ausgelassene Items werden keine Punkte vergeben. Folglich bewegt sich der mögliche Rohscore für die

MAZE Texte *Hase*, *Haushuhn* und *Elefant* zwischen 0 und 36 Punkten. Der Text über den *Wolf* hat dagegen ein Item weniger und somit beträgt auch der maximale Rohscore nur 35 Punkte.

6.1.1 Leistungsverlauf

Bei Betrachtung der Tabelle 1 fällt auf, dass sich die durchschnittlichen Testleistungen der Schüler vom ersten zum zweiten Testzeitpunkt kaum verändern und anschließend deutlich zunehmen.

		Hase_T1_Ges	Wolf_T2_Ges	Huhn_T3_Ges	Elefant_T4_Ges
N	Gültig	105	105	105	105
	Fehlend	0	0	0	0
Mittelwert		15,76	15,71	16,90	19,10
Median		16,00	15,00	17,00	19,00
Standardabweichung		4,679	5,296	5,420	5,262

Tab.1: Mittelwerte der MAZE Texte von T1 bis T4

Auch in der graphischen Darstellung wird dieser Verlauf offensichtlich:

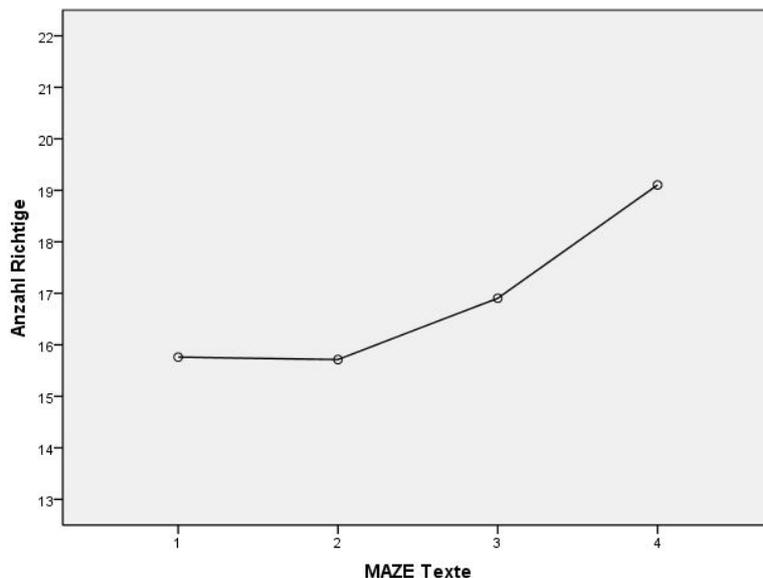


Abb. 2: Mittelwerte der MAZE Texte von T1 bis T4

Ob personenbezogene Faktoren, wie Geschlecht oder Mehrsprachigkeit diesen Leistungsverlauf mit beeinflussen, wird in Kap. 6.1.4 geklärt. Vorerst soll jedoch geklärt werden, ob unabhängig von diesen Faktoren, die Leistungen von Pbn am unteren und oberen Ende des Leistungsspektrums einen ähnlichen Verlauf nehmen.

Zunächst werden von den Ergebnissen im ersten MAZE Text die Werte für das 25ste und 75ste Perzentil ermittelt, um anschließend Gruppen für das oberste und das unterste Quartil bilden zu können. Bei der graphischen und tabellarischen Gegenüberstellung dieser Extremgruppen bleiben die mittleren 50 % der Pbn unberücksichtigt.

		Hase_T1_Ges	Wolf_T2_Ges	Huhn_T3_Ges	Elefant_T4_Ges
PR <= 25	Mittelwert	10,47	11,63	12,30	14,77
PR >= 75	Mittelwert	21,67	20,48	21,59	24,48

Tab. 2: Mittelwerte für das oberste und unterste Quartil

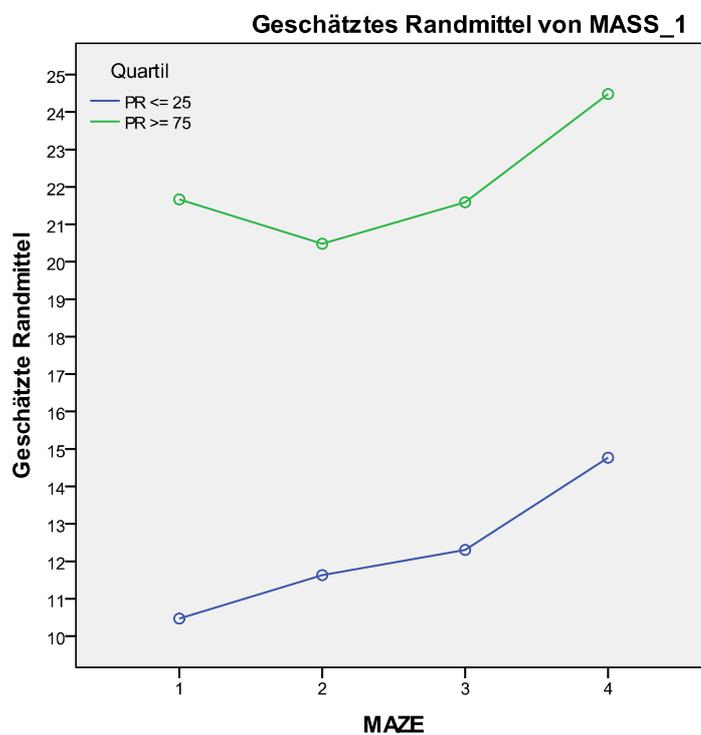


Abb. 3: Mittelwerte für das oberste und unterste Quartil

Sowohl für das oberste, wie auch für das unterste Quartil der Leistungsverteilung ergeben sich optisch und numerisch feststellbare Leistungsverbesserungen vom ersten zum vierten Testtermin. Der Leistungsunterschied zwischen den beiden Gruppen beträgt dabei über den Untersuchungszeitraum hinweg ca. 9 bis 11 Items. Um diese Leistungsverläufe genauer zu analysieren, wird eine Varianzanalyse mit Messwiederholung für die beiden angeführten Quartilgruppen durchgeführt. Zunächst wird daher die Normalverteilung der Daten für beide Gruppen pro MAZE Text mit dem Kolmogorov-Smirnov Test überprüft.

MAZE Text	Quartil	Kolmogorov-Smirnova		
		Statistik	df	Signifikanz
Hase_T1_Ges	1	,162	30	,043
	3	,199	27	,008
Wolf_T2_Ges	1	,111	30	,200
	3	,188	27	,015
Huhn_T3_Ges	1	,102	30	,200
	3	,175	27	,033
Elefant_T4_Ges	1	,093	30	,200
	3	,129	27	,200

Tab. 3: Test auf Normalverteilung für das unterste (1) und oberste (3) Quartil

In Tabelle 3 ist zu erkennen, dass in genau einem Fall das Signifikanzniveau von $\alpha = 0,01$ unterschritten wurde. Da in allen anderen Fällen die Annahme der Normalverteilung nicht verletzt wurde und die Varianzanalyse darüber hinaus als recht robust gegenüber Verletzungen der Normalverteilung gilt (vgl. Schmider, Ziegler, Danay, Beyer & Bühner, 2010), sollte die Varianzanalyse dennoch zu verwertbaren Ergebnissen führen.

Als weitere Voraussetzung für dieses Auswertungsverfahren, wurde die Sphärizität mit dem Mauchly-Test überprüft.

Mauchly-Test auf Sphärizität

Maß: MASS_1

Quartil	Innersubjekteffekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon		
						Greenhouse-Geisser	Huynh-Feldt	Untergrenze
1	MAZE_Text	,818	5,584	5	,349	,884	,981	,333
3	MAZE_Text	,898	2,668	5	,751	,932	1,000	,333

Tab. 4: Mauchly Test auf Sphärizität für das unterste (1) und oberste (3) Quartil

Die Ergebnisse von Mauchlys Test sind für beide Quartilgruppen mit $\chi^2(5) = 5.58$, $p > .05$ sowie $\chi^2(5) = 2.67$, $p > .05$ nicht signifikant, womit die Annahme der Sphärizität der Daten beibehalten werden kann.

Die Leistungsveränderungen über den gesamten Untersuchungszeitraum waren sowohl für das unterste Leistungsquartil mit $F(3, 87) = 15.60$, $p < .05$ wie auch für das oberste Leistungsquartil mit $F(3, 78) = 10.48$, $p < .05$ signifikant.

Erstellt man Kontraste für jeweils zeitlich benachbarte Testtermine, so stellt sich heraus, dass für beide Quartilgruppen jeweils nur zwischen dem dritten und vierten Testtermin signifikante Veränderungen der Testleistungen mit $F(1, 29) = 19.21$, $p < .05$ (unterstes Quartil) und $F(1, 27) = 17.10$, $p < .05$ (oberstes Quartil) auftreten.

Mit dem MAZE Test lassen sich also durchaus Leistungsveränderungen bei kurzen Messzeitintervallen feststellen und auch anschaulich grafisch darstellen (Hypothese 1a), jedoch erwiesen sich nicht alle optisch/numerisch auffälligen Veränderungen als signifikant. Ein durchwegs paralleler oder gar scherenförmiger Leistungsverlauf zwischen dem untersten und obersten Leistungsquartil konnte nicht festgestellt werden (Hypothese 1b), die signifikanten Veränderungen treten aber für beide Gruppen beim gleichen Zeitintervall auf.

6.1.2 Paralleltestreliabilität des MAZE Tests

Da dieser Test auf Leistungsmessungen bei Kindern im österreichischen - und somit deutschsprachigen - Schulsystem ausgelegt ist, werden bei der Bestimmung des Testgütekriteriums Reliabilität nur die Testdaten von Pbn mit ausschließlich deutscher Muttersprache herangezogen. Die Inklusion mehrsprachiger Pbn könnte die Kennwerte für die Testgenauigkeit in unbekannter Richtung verzerren.

Die Rohscores wurden für die Berechnung der Paralleltestreliabilitäten nur aus den ersten 35 Items des jeweiligen Textes (bei Erstvorgabe) errechnet. Auf das 36te und somit letzte Item von drei Texten, musste im Sinne eines homogenen Rohwertmaximums verzichtet werden. Der Informationsverlust ist dabei jedoch äußerst gering, da von den ausschließlich Deutsch sprechenden Kindern nur ein einziger Proband eines der drei Items bearbeitet hatte. Die anderen beiden Items wurden von keinem Pbn bearbeitet.

Die auf diese Weise ermittelten Rohscores wurden paarweise miteinander korreliert. In der folgenden Tabelle werden die Pearson-Korrelationen zwischen den MAZE Texten (inkl. Angabe der Kalenderwoche in der diese vorgegeben wurden) wiedergegeben.

Text		Hase	Wolf	Haushuhn	Elefant
	Zeitpunkt	KW 9	KW 12	KW 16	KW 20
Hase	KW 9	1,000	0,689	0,653	0,629
Wolf	KW 12	0,689	1,000	0,814	0,781
Haushuhn	KW 16	0,653	0,814	1,000	0,826
Elefant	KW 20	0,629	0,781	0,826	1,000

Tab. 5: Pearson-Korrelationen der MAZE Texte (bei Erstvorgabe)

Wie in der Tabelle zu sehen ist, wird nur in 2 von 6 Fällen die von Bortz und Döring (2006) geforderte untere Grenze der Reliabilität für Leistungstests von 0,80 erreicht, womit der MAZE Test hinter den Erwartungen (Hypothese 2a) zurückbleibt, der Durchschnitt liegt bei $r = 0,732$. Um dieses Ergebnis abzusichern wird die Äquivalenz der MAZE Texte aus den Wiederholungsmessungen überprüft. Aufgrund der einheitlich gewählten Zeitintervalle, sollten mögliche Erinnerungseffekte bei allen Texten gleich starke Auswirkungen auf den Testwert haben und somit vernachlässigbar sein.

Text		Hase	Wolf	Haushuhn	Elefant
	Zeitpunkt	KW 12	KW 16	KW 20	KW 24
Hase	KW 12	1,000	0,845	0,827	0,802
Wolf	KW 16	0,845	1,000	0,835	0,838
Haushuhn	KW 20	0,827	0,835	1,000	0,914
Elefant	KW 24	0,802	0,838	0,914	1,000

Tab. 6: Pearson-Korrelationen der MAZE Wiederholungsmessungen

Werden die Testwerte der Wiederholungsmessungen zur Berechnung herangezogen, so bleiben die Werte durchgängig über der Untergrenze von $r = 0,80$. Auch die in Tabelle 3 auffällig niedrigeren Korrelationen der anderen drei Texte mit dem Hasen-Text lassen sich in Tabelle 4 nicht mehr nachweisen.

Da der jeweils ersten Vorgabe der MAZE Texte das höhere Gewicht beizumessen ist, muss die Annahme einer zufriedenstellenden Äquivalenz für diese Untersuchung zwar verworfen werden, jedoch weisen die Ergebnisse in Tabelle 2 darauf hin, dass dies nicht zwangsläufig an der Qualität der Texte liegt, sondern womöglich an Schwächen im Untersuchungsdesign (siehe Kap. 7.2.1).

6.1.3 Retestrelabilität des MAZE Tests

Als nächstes wird die Retestrelabilität der MAZE-Texte berechnet. Dabei werden für jeden Text die Rohscores sämtlicher Items aus erster und zweiter Vorgabe miteinander korreliert. Die folgende Tabelle gibt wiederum die Pearson-Korrelationen, sowie die Anzahl berücksichtigter Pbn wieder:

	Hase	Wolf	Haushuhn	Elefant
Pearson	0,828	0,886	0,869	0,918
N =	69	79	84	81

Tab. 7: Retestrelabilität der einzelnen MAZE Texte

Für die Retestreliaibilität ergeben sich bei sämtlichen Texten durchwegs befriedigende Werte, deutlich über der Schwelle von 0,80. Die MAZE Texte erweisen sich somit im Einzelnen als ausreichend stabil für die Leistungsmessung (Hypothese 2b).

6.1.4 Leistungsveränderungen und Gruppenvergleiche

Als nächstes sollen die Leistungsverläufe und die Effekte voneinander unabhängiger Personenvariablen auf die Testleistung untersucht werden. Hierzu wird eine *mehrfaktorielle Varianzanalyse mit Messwiederholung* durchgeführt. Die vier MAZE Texte – bei ihrer jeweils erstmaligen Vorgabe – werden für diese Auswertungsmethode als zueinander äquivalent betrachtet. Analog zur Berechnung der Paralleltestreliaibilität, werden die Rohscores wiederum nur aus den ersten 35 Items pro Text gebildet.

Vor der eigentlichen Hypothesenprüfung muss jedoch festgestellt werden, ob für die zu vergleichenden Gruppen eine Normalverteilung (NV) in sämtlichen abhängigen Variablen vorliegt. Diese Überprüfung wurde mit dem *Kolmogorov-Smirnov* Test (K-S Test) mit Signifikanzkorrektur nach Lilliefors durchgeführt. Als Signifikanzniveau wurde $\alpha = 0,01$ festgelegt.

In der folgenden Tabelle werden die Signifikanzwerte für sämtliche Kombinationen von Gruppen und MAZE Texten wiedergegeben (siehe unten). Hier gilt zu beachten, dass nach dem listenweisen Fallausschluss nur mehr N=105 Personen für die Gruppenvergleiche sowie für die Überprüfung der NV und Sphärizität zur Verfügung stehen.

Geschlecht	Mehrsprachig		Kolmogorov-Smirnov		
			Statistik	df	Signifikanz
männlich	nein	Hase_T1_Ges	,098	36	,200
		Wolf_T2_Ges	,168	36	,012
		Huhn_T3_Ges	,149	36	,042
		Elefant_T4_Ges	,083	36	,200
	ja	Hase_T1_Ges	,165	19	,186
		Wolf_T2_Ges	,177	19	,119
		Huhn_T3_Ges	,143	19	,200
		Elefant_T4_Ges	,127	19	,200
weiblich	nein	Hase_T1_Ges	,158	29	,062

	Wolf_T2_Ges	,182	29	,015
	Huhn_T3_Ges	,166	29	,040
	Elefant_T4_Ges	,160	29	,056
ja	Hase_T1_Ges	,129	21	,200
	Wolf_T2_Ges	,183	21	,065
	Huhn_T3_Ges	,129	21	,200
	Elefant_T4_Ges	,147	21	,200

Tab. 8 Test auf NV mit den Faktoren Geschlecht und Mehrsprachigkeit

Wie in der markierten Spalte rechts außen zu sehen ist, wurde das Signifikanzniveau von $\alpha = 0,01$ in keinem einzigen Fall unterschritten. Die Annahme der Normalverteilung der Rohwerte kann daher für alle vier MAZE Texte beibehalten werden.

Anschließend wurde mit *Mauchly's Test* die Sphärizität, also die Homogenität der Varianzen der Messwertdifferenzen, überprüft.

Mauchly-Test auf Sphärizität

Maß:MASS_1

Innersubjekteffekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Untergrenze
MAZE	,886	12,030	5	,034	,921	,977	,333

Tab. 9: Mauchly-Test auf Sphärizität

Das signifikante Ergebnis bei Mauchly's Test $\chi^2(5) = 12.03$, $p < .05$ legt nahe, dass die Annahme der Sphärizität der Daten nicht aufrecht erhalten werden kann. Bei Field (2009) findet sich die Empfehlung (S. 461), bei einem Epsilon (ϵ) $> 0,75$ die Huynh-Feldt Korrektur der Freiheitsgrade anzuwenden. Signifikante Innersubjekteffekte werden daher mit dem nach Huynh-Feldt korrigierten Wert bei einem Signifikanzniveau von 0.05 berichtet.

Die durchschnittlichen Testleistungen der Pbn im MAZE Test veränderten sich innerhalb des Untersuchungszeitraumes signifikant, bei $F(2.93, 296.09) = 36.77$, $p < .05$. Daher kann die H_0 verworfen und die Alternativhypothese, dass sich die Testwerte insgesamt über die Zeit verändern, angenommen werden (Hypothese 3).

Um diesen Effekt genauer zu untersuchen, wurden Kontraste für jeweils zeitlich benachbarte Testtermine erstellt. Es stellte sich heraus, dass für die Gesamtstichprobe ($n = 105$) nur zwischen den Messzeitpaaren $T_2 - T_3$ bei $F(1, 101)$

= 13,95, $p < .05$ sowie $T3 - T4$ bei $F(1, 101) = 41.23$, $p < .05$ signifikante Veränderungen der Testleistungen auftraten, zwischen den ersten beiden Messzeitpunkten waren die Veränderungen nicht signifikant. In der Tabelle werden die Mittelwerte der Rohscores nach Testzeitpunkt bzw. MAZE Text aufgeschlüsselt. Im Durchschnitt scheinen sich die Werte vom ersten zum zweiten Testtermin nur sehr geringfügig, und noch dazu in negativer Richtung zu verändern. Im weiteren Verlauf erzielen die Pbn dann jedoch durchschnittlich zunehmend höhere Werte als beim ersten Testtermin.

	T1	T2	T3	T4
	Hase	Wolf	Haushuhn	Elefant
Mittelwert	15,76	15,71	16,90	19,10
Std-abw.	4,68	5,30	5,42	5,26

Tab. 10: Mittelwerte der Gesamtstichprobe (N=105)

Des weiteren zeigt sich ein signifikanter Wechselwirkungseffekt von MAZE x Geschlecht x Mehrsprachigkeit bei $F(2.93, 296.09) = 2.85$, $p < .05$. Bei Kontrasten für zeitlich benachbarte Testtermine konnte eine signifikante Wechselwirkung jedoch nur beim Messzeitpaar $T1 - T2$ bei $F(1, 101) = 4.87$, $p < .05$ festgestellt werden. Dies deutet darauf hin, dass es bei diesem Messzeitpaar Unterschiede in den Leistungsverläufen zwischen Buben und Mädchen in Abhängigkeit von der Sprachkenntnis gibt.

Als nächstes werden die Hypothesen 4a bis 4c überprüft. Levene's Test auf Homogenität der Varianzen ergibt für alle MAZE Texte nicht signifikante Werte bei $\alpha = 0.05$, wodurch die Annahme homogener Varianzen beibehalten werden kann. Weder für das Geschlecht $F(1, 101) = 0.85$, *ns* (nicht signifikant), noch für die Mehrsprachigkeit $F(1, 101) = 2.53$, *ns*, konnte ein signifikanter Haupteffekt festgestellt werden. Infolgedessen werden die Nullhypothesen für 4a und 4b beibehalten.

Es ergab sich jedoch ein signifikanter Wechselwirkungseffekt für *Geschlecht x Mehrsprachigkeit* bei $F(1, 101) = 4.57$, $p < .05$ (Alternativhypothese 4c)! Da eine solche Interaktion bereits in Kombination mit der Messwiederholung stattfindet, soll diese nun genauer untersucht und veranschaulicht werden.

Stellt man die mittleren Rohwerte für mehrsprachig und einsprachig aufwachsende Kinder tabellarisch dar, so zeigt sich für beide Gruppen ein zunehmend positiver

Leistungsverlauf (Tab. 11.1). Dabei scheinen Kinder mit ausschließlich deutscher Muttersprache zu jedem Testzeitpunkt im Durchschnitt mehr Items richtig zu lösen, als Kinder die in einem mehrsprachigen Haushalt aufwachsen. Wird nur nach dem Geschlecht verglichen (Tab. 11.2), so scheinen Mädchen den Buben zu sämtlichen Testzeitpunkten überlegen zu sein. Auffällig ist hier jedoch, dass die Leistung der Mädchen im Durchschnitt vom ersten zum zweiten Testtermin abnimmt, während sie im gleichen Zeitraum bei den Buben leicht zunimmt.

Sprache	T1	T2	T3	T4
deutsch	16,26	16,2	17,35	19,69
mehrspr.	14,95	14,93	16,18	18,15

Tab. 11.1: MW nach Sprachkenntnis

Geschlecht	T1	T2	T3	T4
Mädchen	16,28	16,12	17,66	19,92
Buben	15,29	15,35	16,22	18,36

Tab. 11.2: MW nach Geschlecht

Die Wechselwirkung von Geschlecht und Mehrsprachigkeit wird erst sichtbar, sobald man separate Diagramme für Buben und Mädchen, mit jeweils eigenen Linien für mehrsprachige und einsprachige Pbn erstellt (siehe unten).

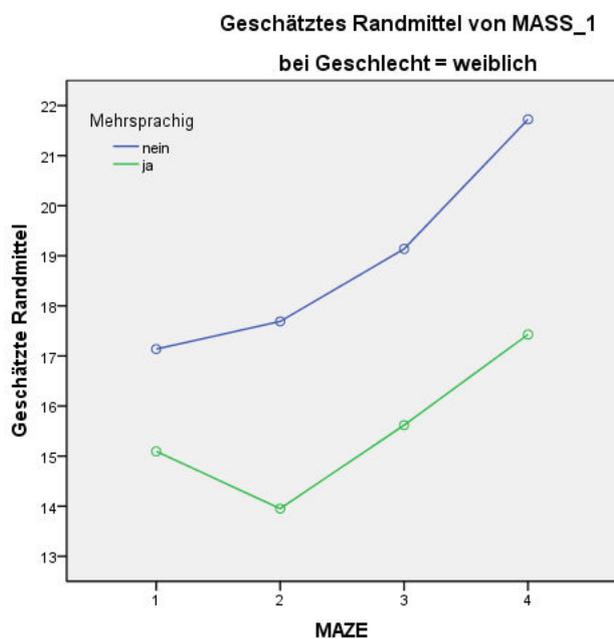


Abb. 4.1: MW für Mädchen, getrennt nach Sprachkenntnis

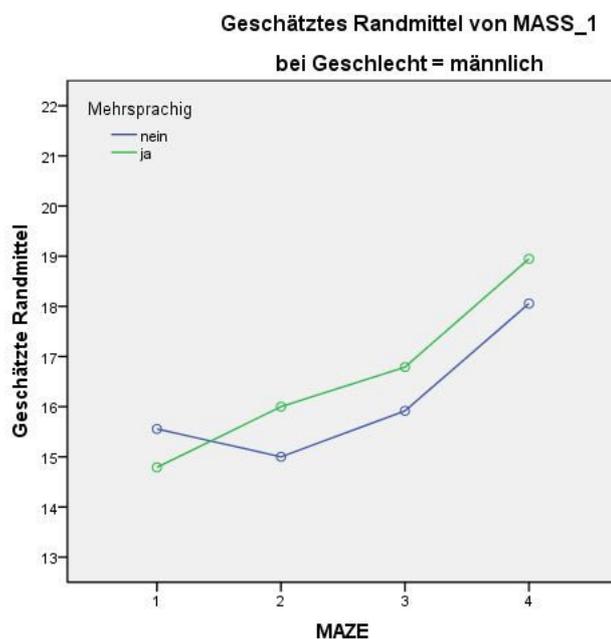


Abb. 4.2: MW für Buben getrennt nach Sprachkenntnis

Mädchen	T1	T2	T3	T4
deutsch	17,14	17,69	19,41	21,72
mehrspr.	15,10	13,95	15,62	17,43

Tab. 12.1: MW für Mädchen, getrennt nach Sprachkenntnis

Buben	T1	T2	T3	T4
deutsch	15,56	15,00	15,92	18,06
mehrspr.	14,79	16,00	16,79	18,95

Tab. 12.2: MW für Buben, getrennt nach Sprachkenntnis

Zum einen unterscheiden sich Mädchen in Abhängigkeit von Mehrsprachigkeit deutlicher in der Anzahl richtig gelöster Items, als das bei Buben der Fall ist. Zum anderen kehrt sich bei den Buben der Leistungsvorsprung exklusiv deutschsprachiger gegenüber mehrsprachig aufwachsender Pbn vom ersten zum zweiten Testtermin um und bleibt fortan bestehen! Sodass bereits ab dem zweiten Testtermin mehrsprachig aufwachsende Buben geringfügig mehr Items richtig lösen. Die Mittelwerte der Rohscores finden sich unter dem jeweiligen Diagramm.

6.1.5 Itemanalyse

Aufgrund des Austausches jedes siebten Wortes durch ein MC-Testitem, sollte über den gesamten MAZE Text hinweg grundsätzlich ein recht ausgewogenes Verhältnis von Wortarten in den Items bestehen. Dabei ist nicht auszuschließen, dass gewisse Items aus inhaltlichen oder syntaktischen Gründen, sowie aufgrund der Wortart schwieriger zu lösen sind als andere. Da es sich bei den Items um Mehrfachwahl(MW)-Aufgaben handelt, muss auch ein gewisser Zufallseinfluss bei der Aufgabenbearbeitung einkalkuliert werden. Lienert und Raatz (1998) empfehlen in diesem Zusammenhang eine zufallskorrigierte Formel (S. 75) für den Schwierigkeitsindex:

$$P = 100 \frac{N_R - \frac{N_F}{m-1}}{N}$$

Variablen in dieser Formel:

N_R = Anzahl der Pbn, die das Item richtig gelöst haben

N_F = Anzahl der Pbn, die das Item falsch beantwortet haben

N = Gesamtzahl der Pbn

m = Anzahl der Wahlmöglichkeiten der MW-Aufgabe

Je mehr Pbn ein Item richtig gelöst haben, desto näher liegt der resultierende Wert beim Maximum von 100. Um den Einfluss mangelnder Sprachkenntnis auf den Schwierigkeitsindex zu minimieren, wurden nur die Daten von Kindern mit ausschließlich deutscher Muttersprache zur Itemanalyse herangezogen.

In den folgenden Tabellen werden nur jene Items aufgeführt, die von zumindest 10 Pbn bearbeitet wurden. Solange nichts über die Arbeitsstile der Pbn bekannt ist, wird davon ausgegangen, dass besonders gute Leser den MAZE Test im Allgemeinen schnell UND genau, also mit wenigen Fehlern bearbeiten können. Daher würde der Schwierigkeitsgrad jener Items, die später im Text vorkommen und somit nur noch von sehr wenigen, leistungsstarken Pbn bearbeitet wurden womöglich unterschätzt werden. Grundsätzlich sollte bei genauem Lesen, für Kinder der vierten Schulstufe jedes Item lösbar sein. Dementsprechend sind tendenziell recht hohe Werte (= geringe Schwierigkeit) von mindestens 80,0 zu erwarten.

Hase

Hase	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12
P =	94,1	80,3	90,1	100,0	84,2	100,0	94,1	92,0	70,8	100,0	100,0	97,8
n =	76	76	76	76	76	75	76	75	72	74	71	68

h13	h14	h15	h16	h17	h18	h19	h20	h21	h22
97,8	95,2	97,2	100,0	96,3	89,3	100,0	100,0	100,0	70,0
67	62	54	52	40	28	27	21	13	10

Tab. 13.1: Item-Lösungswahrscheinlichkeiten für MAZE Text "Hase"

Wolf

	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12
P =	98,2	100,0	81,7	88,9	94,5	98,2	98,1	92,6	98,1	98,1	98,1	86,2
n =	82	82	82	81	82	82	81	81	80	79	78	76

w13	w14	w15	w16	w17	w18	w19	w20	w21	w22	w23
93,9	89,1	56,0	91,2	100,0	100,0	100,0	100,0	100,0	70,0	100,0
74	69	58	51	45	34	29	23	18	15	13

Tab. 13.2: Item-Lösungswahrscheinlichkeiten für MAZE Text "Wolf"

Haushuhn

	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12
P =	89,7	96,6	96,6	86,0	100,0	89,7	92,9	98,3	94,8	94,8	100,0	98,2
n =	87	87	87	86	86	87	85	86	86	86	84	82

c13	c14	c15	c16	c17	c18	c19	c20	c21	c22	c23	c24	c25
52,5	84,8	95,9	89,1	97,4	87,2	84,6	45,0	100,0	78,6	100,0	76,9	85,0
79	79	73	69	58	47	39	30	24	21	18	13	10

Tab. 13.3: Item-Lösungswahrscheinlichkeiten für MAZE Text "Haushuhn"

Elefant

	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12
P =	98,3	100,0	93,0	96,5	92,9	98,3	84,1	94,8	98,3	96,5	77,3	96,4
n =	86	86	86	86	85	86	85	86	86	86	86	84

e13	e14	e15	e16	e17	e18	e19	e20	e21	e22	e23	e24	e25
96,4	92,7	94,4	93,9	97,9	92,4	94,9	79,0	74,4	78,6	100,0	93,8	85,7
83	82	81	74	70	59	59	50	41	35	30	24	21

e26	e27	e28
100,0	67,9	100,0
15	14	11

Tab. 13.4: Item-Lösungswahrscheinlichkeiten für MAZE Text "Elefant"

Wie erwartet, zeigen sich bei den angeführten Items zum überwiegenden Teil Schwierigkeitsparameter über 80,0. Kleinere Abweichungen nach unten lassen sich wahrscheinlich noch durch den Zeitdruck in der Testsituation, oder einen Arbeitsstil der mehr auf Geschwindigkeit als auf Genauigkeit ausgelegt ist erklären. Es konnten jedoch drei Items identifiziert werden, deren Schwierigkeitsparameter zwischen 45,0 und 56,0 liegt (in der Tabelle grau hinterlegt). Diese Items werden daher auf inhaltliche Unstimmigkeiten untersucht.

Bei Item *w15* (Wolf-Text) dürfte den Kindern womöglich das Wort „Unterarten“ noch nicht geläufig sein:

... es bei den Wölfen sehr viele **Unterarten - vermarkten - Kanten** gibt, sehen sie in manchen Ländern...

Bei *c13* (Haushuhn) wurde vielleicht der Plural von „Fasanen“ oft nicht beachtet. Wurde im Wortspeicher des Pb nur der Singular „Fasan“ aktiviert, so würde die Wahl des Distraktors „einem“ den Satz zumindest grammatikalisch sinnvoll ergänzen:

Sie sind zum Beispiel mit **stehen - einem - den** Fasanen und Pfauen verwandt.

Die hohe Schwierigkeit von c20 (Haushuhn) ist schwer zu erklären, vor allem, da die beiden vorangehenden Items im selben Satz erwartungsgemäß häufig gelöst wurden. Inhaltlich erscheint der Teilsatz schlüssig, womit die Fehlerhäufigkeit wahrscheinlich auf die hohe Satzlänge oder ein Missverständnis des rückbezüglichen „die“ rückführbar ist.

*...eine Lautäußerung, die man bei **[keinem - allen - Quallen]** Hühnern häufig hört, und vielfältige Bedeutungen **[ist - matt - hat]**.*

Die durchschnittlichen Lösungswahrscheinlichkeiten pro MAZE Text zeigt die folgende Tabelle:

	Hase T1	Wolf T2	Haushuhn T3	Elefant T4
P =	93,1	92,7 (94,4)	88,6 (92,0)	91,7

Tab. 14: durchschnittliche Lösungswahrscheinlichkeit der Items pro Text; Werte nach Ausschluss der Ausreißer befinden sich in Klammern

Bleiben bei der Berechnung der durchschnittlichen Lösungswahrscheinlichkeit die drei Ausreißer unberücksichtigt, so ergeben sich sehr homogene Werte für die nach Texten gruppierten Items, von P = 91,7 bis 93,1.

6.1.6 Validität

Es werden die Ergebnisse des ersten und letzten MAZE Texts mit den Rohscores von beiden Messzeitpunkten des SLS 1-4 korreliert. Nach Bortz und Döring (2006) sind Koeffizienten über 0,60 für die Kriteriumsvalidität als hoch zu bewerten (S. 202). Es ergibt sich ein signifikanter Zusammenhang, sowohl zwischen dem ersten MAZE Text (Hase) und der ersten SLS 1-4 Vorgabe, $r = .73$, $p < .05$, als auch zwischen dem letzten MAZE Text (Elefant) und der zweiten SLS 1-4 Vorgabe, $r = .77$, $p < .05$.

Im Unterschied zum SLS sollten Leistungssteigerungen im MAZE-Test aber nicht gänzlich durch Änderungen in der Lesegeschwindigkeit zustande kommen, sondern hauptsächlich durch eine Verbesserung in einer anderen Teilfertigkeit des Lesens, nämlich dem Leseverständnis. Eine quantitativ äquivalente Leistungszunahme dieser beiden Teilfähigkeiten über den Untersuchungszeitraum wäre zwar möglich, aber statistisch gesehen eher unwahrscheinlich.

Um den Zusammenhang zwischen Änderungen in der Lesegeschwindigkeit und Änderungen im Leseverständnis zu ermitteln, werden die Leistungsdifferenzen zwischen dem ersten und letzten MAZE Test, sowie zwischen erster und zweiter Vorgabe des SLS 1-4 errechnet und miteinander korreliert.

Daraus ergibt sich eine signifikante, wenn auch niedrige Korrelation zwischen den beiden Veränderungsmaßen, $r = .29$, $p < .05$.

Fasst man die Ergebnisse der Validierung zusammen, so zeigt sich, dass der MAZE Test einerseits in hohem Maße mit dem SLS 1-4 als Beispiel für einen anerkannten Lesetest korreliert (Hypothese 5), und andererseits die *Leistungsveränderungen* im MAZE Verfahren nur zu 8,64 % durch Änderungen in der Lesegeschwindigkeit (SLS 1-4) mitbedingt werden. Dass ein Großteil der Veränderungsvarianz hierbei unerklärt bleibt, kann als Hinweis gedeutet werden, dass der MAZE Test tatsächlich einen von der Lesegeschwindigkeit weitgehend unabhängigen Aspekt der Lesefertigkeiten erfasst!

7 Diskussion

Ziele dieses Diplomprojekts waren die Entwicklung und Erprobung eines neuen, engmaschig einsetzbaren Testverfahrens zur Messung des Leseverständnisses. Vor allem sollte hiermit eine Möglichkeit aufgezeigt werden, auch über längere Zeiträume die Entwicklung von Lesefähigkeiten objektiv und verlässlich erfassen zu können. Die Einhaltung der strengen Richtlinien des CBM als Garant für Effizienz und Zuverlässigkeit wurden dabei eingehend geprüft.

7.1 Etablierung im Unterricht

In den ersten Gesprächen mit den Schulleitungen und dem Lehrpersonal wurden wiederholt Fragen und Bedenken bzgl. des Zeitaufwands für dieses Projekt geäußert. Neben dem regulären Unterricht müssen die Klassenlehrerinnen organisatorische Tätigkeiten, schulinterne Projekte, wie auch regelmäßig nationale und internationale Lernstandserhebungen durchführen, womit für weitere Projekte generell wenig Zeit bleibt. Umso wichtiger war es, die Vorzüge des MAZE Tests, insbesondere dessen einfache und schnelle Anwendbarkeit, in der Praxis zu

demonstrieren und somit eventuell den Weg für weitere Projekte zu ebnen. Bereits ab dem zweiten Testtermin, also nach Beendigung der Einschulungsphase, wurde wiederholt positives Feedback von Seiten des Lehrpersonals bzgl. des verhältnismäßig geringen Zeitaufwands und der Abwechslung für die Schüler gegeben. Bei einem Großteil der Pbn konnte ein deutliches Leistungsstreben und auch Wettbewerbsdenken bei der Testbearbeitung beobachtet werden, was für eine hohe Akzeptanz des MAZE Testformats spricht. Auch die Anzahl von Pbn, die den Test unsachgemäß bearbeitet hatten, hielt sich in äußerst geringen Grenzen und die betreffenden Testbögen ließen sich aufgrund der damit einhergehenden großen Zahl von Auslassungen und Fehlern eindeutig identifizieren und aussortieren.

Es konnte statistisch nachgewiesen werden, dass es zwischen mehreren Paaren von Messzeitpunkten zu signifikanten Leistungssteigerungen der Gesamtstichprobe kommt, womit dem MAZE Test eine hinreichende Sensibilität für Leistungsveränderungen bei kurzen Zeitintervallen (4 Wochen) attestiert werden kann!

7.2 Auffälligkeiten im Leistungsverlauf

Bei diversen Gruppenvergleichen (Quartile, Geschlecht, Sprachkenntnis) zeigten sich jedoch teils unerwartete Leistungsverläufe, vor allem zwischen den ersten beiden Testterminen. Das oberste Leistungsquartil, wie auch die Gruppen mehrsprachig aufwachsender Mädchen und einsprachig aufwachsender Buben wiesen vom ersten zum zweiten Testtermin einen negativen Leistungsverlauf auf. Bei den beiden Folgeterminen nimmt die Leistung hingegen erwartungsgemäß zu.

7.2.1 Qualität der MAZE Texte

Ein möglicher Grund für diesen „Knick“ im Leistungsdiagramm könnte in der (mangelnden) Äquivalenz der MAZE Texte liegen. Die Kennwerte der Paralleltestreliabilität sprachen auf den ersten Blick für eine geringfügig niedrigere Übereinstimmung des MAZE Texts „Hase“ mit den übrigen Texten. Unter der Annahme, dass die Erinnerungseffekte bei allen Texten in etwa gleich stark zum tragen kamen, wurden diese Kennwerte ebenfalls für die Wiederholungsmessungen ermittelt, wobei keine Benachteiligung des „Hase“ Texts mehr feststellbar war.

Es erscheint somit naheliegend, dass sich die Arbeitsstrategien einiger Pbn vom ersten zum zweiten Testtermin verändert hatten. Womöglich tendierten sie beim ersten Termin noch zu einem eher langsamen und sorgfältigen Arbeitsstil, der aufgrund der im Durchschnitt äußerst niedrigen Itemschwierigkeit bereits ab dem zweiten Termin zu Gunsten eines mehr auf Geschwindigkeit bedachten Stils verworfen wurde. Dies würde die anfänglich geringere Äquivalenz des „Hase“ Texts mit den anderen Parallelformen erklären, und kann als Beleg gewertet werden, dass der Übungstext, welcher ohne Zeitgrenze gemeinsam mit dem Testleiter bearbeitet wurde, nicht ausreichte um die Pbn hinreichend auf die eigentliche Testsituation vorzubereiten. Für künftige Untersuchungen mit dem MAZE Verfahren wird daher explizit die Verwendung eines Übungstexts von gleicher Länge und unter denselben Bedingungen wie bei der eigentlichen Testung empfohlen. Die Pbn hätten somit Gelegenheit, Erfahrungen mit den spezifischen Anforderungen des Testformats zu sammeln und sich geeignete Strategien zu Recht zu legen.

Um die Leistungsstagnation hin zum zweiten Testtermin gänzlich aufzuklären, bedarf es noch einer genaueren Betrachtung der Ergebnisse der Itemanalyse. Dabei erwiesen sich die Items Nummer 15 beim „Wolf“-Text (T2), sowie die Nummer 13 und 20 beim „Haushuhn“-Text (T3) als übermäßig schwierig im Vergleich mit dem übrigen Itempool. Betrachtet man nun die Mittelwerte des Quartilvergleichs (siehe Tab. 2), so wird ersichtlich, dass ein Großteil der Pbn des obersten Leistungsviertels alle diese Items bearbeitet haben muss. Da aber nur rund die Hälfte der betreffenden Pbn diese richtig lösen konnten, führte das zunächst zu einem Knick in der Leistungskurve (T2), und bei T3 trotz einer höheren Anzahl bearbeiteter Items nur zu einem Testwert ähnlich dem Ausgangswert. Aus dem untersten Leistungsviertel nahmen hingegen nur wenige Pbn diese Items überhaupt in Angriff, womit sich deren Testleistungen im Gruppenschnitt durchwegs in positiver Richtung entwickeln konnten. Des weiteren lässt sich auch der verhältnismäßig große Leistungssprung der Gesamtstichprobe von T3 zu T4 durch den störenden Einfluss der beiden äußerst schwierigen Items im „Haushuhn“-Text erklären.

Das Ziel, eine Paralleltestreliabilität von mindestens $r = 0,80$ zu erreichen, konnte zwar bei diesem Projekt noch nicht gänzlich realisiert werden, sollte aber durch oben genannten Maßnahmen zur Verbesserung der Übungsphase und eine Überarbeitung der auffallend schwierigen Items mit relativ geringem Aufwand zu bewerkstelligen sein. Trotz der genannten Schwächen korrelieren die MAZE Texte mit durchschnittlich $r = 0,732$ bereits signifikant hoch miteinander, was in Anbetracht des unterschiedlichen Inhalts und variierender Satzkonstruktionen innerhalb und zwischen den Texten durchaus beachtlich ist.

Dies spricht auch indirekt für die Qualität der 3.nWS von Bamberger und Vanecek (1998). Dem Anschein nach reichen zumindest bei Texten auf Volksschulniveau die Kennwerte *Satzlänge* und *Anzahl Mehrsilber* aus, um verlässliche Aussagen über den Schwierigkeitsgrad eines Texts machen zu können. Da die meisten Textverarbeitungsprogramme mittlerweile dazu in der Lage sind, Wörter im Text zu zählen, muss lediglich die Anzahl der Sätze und der Mehrsilber manuell vorgenommen werden, wodurch sich diese Formel äußerst zeiteffizient einsetzen lässt.

7.2.2 Einfluss von Geschlecht und Sprachkenntnis

Der in Bildungsdiskussionen oft vorgebrachte Risikofaktor „Mehrsprachigkeit“ konnte in dieser Untersuchung nicht bestätigt werden. Zwar wurden Unterschiede in den durchschnittlichen Testleistungen mehrsprachig und einsprachig aufwachsender Kinder festgestellt, diese erwiesen sich aber als nicht signifikant. Die geschlechtsspezifischen Unterschiede fielen ebenfalls nicht signifikant und zudem geringer aus. Auffällig waren jedoch die zueinander konträren Leistungsdifferenzen zwischen mehrsprachig und einsprachig aufwachsenden Kindern in Abhängigkeit vom Geschlecht. Während einsprachige Mädchen den mehrsprachig aufwachsenden je nach Testtermin im Durchschnitt ca. 2 bis 4 Punkte überlegen waren, verloren die einsprachigen Buben ihren (wenn auch nur geringen) Leistungsvorsprung vom ersten zum zweiten Testtermin und erreichten auch bei den folgenden Terminen nicht das Leistungsniveau der mehrsprachigen Buben. Ob dieses Ergebnis nun eher im Sinne einer erfolgreichen Bewältigung des multiplen Spracherwerbs seitens mehrsprachig aufwachsender Buben oder einer zufallsbedingt leistungsschwachen

Stichprobe exklusiv deutsch sprechender Buben zu interpretieren ist, bleibt allerdings offen.

Dabei sollte die Mehrsprachigkeit allgemein nicht als isoliert wirkender Faktor in der Entwicklung der Lesefertigkeiten betrachtet werden. Schließlich konnte bei PISA 2009 gezeigt werden, dass diesbezügliche Leistungsunterschiede, sofern überhaupt vorhanden, je nach Land unterschiedlich ausgeprägt sind (Herzog-Punzenberger & Schnell, 2012). Die Autoren fanden heraus, dass in Österreich etwas weniger als die Hälfte der sprachabhängigen Leistungsdifferenz bei der Lesekompetenz durch sozioökonomische Hintergrundfaktoren erklärbar sind. Für zukünftige Projekte mit dem MAZE Verfahren wäre es daher angebracht, auch die Lebensumstände und den Bildungshintergrund der Eltern mit zu erheben, um noch systematischer die Ursachen für Leistungsunterschiede in den Lesefertigkeiten nachweisen zu können.

7.3 Ausblick

Dieser (nach Kenntnisstand des Autors) erste wissenschaftliche Einsatz der MAZE Prozedur in Österreich, lieferte wichtige Erkenntnisse zu den Qualitäten und Potentialen dieses Verfahrens. Zwar werden die Grundregeln zur Konstruktion äquivalenter Parallelversionen erst mit zunehmender Erfahrung und Geübtheit einfacher anzuwenden, sodann ergeben sich aber aufgrund der Möglichkeit von Langzeitstudien deutlich mehr Anwendungsfelder als bei vielen anderen Lesetests, von denen nur ein oder zwei Versionen vorliegen. Durch eine Erweiterung des MAZE Textpools könnten die Leseleistungen objektiv und engmaschig über ein ganzes Schuljahr oder sogar noch länger beobachtet werden und den Lehrpersonen zeitnah Hinweise liefern, von welchen Lehrmethoden ihre Schüler besonders profitieren. Die hohe Augenscheinvalidität dieses Testverfahrens für Lehrpersonen, sowie der geringe Zeitaufwand liefern darüber hinaus eine gute Grundlage für dessen flächendeckende Etablierung im Unterricht.

Sollte sich der MAZE Test (z.B. durch Extremgruppenvalidierung) als geeignetes Instrument zur Feststellung einer Leseschwäche erweisen, so könnte man ihn als äußerst zeiteffizientes Screeningverfahren großflächig einsetzen. Zudem verlangen Anwendung und Auswertung dieses Verfahrens kaum nach wissenschaftlich-psychologischem Vorwissen und können somit nach einer entsprechenden Einschulung dem Lehrpersonal selbst überlassen werden.

Derartigen Projekten müsste jedenfalls noch eine ausführlichere Validierung anhand von differenzierten Leseverständnistests vorangehen, um aufzuklären, welche Teilaspekte des Leseverständnisses mit MAZE vorwiegend erfasst werden können.

8 Literaturverzeichnis

- Amstad, T. (1978). *Wie verständlich sind unsere Zeitungen?* Diss., Universität, Zürich.
- Bamberger, R. & Vanecek, E. (1984). *Lesen-Verstehen-Lernen-Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Wien: Jugend und Volk.
- Barnitzky, H. & Bunk, H. (Hrsg.). (2007a). *Lasso Lesebuch 3*. Wien: Österreichischer Bundesverlag Schulbuch GmbH & Co. KG.
- Barnitzky, H. & Bunk, H. (Hrsg.). (2007b). *Lasso Lesebuch 4*. Wien: Österreichischer Bundesverlag Schulbuch GmbH & Co. KG.
- Bergmüller, S. & Wintersteller, A. (2012). Lesen: Verteilung der Schüler/innen auf Kompetenzstufen. In B. Suchań, C. Wallner-Paschon, S. Bergmüller & C. Schreiner (Hrsg.), *PIRLS & TIMSS 2011 Schülerleistungen in Lesen, Mathematik und Naturwissenschaft in der Grundschule* (S. 14-15). Graz: Leykam.
- Best, K. (2006). Sind Wort- und Satzlänge brauchbare Kriterien zur Bestimmung der Lesbarkeit von Texten? In S. Wichter & A. Busch (Hrsg.), *Wissenstransfer – Erfolgskontrolle und Rückmeldungen aus der Praxis* (S. 21-31, Transfer Wissenschaften, Bd. 5). Frankfurt am Main: Peter Lang GmbH.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4., überarbeitete Aufl.). Heidelberg: Springer Medizin Verlag.
- Bradley, J. M., Ackerson, G. & Ames, W. S. (1978). The reliability of the Maze procedure. *Journal of Literacy Research*, 10 (3), 291-296.
- 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*, 22 (6), 1197-1211.
- Deno, S. L. (2003). Curriculum-Based Measures: development and perspectives. *Assessment for Effective Intervention*, 28 (3), 3-12.
- Field, A. (2009). *Discovering statistics using SPSS. Third edition*. Los Angeles: Sage.

- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32 (3), 221-233.
- Herzog-Punzenberger, B. & Schnell, P. (2012). Die Situation mehrsprachiger Schüler/innen im österreichischen Schulsystem – Problemlagen, Rahmenbedingungen und internationaler Vergleich. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012. Band 2.* (S. 229-267). Graz: Leykam.
- Kintsch, W. (1998). *Comprehension – A paradigm for cognition.* Cambridge: Cambridge University Press.
- Klicpera, C. & Gasteiger-Klicpera, B. (1993). *Lesen und Schreiben - Entwicklung und Schwierigkeiten.* Bern: Huber.
- Klicpera, C. & Gasteiger-Klicpera, B. (1995). *Psychologie der Lese- und Schreibschwierigkeiten - Entwicklung, Ursachen, Förderung.* Weinheim: Beltz, Psychologie Verlags Union.
- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens.* Göttingen: Hogrefe.
- Lenhard, W. (2013). *Leseverständnis und Lesekompetenz.* Stuttgart: Kohlhammer.
- Lenhard, W. & Schneider, W. (2006). *Ein Leseverständnistest für Erst- bis Sechstklässler.* Göttingen: Hogrefe.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Mayringer, H. & Wimmer, H. (2003). *Salzburger Lese-Screening (SLS).* Bern: Huber.
- Moll, K & Landerl, K. (2010). *SLRT II – Lese- und Rechtschreibtest. Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT).* Bern: Huber.
- Proyer, R., Wagner-Menghin, M. & Grafinger, G. (2006). *Leseverständnistest.* Mödling: Schuhfried.
- Schabmann, A., Landerl, K., Bruneforth, M. & Schmidt, B. A. (2012). Lesekompetenz, Leseunterricht und Leseförderung im österreichischen Schulsystem. Analysen zur pädagogischen Förderung der Lesekompetenz. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012. Band 2.* (S. 17-70). Graz: Leykam.

- Schmider, E., Ziegler, M., Danay, E., Beyer, L. & Bühner, M. (2010). Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*, 6(4), 147-151.
- Shin, J., Deno, S. L. & Espin, C. (2000). Technical adequacy of the Maze task for Curriculum-Based Measurement of reading growth. *The Journal of Special Education*, 34 (3), 164-172.
- Walter, J. (2008). Curriculumbasiertes Messen (CBM) als lernprozessbegleitende Diagnostik: Erste deutschsprachige Ergebnisse zur Validität, Reliabilität und Veränderungssensibilität eines robusten Indikators zur Lernfortschrittsmessung beim Lesen. *Heilpädagogische Forschung*, 34 (2), 62-79.
- Walter, J. (2009). Eignet sich die Messtechnik „MAZE“ zur Erfassung von Lesekompetenzen als lernprozessbegleitende Diagnostik? *Heilpädagogische Forschung*, 35 (2), 62-75.
- Walter, J. (2011). Die Entwicklung eines auch computerbasiert einsetzbaren Instruments zur formativen Messung der Lesekompetenz. *Heilpädagogische Forschung*, 37 (3), 106-126.
- Wiley, H. I. & Deno, S. L. (2005). Oral reading and MAZE measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education*, 26 (4), 207-214.
- Wikipedia. Elefanten [online]. URL: <http://de.wikipedia.org/wiki/Elefant> [14.12.2012].
- Wikipedia. Flusspferd [online]. URL: <http://de.wikipedia.org/wiki/Nilpferd> [04.01.2013].
- Wikipedia. Hasen [online]. URL: <http://de.wikipedia.org/wiki/Hasen> [14.12.2012].
- Wikipedia. Haushuhn [online]. URL: <http://de.wikipedia.org/wiki/Huhn> [14.12.2012].
- Wikipedia. Wolf [online]. URL: <http://de.wikipedia.org/wiki/Wolf> [14.12.2012].

9 Anhang

9.1 Zusammenfassung (deutsch)

In diesem Projekt wurde ein neuer Leseverständnistest auf Basis der im deutschen Sprachraum bisher wenig beachteten MAZE Technik erstellt, und auf seine Eignung für die schulische Lernverlaufsdiagnostik bei 149 Schülern der vierten Schulstufe überprüft. Dabei handelt es sich um eine spezielle Form von Lückentexten, die im Multiple-Choice Format von den Probanden gelöst werden sollen. Die Entwicklung erfolgte nach den Grundsätzen des Curriculum-Basierten Messens nach Deno (2003), mit dem Ziel, auch kleinere Leistungsveränderungen, wie sie sich in Zeitintervallen von nur vier Wochen ergeben können, zeiteffizient und genau zu erfassen. Um die Äquivalenz der dafür notwendigen Parallelversionen zu gewährleisten, wurde die Lesbarkeit der Texte mit Hilfe der 3. neuen Wiener Sachtextformel (Bamberger & Vanecek, 1984) quantifiziert und auf ein homogenes Niveau gebracht.

Der MAZE Test erwies sich als hinreichend sensibel, um signifikante Leistungsveränderungen auch bei kurzen Zeitintervallen festzustellen. Während jeder einzelne Text die für Leistungstests geforderte Retestreliaibilität von $r = 0,80$ mehr als erfüllen konnte, lagen die Werte für die Paralleltestreliaibilität mit durchschnittlich $r = 0,732$ hinter den Erwartungen zurück. Dies lag vermutlich an einer zu kurz gehaltenen Übungsphase und einzelnen Items mit unerwartet niedriger Lösungswahrscheinlichkeit. Des weiteren zeigte sich eine hohe Korrelation zwischen dem MAZE Test und dem Salzburger Lese-Screening (SLS), als Indikator für die basale Lesefähigkeit.

Außerdem ergab sich eine signifikante Wechselwirkung zwischen dem Geschlecht und der Sprachkenntnis der Pbn, welche sich in deutlich unterscheidbaren Leistungsverläufen und Leistungsdifferenzen zwischen ein- und mehrsprachig aufwachsenden Kindern in Abhängigkeit vom Geschlecht äußerte.

9.2 Abstract (english)

In this project, a new reading comprehension test was developed with the MAZE procedure, which is still hardly recognized in German applied research. Its adequacy for measuring reading comprehension growth over time was tested on a sample of 149 students of 4th class primary school. The MAZE procedure is a particular kind of cloze, which includes a high number of multiple choice items. The development of this test was based on the principles of Curriculum-Based Measurement (Deno, 2003), with the goal of being able to identify even very small changes in reading skills reliably and efficiently, that occur within short intervals of time (4 weeks). To ensure equivalence of the test forms (of the same task), the readability of the texts was assessed, using the "3. neue Wiener Sachtextformel" (Bamberger & Vanecek, 1984) – a formula for quantifying the reading ease through several text related factors.

The MAZE test proved to be sufficiently sensitive for identifying significant changes in students reading skills for short intervals if time! A retest-reliability of above $r = 0,80$ was achieved for every single test form, but on the other hand the equivalence of the four forms didn't fulfill the expectations with a mean of $r = 0,732$, being below the aforementioned limit. This was probably due to the (insufficiently) short introduction to the new test procedure and a few exceedingly difficult items.

The MAZE Test also showed a high positive correlation to the SLS, a test for measuring reading speed and also an indirect indicator for reading comprehension.

Furthermore there seemed to be an interaction between sex and the habitual language use, resulting in a differentiated skill progression for single- and multi-language users according to sex.

9.3 Fragenkatalog für Schüler

Bitte lies Dir diese Fragen genau durch und beantworte sie:

Wie viele Jahre bist du alt?

Bist du ein Bub oder ein Mädchen?

Welche Sprachen sprichst du?

Welche Sprache(n) sprichst du mit deiner Mutter?

Welche Sprache(n) sprichst du mit deinem Vater?

9.4 MAZE – Instruktion

(Übungstext „Nilpferd“ austeilen)

Ihr habt in der Schule sicher schon mal Lückentexte gesehen, wo ihr an bestimmten Stellen das passende Wort reinschreiben musstet. Heute machen wir etwas Ähnliches.

Wer von euch will mir den ersten Satz des Textes vorlesen, den ich euch mitgebracht habe? Ja bitte, und die anderen lesen bitte still mit.

„Dankeschön!“

Wie ihr sehen könnt steht ungefähr in der Mitte des Satzes eine eckige Klammer in der 3 Wörter drinstehen. Eure Aufgabe ist es, genau das Wort einzuringen, das am Besten an diese Stelle des Satzes passt.

Also wer kann mir sagen welches Wort hier am besten reinpasst?

„Ja richtig!“

Wer will mir den zweiten Satz vorlesen?

Welches Wort passt am Besten in die erste Klammer?

....

Welches in die Zweite?

=> bei Problemen: Hinweis dass manche Klammern schon recht schwierig sind.

Denkt daran: in jeder Klammer ist immer nur 1 Wort wichtig!

Den Rest dieses Übungstextes bearbeitet jetzt bitte jeder alleine und wir vergleichen dann die Lösungen.

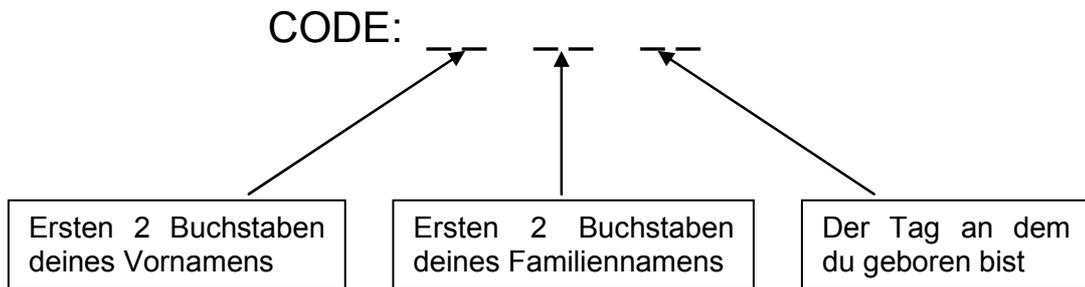
Es kann passieren, dass ihr plötzlich merkt, dass ihr eigentlich ein anderes Wort in einer Klammer einringen wolltet. In diesem Fall streicht das falsch eingeringte Wort durch und ringt das Richtige ein.

....nach dem Vergleich....

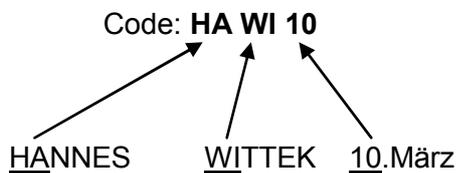
Wie ich sehe habt ihr alle verstanden wie der Text funktioniert. Für den Text, den ihr jetzt bekommt habt ihr 2:30. Minuten Zeit. Es ist völlig normal, wenn ihr in dieser Zeit nicht fertig werdet, denn der Text ist wirklich **sehr** lang. (Auch Erwachsenen gelingt das meistens nicht.). Versucht also möglichst genau, aber auch schnell zu arbeiten!

Auf „LOS“ dreht ihr bitte das Blatt um und beginnt zu arbeiten. Und auf „STOP“ hört ihr auf und legt sofort die Stifte weg. Anschließend sollt ihr dann noch euren Code unten rechts eintragen.

9.5 Probandencode und Übungs-MAZE Text „Nilpferd“



Beispiel:



Bitte schreibe jetzt oben deinen Code hin!

Übungstext: **Das Nilpferd**

Das Nilpferd ist eines der größten **[am - zum - Kamm]** Land lebenden Säugetiere. Sie leben in **[ein - wer - der]** Nähe von Gewässern im mittleren und **[natürlich - südlichen - blauen]** Teil Afrikas. Den Großteil des Tages **[klettern - gelingen - verbringen]** Nilpferde schlafend oder ruhend im Wasser. **[Mai - Dabei - Um]** sieht man zumeist ihre Augen, Ohren **[und - rund - weil]** Nasenlöcher aus dem Wasser herausragen. In **[das - wer - der]** Nacht kommen die Nilpferde aus dem **[lassen - Wasser - Buch]** heraus und suchen nach Nahrung. Sie **[entfernen - verwandeln - Scheren]** sich manchmal mehrere Kilometer vom Wasser **[dumm - da - um]** ausreichend Gräser zum Fressen zu finden.

9.6 MAZE Text „Hase“

Der Hase

Hasen gehören zu den Säugetieren und **[man - kann - jemand]** findet sie auf fast allen Kontinenten. **[Bauch - Auch - Sowie]** in Österreich sind sie heimisch. In **[einem - er - der]** freien Natur werden sie Feldhasen genannt. **[Sie - Nie - Wir]** können bis zu 68 Zentimeter lang **[Pferden - werden - sehen]**, haben große Ohren und lange Hinterbeine. **[Damit - Ohne - Schnitt]** können sie sehr schnell laufen und **[von - Riss - bis]** zu 2 Meter hoch springen. Ihr **[überstellen - Rückenfell - Kran]** ist zumeist braun, grau und etwas **[merklich - kühl - gelblich]**. Die Farben können sich aber auch **[vermischen - gehen - fischen]**. Kopf und Brust sind hellbraun und **[eine - Fass - das]** Fell am Bauch ist cremeweiß. Wenn **[sie - wie - euch]** eine Gefahr erkennen, dann drücken sie **[mir - Licht - sich]** zum Schutz in eine Erdmulde und **[Kuchen - versuchen - schauen]** möglichst ruhig zu sitzen. Nur wenn **[die - Vieh - ein]** Gefahr zu groß wird, ergreifen sie **[dem - die - Knie]** Flucht. Sie sind meist in der **[lacht - Zeit - Nacht]** und zur Dämmerung aktiv. Im Spätwinter **[Hund - und - aber]** Frühling sieht man sie auch am **[Tag - lag - Monat]**. Sie fressen die grünen Teile von **[Glanz - Dingen - Pflanzen]**, aber auch Knollen und Wurzeln. Im **[hinter - Winter - Woche]** ernähren sie sich von der Rinde **[junger - Hunger - eckiger]** Bäume.

Die Feldhasen leben normalerweise als **[verlängern - Einzelgänger - Leute]**, aber in der Paarungszeit suchen die **[Männchen - scheren - Vögel]** nach einer Partnerin. Dabei kann es **[neben - mischen - zwischen]** den Männchen zu Streitereien kommen. Sie **[Magen - rennen - jagen]** sich gegenseitig und boxen einander mit **[den - einen - wen]** Vorderpfoten. Die Feldhasen sind auch mit **[Dänen - die - den]** Hauskaninchen verwandt. Diese werden von vielen **[Katzen - Märchen - Menschen]** daheim als Haustiere gehalten. Anders als **[das - die - wie]** Feldhasen leben Kaninchen lieber gemeinsam als **[Paar - nah - Packung]** oder in Gruppen. Man hält sie **[darin - im - nimm]** Käfig oder draußen in einem Gehege. **[riechen - Kaninchen - Bienen]** gibt es in sehr vielen Farben **[und - Mund - damit]** mit

unterschiedlich langem Fell. Sie sind **[auch - Bauch - da]** deutlich kleiner als ihre frei lebenden **[Verwandten - ernannten - Dinge]**.

9.7 MAZE Text „Wolf“

Der Wolf

Der Wolf gehört zu den Raubtieren **[sowie - und - Mund]** ist mit den Hunden verwandt. Wölfe **[springen - leben - Mähnen]** im Rudel, welches zumeist aus einem **[Wecker - Elternpaar - elegant]** und dessen Nachkommen besteht. Das Elternpaar **[ist - tut - Mist]** im Regelfall dominant gegenüber den Jungtieren. **[Mager - Weil - Daher]** kommt es auch zu keinen Rankämpfen **[zwischen - Kirschen - über]** den Tieren. Sobald die Jungtiere geschlechtsreif **[können - Rind - sind]**, verlassen sie ihre Eltern. Anschließend machen **[er - sie - nie]** sich auf Partnersuche, um selbst eine **[Familie - fliegen - Lilie]** zu gründen. Sie sehen so ähnlich **[Maus - darin - aus]** wie große Haushunde, haben aber einen **[längeren - Kerker - hölzernen]** Körper und ihre Brust liegt höher. **[Wir - Sie - Knie]** haben einen recht großen Kopf mit **[einer - deiner - Leimen]** breiten Stirn, sowie einen buschigen Schwanz. **[Ohne - Hahn - Da]** es bei den Wölfen sehr viele **[Unterarten - vermarkten - Kanten]** gibt, sehen sie in manchen Ländern **[gehen - schleunig - deutlich]** anders aus, als in Europa. Zum **[Beispiel - Freispiel - Sache]** hat der Polarwolf in Grönland ein **[Heuriger - deutlich - brennend]** längeres, weißes Fell. Damit ist er **[im - nimm - links]** Schnee gut getarnt und kommt mit **[das - Meer - der]** Kälte gut zurecht.

Wölfe sind Fleischfresser **[und - Hund - weil]** gehen im Rudel auf die Jagd. **[Kuh - Zu - An]** ihrer Beute zählen zum Beispiel Hirsche **[und - Schlund - weil]** Elche. Aber auch Hasen und andere **[Beine - kleine - leere]** Pflanzenfresser werden von ihnen erlegt. Wenn **[Vieh - uns - sie]** nicht genug Nahrung in ihrem Revier **[finden - saugen - lindern]**, dann fressen sie auch Abfall oder **[die - den - Mine]** Überreste bereits toter Tiere. Sie verlassen **[euch - nicht - sich]** bei der Nahrungssuche sehr stark auf **[euer - ihren - wirren]** guten Geruchssinn. Im Zoo verhalten sich **[Wölfe - Höhlen - Leitern]** im Rudel oft anders, als in **[Eier - glatter - freier]** Wildbahn. Es kann zu Rankämpfen in **[Meer - der - den]** Gruppe kommen. Das Rudel wird dann **[schlicht - schon - nicht]** mehr unbedingt von einem Elternpaar angeführt.

[Hier - Bier - Unter] wird das Tier zum Anführer, das **[riecht - sich - ihn]** gegen die anderen im Kampf durchsetzt.

9.8 MAZE Text „Haushuhn“

Das Haushuhn

Das Haushuhn wird bei uns meistens **[schwierig - einfach - Eintracht]** als Huhn bezeichnet, und auf Bauernhöfen **[gehalten - verhindert - verwalten]**. Das Männchen wird Hahn genannt und **[den - was - das]** Weibchen nennt man Henne. Die Federn **[der - mehr - dessen]** Hühner können viele verschiedene Farben haben. **[Gegen - Kanzlei - Dabei]** ist das Gefieder der Hähne prächtiger, **[falls - als - von]** das der Hennen. Auch der rote **[wann - Mais - Kamm]** am Kopf der Tiere ist bei **[den - wehen - einer]** Männchen merkbar größer. Außerdem unterscheidet sich **[er - eine - der]** Hahn von der Henne durch seinen **[sichelförmigen - lauten - Stein]** Schwanz. Weltweit sind mehr als hundert **[Terrassen - Dinge - Hühnerrassen]** bekannt und es werden immer wieder **[neue - Reue - schiefe]** gezüchtet. Sie sind zum Beispiel mit **[stehen - einem - den]** Fasanen und Pfauen verwandt. Das Krähen **[des - dessen - einer]** Hahnes dient zur Markierung seines Reviers **[Schlund - und - hingegen]** man hört ihn zumeist zum Sonnenaufgang, **[zur - Lurch - am]** Mittagszeit und am Abend. Grundsätzlich können **[ihr - nie - sie]** aber zu jeder Tageszeit krähen. Das **[Gackern - albern - Laufen]** ist eine Lautäußerung, die man bei **[keinem - allen - Quallen]** Hühnern häufig hört, und vielfältige Bedeutungen **[ist - matt - hat]**. Sie können damit einander vor Gefahren **[warnen - singen - Schaben]**, anderen Hühnern drohen oder sie anlocken. **[Darunter - Mischen - Zwischen]** den Hühnern herrscht eine Hackordnung. Dabei **[dämpfen - kämpfen - lesen]** sie zum Beispiel um den höchsten **[Schlafplatz - Katze - Schatz]** im Stall. Um solche Rankämpfe zu **[lesen - vermeiden - erteilen]**, werden ihre Sitzstangen in möglichst gleicher **[Höhe - Löwe - Teller]** montiert. In der freien Wildbahn fressen

[früher - Bilder - Hühner] sowohl Gras und Körner, wie auch **[Insekten - erwecken - Wolken]** und Schnecken. Sogar Mäuse zählen manchmal **[Kuh - zu - am]** ihren Beutetieren. Bei der Suche nach **[Tasten - Nahrung - Mahnung]** scharren sie oft mit den Füßen **[daneben - Kauf - auf]** dem Boden. Auf Bauernhöfen werden sie **[dagegen - erlegen - oder]** meist mit Körnerfutter ernährt. Hühner können **[Riss - bis - von]** zu 300 Eier im Jahr legen, **[wenn - renn - dabei]** man sie ihnen regelmäßig wieder wegnimmt. **[Fenster - Kosten - Ansonsten]** brüten sie ihre Eier aus, was **[ungefähr - Verkehr - ganz]** drei Wochen dauert.

9.9 MAZE Text „Elefant“

Der Elefant

Elefanten sind die größten, am Land **[schreibenden - lebenden - elende]** Säugetiere. Schon bei der Geburt wiegen **[sie - Knie - wir]** knapp hundert Kilo. Im Erwachsenenalter können **[nie - ihr - sie]** eine Schulterhöhe bis zu vier Metern **[erreichen - trinken - Eiche]** und fünf Tonnen schwer werden. Man **[schreibt - erkennt - entkernt]** sie vorallem an ihrem langen Rüssel **[und - Mund - aber]** an den Stoßzähnen. Der Rüssel wird **[auf - Mohn - von]** ihnen auf viele verschiedene Weisen genutzt. **[Es - Sie - Reh]** atmen nicht nur durch ihn, sondern **[nutzen - Stutzen - zaubern]** ihn auch zum Riechen, Tasten und **[Schleife - lesen - ergreifen]** von Nahrung. Zum Beispiel reißen sie **[um - damit - Bandit]** Grasbüschel aus dem Boden, um sie **[zu - Kuh - auf]** fressen. Aufgrund der starken Muskeln im **[küssen - Glas - Rüssel]** können sie sich sogar gegen Feinde **[verteidigen - geheiligt - stehlen]**. Ihre langen Stoßzähne bestehen aus Elfenbein **[rund - weil - und]** werden ebenfalls als Waffe eingesetzt. Außerdem **[sitzen - können - föhnen]** sie damit die nahrhafte Rinde einiger **[Fenster - Baumsorten - Traumort]** abkratzen.

Wegen ihrer Größe und dem **[neben - Haus - Leben]** in einer Herde, haben sie nur **[wenige - ewige - Leder]** natürliche Feinde. Eine Herde besteht aus **[Kühen - Liedern - Mühe]** und deren Kälbern. Sie wird von **[leider - keinem - einer]** älteren und erfahrenen Leitkuh angeführt, die **[ein - der - wer]** Herde als Vorbild dient. Die Männchen **[trennen - Menge - fliegen]** sich im Alter von etwa zwölf **[planen - Jahren - Haare]** von der Herde und leben dann **[Schmalz - gleich - als]** Einzelgänger oder in losen Gruppen. In **[der - ein - schwer]** Brunftzeit suchen sie wieder eine Herde **[im - Hauch - auf]**, um sich zu paaren. Die Größe **[bunt - und -**

aber] das Wanderverhalten der Herde richten sich **[Krach - auf - nach]** dem Nahrungsangebot. Wenn es nur wenig **[Nahrung - warum - Zahn]** in ihrer Nähe gibt, dann bilden **[ihr - oder - sie]** eher kleinere Gruppen. Sind Gräser und **[Blätter - reden - Kabel]** im Überfluss vorhanden, dann bilden mehrere **[Miene - Familien - viele]** eine große Herde. In trockenen Gebieten **[legen - Regen - kauen]** sie oft lange Strecken zurück, um **[Rechner - Nahrung - Mahnung]** zu finden. Diese Wege werden auch **[Elefantenstraßen - verlassen - Katzen]** genannt.

9.10 Lesbarkeitsindex für Texte aus Schulbüchern

Variable	LASSO Lesebuch 4					LASSO Lesebuch 3				
	4.Kl.					3.Kl.				
	S. 72-73	S.144	S.120	S. 35-36	S. 89	S. 68-69	S. 130	S.19	S. 48	S. 107
Anzahl Sätze	65	25	26	38	15	51	46	30	19	29
Anzahl Silben	720	600	627	723	345	593	626	381	324	649
Anzahl Wörter	525	372	329	424	216	352	419	242	197	381
Wörter mit mind. 3 Silben	29	52	53	71	30	39	29	33	26	58

Satzlänge (Wörter):	8,08	14,88	12,65	11,16	14,40	6,90	9,11	8,07	10,37	13,14
Wortlänge (in Silben):	1,37	1,61	1,91	1,71	1,60	1,68	1,49	1,57	1,64	1,70
Anteil Mehrsilbiger (>=3 Silben):	5,52	13,98	16,11	16,75	13,89	11,08	6,92	13,64	13,20	15,22

Flesch (Amstad) =	91,69	70,77	55,86	69,09	72,16	74,55	83,49	79,83	73,42	67,21
3.nWS =	2,06	5,86	6,07	5,97	5,74	3,48	2,67	4,46	4,77	5,90

Durchschnitt 3.Klasse Flesch 75,70
WSTF 4,26

Durchschnitt 4.Klasse Flesch 71,91
WSTF 5,14

9.11 Lesbarkeitsindex für MAZE Texte

Variable	HASE	WOLF	HAUSHUHN	ELEFANT
Anzahl Sätze	23	20	21	20
Anzahl Silben				
Anzahl Wörter	252	251	255	253
Wörter mit mind. 3 Silben	36	37	36	34

Satzlänge (Wörter):	10,96	12,55	12,14	12,65
Anteil Mehrsilbiger (>=3 Silben):	14,29	14,74	14,12	13,44

3.nWS =	5,21	5,64	5,38	5,28
----------------	-------------	-------------	-------------	-------------

9.12 LEBENS LAUF

Persönliche Daten

Name	Hannes WITTEK
Geburtsdatum, Geburtsort	10. März 1985, Wiener Neustadt
Staatsbürgerschaft	Österreich

Ausbildung

seit 2004	Diplomstudium Psychologie Schwerpunkte: Angewandte Entwicklungspsychologie und Psychologische Diagnostik
1995 – 2003	BRG Gröhrmühlgasse, Wr. Neustadt Schwerpunkt: Naturwissenschaften
1991 – 1995	Volksschule Sta. Christiana Wr. Neustadt

Berufserfahrung

Dez. 2011 – Feb. 2012	Psychologisches Praktikum am KH Hietzing mit Neurologischem Zentrum Rosenhügel - Neuropsychiatrische Abteilung für Kinder und Jugendliche mit BHZ
seit 2008	Ehrenamtlicher Trainer beim 1.Karate-Kickbox- Club Wr. Neustadt
seit 2005	Geringfügige Beschäftigung im EDV Bereich
Okt. 2003 – Sep. 2004	Zivildienst im LPH Wiener Neustadt