

Interkantonale Hochschule für Heilpädagogik Zürich
Departement 2, Studiengang Logopädie

Auf dem Weg zu einem neuen Wortschatztest

Bachelor-Arbeit

Eingereicht von: Esther Glaus
Begleitung: Prof. Dr. Anke Sodogé

Eingereicht am 20. 02.2017

Abstract

Für sehr viele Bereiche der Sprach-Diagnostik sind keine normierten, an die Schweiz angepassten Tests vorhanden, so auch nicht für die Überprüfung des rezeptiven Wortschatzes von Kindergartenkindern. Das Ziel dieser Arbeit ist es, die Entwicklung eines solchen Tests zu beginnen, eine erste Version zu erproben und die erhaltenen Daten zu analysieren.

In der Planungsphase wurden 57 Items generiert und zusammen mit je drei Ablenkern auf Vierfeldertafeln bildlich dargestellt. Unter den Ablenkern befinden sich jeweils ein semantischer und ein phonologischer Ablenker, um später in einer Fehleranalyse qualitative Informationen über die Beeinträchtigung gewinnen zu können. Das Verfahren wurde an 140 Kindergartenkindern aus fünf Gemeinden der Kantone Zürich und St. Gallen erprobt. Die Ergebnisse dieser Untersuchung zeigen, dass die Gütekriterien Reliabilität und Objektivität erfüllt sind und dass die Itemschwierigkeiten und Trennschärfen in einem akzeptierbaren Bereich liegen. Es besteht allerdings ein Überangebot an zu leicht zu lösenden Items.

In einer nächsten Phase der Entwicklung können Daten für eine Normierung und die Berechnung der Validität erhoben werden.

Dank

Ohne die Unterstützung von über hundert Kindergärtnern, von vielen Kindergärtnerinnen und mehreren Schulleitungen wäre diese Bachelor-Arbeit nicht möglich gewesen. Ihnen allen danke ich für ihren Einsatz, die freundliche Aufnahme im Kindergarten und die Zusage zum Projekt. Für die Kindergärtnerinnen war es nicht immer einfach eine Lücke in ihrem gut ausgelasteten Kalender zu finden. Deshalb geht ein besonderes Dankeschön an sie alle, d.h. an Bea Brägger, Gerarda Caputo, Franziska Gubser, Brigitta Landolt, Marisa Müller, Natascha Imhof, Barbara Keller, Marianne Zoller, Manuela Aschwanden, Ivana Gojkovic, Anita Kessler, Fatima Martin, Nadja Schmalz, Bettina Solinger und Verena Stenz.

Inhaltsverzeichnis

1. Einleitung.....	1
2. Theoretischer Hintergrund.....	3
2.1 Semantisch-lexikalische Störungen	3
2.1.1 Einleitung zu semantisch-lexikalischen Störungen im Kindergartenalter	3
2.1.2 Diagnostik bei semantisch-lexikalischen Störungen.....	5
2.1.3 Bedeutung von Wortschatz-Tests.....	5
2.1.4 Itembasierte Wortschatztests.....	6
2.1.5 Verschiedene Wortarten in einem Wortschatztest.....	6
2.1.6 Diagnostische Verfahren bei semantisch-lexikalischen Störungen.....	6
2.2 Deutschschweizer Dialekte.....	10
2.3 Theorie zu Testentwicklung und Testanalyse.....	11
2.3.1 Einteilung diagnostischer Verfahren.....	11
2.3.2 Itemverteilungen, Itemanalyse und Distraktorenanalyse	12
2.3.3 Messgütekriterien für diagnostische Verfahren	13
2.3.4 Normierung oder Cut-Off.....	14
3. Methodik	15
3.1 Auswahl und Anzahl der Test-Items	16
3.2 Anzahl und Auswahl der Ablenker.....	17
3.3 Darstellung des Tests und der Items	18
3.4 Auswahl der Dialektregionen für die Testkonstruktion	19
3.5 Die Rolle des Dialekts bei der Itemauswahl und der Itempräsentation.....	21
3.6 Auswahl der Kindergärten für die Konstruktionsstichprobe.....	21
3.7 Durchführung der Erprobung	22
3.8 Grösse der Konstruktions-Stichprobe	23

3.9 Auswertung der Daten	23
4. Ergebnisse.....	24
4.1 Zusammensetzung der Konstruktionsstichprobe.....	24
4.2 Itemanalyse.....	26
4.2.1 Itemschwierigkeit.....	26
4.2.2 Trennschärfe	29
4.3 Itemausschluss.....	32
4.4 Distraktoren-Analyse	32
4.5 Gütekriterien	39
4.6 Normierung.....	39
4.7 Cut-Off	40
4.8 Wortliste	41
5. Diskussion.....	42
5.1 Itemanalyse.....	42
5.1.1 Itemschwierigkeit.....	42
5.1.2 Trennschärfe	43
5.2 Reliabilität	43
5.3 Distraktoren-Analyse	44
5.4 Verschiedene Wortarten in einem Wortschatztest	45
5.5 Fazit und Ausblick	45
Tabellenverzeichnis.....	47
Abbildungsverzeichnis.....	47
Literaturverzeichnis.....	48
Anhang	

1. Einleitung

Die im deutschsprachigen Raum vorhandenen diagnostischen Verfahren für semantisch-lexikalische Störungen basieren fast ausschliesslich auf Standardsprache. Für die Abklärungen von Kindergartenkindern werden aber vorwiegend Tests oder Screenings in Schweizerdeutsch benötigt. Zwar zeigte eine Studie von Landert (2007), dass Kinder in diesem Alter Standardsprache erstaunlich gut verstehen und teilweise auch sprechen, trotzdem ist anzunehmen, dass sie diese nicht so gut beherrschen wie die deutschen Kinder der jeweiligen Normierungsstichproben. Bei den aktiven Benenntests spielt dies eine untergeordnete Rolle, denn die schweizerdeutschen Antworten werden meist mit jenen der Standardsprache gleichgesetzt und die Normierung wird übernommen. Werden aber Wortverständnis-Tests ins Schweizerdeutsche übersetzt, wirken viele der oft eingesetzten phonologischen Ablenker nicht in der gleichen Weise wie in der Standardsprache, denn die Nähe zwischen Zielitem und phonologischem Ablenker kann sich ändern. Die Normierung kann also nicht einfach für die Schweiz übernommen werden.

Eine umfassende semantisch-lexikalische Diagnostik sollte sowohl den produktiven wie auch den rezeptiven Wortschatz überprüfen und die Lemma- und Lexemebene berücksichtigen. Aus diesem Grund hat Rupp (2008) für ihre modellgeleitete Diagnostik bei kindlichen lexikalischen Störungen eigens zwei Verfahren als Ergänzung zum produktiven AWST (Kiese & Kozelsky 1996) entwickelt. Sie benutzte dazu für ihr Wortverständnis-Screening und ihr wortfreies Semantik-Screening einen Teil der AWST-Items. Diese drei Verfahren zusammen, bildeten ein Instrument für eine umfassende, semantisch-lexikale Abklärung für Kindergartenkinder mit aufeinander abgestimmten Items in Standardsprache. Das Verfahren nach Rupp (2008) ist aber nicht normiert und nicht an die Schweiz angepasst. Beim Wortverständnis-Screening ist die Nähe zwischen phonologischem Ablenker und Zielitem nach der Übersetzung ins Schweizerdeutsche bei zehn von 30 Items deutlich verändert. Inzwischen ist auch die neue Version des AWST, der AWST-R (Kiese-Himmel 2005), mit einer etwas veränderten Itemliste erschienen. Die von Rupp (2008) verwendeten Items stimmen nun nicht mehr mit jenen des AWST-R überein.

Mit der PDSS von Kauschke und Siegmüller (2010) steht uns ein weiteres Verfahren zur Verfügung, das den produktiven und rezeptiven Wortschatz überprüft und das seit der zweiten Auflage auch normiert ist. Leider werden in diesem Verfahren keine phonologischen Ablen-

ker eingesetzt und somit ist keine Fehleranalyse bezüglich der Wortformen (Lexeme) möglich.

Auch der WWT 6-10 von Glück (2011) ist ein normiertes Verfahren, das aus produktiven und rezeptiven Wortschatztests besteht. Er wurde aber für Schulkinder entwickelt und kann für die jüngeren Kindergartenkinder nicht eingesetzt werden. Der WWT 6-10 enthält keine Anpassungen an die Schweiz. Übersetzt man den rezeptiven Untertest ins Schweizerdeutsche, verändert sich bei 34 der 95 Items die Nähe zwischen phonologischem Ablenker und Zielitem oder das Zielitem ist nicht gebräuchlich in der Schweiz.

Nicht nur bei den Zielitems und den phonologischen Ablenkern, sondern auch bei den Bildern werden bei manchen Tests Anpassungen an die Schweizer Begebenheiten nötig. So kommen z.B. im WWT 6-10 Bilder von Lebensmitteln, Lösen und einem Bankomaten vor, die so in der Schweiz nicht bekannt sind.

Es besteht also eindeutig ein Bedarf an rezeptiven Wortschatztests, die speziell für die Schweiz entwickelt und normiert wurden.

Ziel dieser Bachelorarbeit ist es, einen solchen schweizerdeutschen, rezeptiven Wortschatz-Test oder ein Wortschatz-Screening für das Kindergartenalter zu entwickeln, soweit es im Rahmen einer Bachelorarbeit möglich ist. Es sollen phonologische und semantische Ablenker zum Einsatz kommen und die Bilder sollen länderneutral oder an die Schweiz angepasst sein. Als Fernziel sollen später auch eine Normierung und die Evaluation der Gütekriterien möglich sein.

2. Theoretischer Hintergrund

2.1 Semantisch-lexikalische Störungen

2.1.1 Einleitung zu semantisch-lexikalischen Störungen im Kindergartenalter

Der Wortschatzerwerb ist zwar nie ganz abgeschlossen, im Kindergartenalter ist er aber schon sehr weit fortgeschritten. Sechsjährige beherrschen ca. 3000-5000 Wörter produktiv und 9000-14000 rezeptiv (Grimm 2000). Diese Einträge im Lexikon sind aber in diesem Alter oft noch ungenau und werden im sogenannten Slow mapping weiter ausdifferenziert. Die Wortarten sind bereits ausgewogen, d.h. der Muttersprache entsprechend, im mentalen Lexikon von Kindergartenkindern vorhanden.

Bei semantisch-lexikalischen Störungen (SLS) können die Beeinträchtigungen auf verschiedenen Ebenen liegen. Es können der Wortabruf, das Verständnis oder die qualitative und quantitative Abspeicherung betroffen sein. Es ist deshalb wichtig, eine differenzierte Diagnostik durchzuführen. Levelt (1989) versuchte in seinem Modell diese verschiedenen Ebenen darzustellen (siehe Abb. 1).

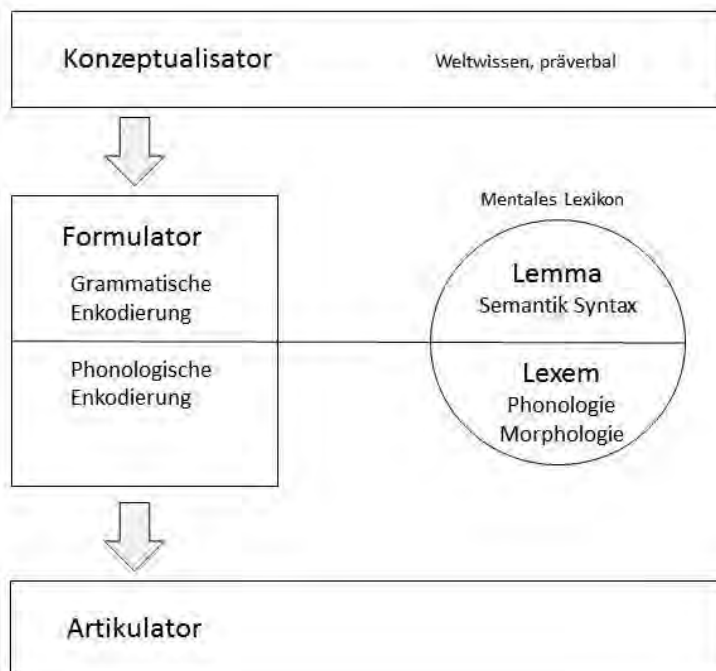


Abb.1: Sprachproduktions-Modell nach Levelt (1989). Vereinfachte Darstellung.

Gemäss dem Levelt-Modell (1989) steht am Anfang der Sprachproduktion eine noch präverbale Äusserungsabsicht, bei der auf die mentalen Konzepte zugegriffen wird. Ein Konzept ist ein Bündel von Wissen oder Vorstellungen zu einem Begriff. Hat ein Kind keine Konzepte von den Dingen dieser Welt, kann es auch keine Worte dazu abspeichern. Auf der folgenden Ebene, jener des Formulators, werden Wortbedeutungen (Lemmata) und Wortformen (Lexeme) aus dem mentalen Lexikon abgerufen und auf der Ebene des Artikulators werden die Worte in ein akustisches Signal umgewandelt.

Eine anderes Modell der Sprachproduktion und des mentalen Lexikons ist jenes von Dell (1986), bei welchem nicht von seriellen, sondern von konnektionistischen Verknüpfungen zwischen den einzelnen Ebenen ausgegangen wird. Das bedeutet, die Verarbeitung verläuft netzartig in mehrere Richtungen gleichzeitig (siehe Abb. 2). Die Semantik-Ebene entspricht den nonverbalen Konzepten im Levelt-Modell und auf der Wort-Ebene entstehen Lemmata und Lexeme.

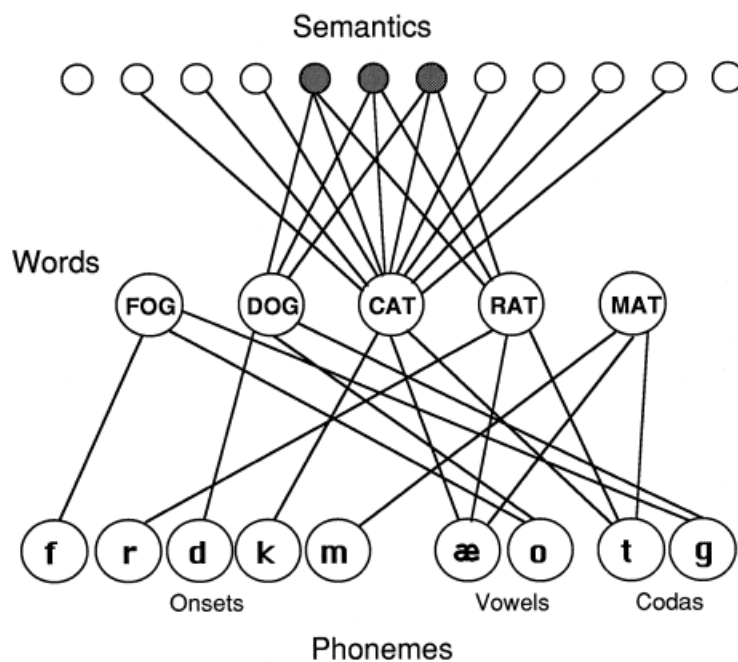


Abb. 2: Lexikon-Modell nach Dell (aus Dell, Chang und Griffin 1999).

Gemäss diesen Modellen können also semantisch-lexikale Störungen unterteilt werden in solche mit zu wenigen Konzepten oder zu wenigen Einträgen im mentalen Lexikon, in solche mit ungenauen Einträgen bezüglich Wortbedeutung und Wortform oder in solche mit Abrufproblemen.

2.1.2 Diagnostik bei semantisch-lexikalischen Störungen

Um spezifische Therapieentscheidungen treffen zu können, müssen die unterschiedlichen Teilleistungen der semantisch-lexikalischen Fähigkeiten abgeklärt und miteinander verglichen werden. Nur eine Kombination der verschiedenen Testresultate erlaubt Rückschlüsse auf die Störungsebene innerhalb einer SLS. Das heisst in erster Linie, nicht nur der produktive, sondern auch der rezeptive Wortschatz müssen abgeklärt werden. Sind beide ungenügend, so sind zu wenige Einträge im Lexikon vorhanden oder es sind noch nicht einmal Konzepte verfügbar. Ist nur der produktive Wortschatz eingeschränkt und der rezeptive in Ordnung, so könnte eine Wortabrufstörung vorliegen. Die Worte sind dann zwar im Lexikon vorhanden, deren Abruf ist aber nicht oder nur inkonstant möglich. Ebenfalls denkbar bei einer Diskrepanz zwischen produktiver und rezeptiver Leistung wäre eine qualitative Störung im Bereich der Wortbedeutung oder der Wortform. Dabei ist der Wortabruf ebenfalls inkonstant und er kann mit semantischen oder phonologischen Abrufhilfen erleichtert werden. Werden also in einem produktiven Tests diese Hilfen angeboten oder in einem rezeptiven Tests mit semantischen und phonologischen Ablenkern gearbeitet, ermöglicht dies wertvolle Rückschlüsse auf die Art der Störung.

2.1.3 Bedeutung von Wortschatz-Tests

Die Erfassung des Wortschatzes ist ein wichtiges Instrument für das Diagnostizieren von semantisch-lexikalischen und Sprachverständnis-Störungen und ist nötig für eine spezifische Therapieplanung. Bei zweijährigen Kindern wird sogar nur über die Erfassung des produktiven Wortschatzes entschieden, ob das Kind ein Risiko für eine Sprachentwicklungsstörung hat oder nicht. Später, im Kindergartenalter, fallen Kinder, die sich sprachlich nicht mitteilen können oder Sprache nicht verstehen durch ein abweichendes Sozialverhalten auf. Sie ziehen sich zurück oder werden aggressiv gegenüber ihren Spielgefährten. Die Geschichten, die die Kindergärtnerin erzählt, verstehen sie nicht, deshalb wirken diese Kinder oft abwesend und unaufmerksam. Später in der Schule, können sie dem Unterricht nicht folgen und die Schere zwischen sprachlich beeinträchtigten und unbeeinträchtigten Kindern geht immer weiter auseinander. Gemäss Mayer (2012) kommen semantisch-lexikalische Störungen auch sehr oft zusammen mit Lese-Rechtschreibstörungen vor.

Wortschatztests und weitere Verfahren für das Erkennen und therapieren von semantisch-lexikalischen Störungen sind deshalb gerade im Kindergartenalter wichtig für das Wohlbefinden des Kindes und seinen späteren Schulerfolg.

2.1.4 Itembasierte Wortschatztests

Ein von Glück (2011) und von Rupp (2008) favorisiertes diagnostisches Vorgehen bei Verdacht auf SLS ist das itembasierte Testen des Wortschatzes. Dabei bestehen die rezeptiven Untertests aus den gleichen Itemsätzen wie die produktiven. Zuerst wird die Produktion überprüft und nur die Items, die produktiv nicht vorhanden waren, werden später auch noch rezeptiv getestet. Es wird angenommen, dass die produktiv erworbenen Wörter auch rezeptiv erworben sind und so werden sie zum rezeptiven Wortschatz dazugezählt. Itembasiertes Testen führt also zu einer Zeitersparnis. Ein Nachteil dieses Vorgehens ist aber, dass das Zusammenstellen der Itemsets nicht einfach ist, denn der rezeptive Wortschatz ist viel grösser als der produktive und muss deshalb mit schwierigeren Items überprüft werden. Das Itemset muss also so unterschiedliche Schwierigkeitsgrade aufweisen, dass produktiv und rezeptiv das ganze Leistungsspektrum der Zielgruppen von unter- bis überdurchschnittlich abgedeckt ist.

2.1.5 Verschiedene Wortarten in einem Wortschatztest

Für eine spezifische Therapieplanung ist es auch wichtig, die Zusammensetzung des mentalen Lexikons bezüglich Wortarten genauer zu kennen. Gemäss Schulz und Grimm (2012) ist bei einer SLS der Verbwortschatz meist stärker betroffen als der Nomenwortschatz. Da in Wortschatztest oft nur die einfacher zu präsentierenden Nomen getestet werden, könnten dadurch SLSs unterdiagnostiziert werden. Es ist deshalb wichtig, Nomen und Verben in einen Wortschatz-Test aufzunehmen.

2.1.6 Diagnostische Verfahren bei semantisch-lexikalischen Störungen

Besteht ein Verdacht auf eine semantisch-lexikalische Störung, stehen mehrere Verfahren zur Verfügung, aber nicht alle sind normiert und nur wenige decken alle oder zumindest mehrere Ebenen einer SLS ab. Eine gute Übersicht über die vorhandenen diagnostischen Verfahren bei SLS findet sich in Rupp (2013). An dieser Stelle sollen nur die wichtigsten für das Kindergartenalter in Frage kommenden Verfahren kurz vorgestellt werden.

WWT 6-10 (Wortschatz- und Wortfindungstest für 6- bis 10-Jährige, Glück 2011)

Dies ist ein sehr umfangreiches, itembasiertes Verfahren für den produktiven und rezeptiven Wortschatz. Im Benenntest werden semantische und phonologische Abrufhilfen angeboten, im rezeptiven Teil kommen semantische, phonologische und neutrale Ablenker zum Einsatz. Die Items bestehen aus Nomen, Verben und Adjektiven. Der Wortabruf wird ebenfalls beachtet. Dieser Test wurde hauptsächlich für Schulkinder entwickelt und ist nur für die älteren Kindergartenkinder geeignet. Das Verfahren erfüllt die Gütekriterien und ist normiert.

Nachteil: Nicht alle phonologischen Ablenker und nicht alle Bilder sind für die Schweiz geeignet.

AWST-R (Aktiver Wortschatztest für 3- bis 5-jährige Kinder, Kiese-Himmel 2005)

Der AWST-R ist ein Bild-Benenntest, der 51 Nomen und 24 Verben beinhaltet. Qualitative Aussagen über Semantik und Wortformen werden durch eine Fehlerauswertung möglich. Der Wortabruf wird durch Beobachten beurteilt. Der Test erfüllt die Gütekriterien und er ist normiert.

Nachteil: Es ist ein rein produktives Verfahren. Es sind keine Normen für einen Grossteil der Kinder im zweiten Kindergartenjahr vorhanden.

Screening nach Rupp (2008)

Bei diesem Verfahren wird die Produktion über den AWST (Kiese & Kozielsky 1996), d.h. die Vorgängerversion des AWST-R (Kiese-Himmel 2005), abgeklärt. Die Rezeption wird mit zwei von Rupp entwickelten Verfahren überprüft, dem Wortverständnis-Screening und dem Semantik-Screening. Im Wortverständnis-Screening werden 30 Nomen zusammen mit je einem phonologischen und einem semantischen Ablenker präsentiert. Das Semantik-Screening überprüft sprachfrei die Konzepte durch Bilder-Zuordnungen. Die Items der drei Verfahren sind aufeinander abgestimmt. Das Wortverständnis- und das Semantikscreening von Rupp (2008) sind analysiert worden, es liegt jedoch keine Normierung vor.

Nachteil: Im Itemset fehlen die für den Wortschatz wichtigen Verben und der Einsatz von nur 2 Ablenkern erhöht die Ratewahrscheinlichkeit. Weil der AWST inzwischen revidiert und die Itemauswahl verändert wurde, ist die Itembasierung nicht mehr völlig vorhanden bei der Kombination mit dem heute gebräuchlichen AWST-R. Es ist keine an die Schweiz angepasste Version vorhanden.

MSVK (Marburger Sprachverständnistest für Kinder, Elben & Lohaus 2000)

Dieses Verfahren misst in sechs Untertests mehrere Dimensionen des Sprachverständnisses. Der rezeptive Wortschatz wird durch das Ankreuzen des Zielitems aus einer Auswahl mit 3 Ablenkern getestet. Die Ablenker sind phonologisch, semantisch oder neutral. Die insgesamt 24 überprüften Items bestehen aus Nomen, Verben und Adjektiven. Durch semantisches Kategorisieren werden die Konzepte überprüft. Satz-, Instruktionsverständnis und Pragmatik werden ebenfalls getestet. Das Verfahren ist für Kindergartenkinder ab 5;0 Jahren und Erstklässler geeignet. Es erfüllt die Gütekriterien und ist normiert.

Nachteil: Der produktive Wortschatz wird nicht überprüft. Eine Itemanzahl von 24 für die Überprüfung von 3 Wortarten ist etwas tief. Bei 13 von den 24 Items ist die Nähe zwischen phonologischem Ablenker und Zielitem im Dialekt verändert, verglichen mit der Standardsprache. Der MSVK hat keine an die Schweiz angepasste Version und nicht alle im Kindergarten vorkommende Alterskategorien sind abgedeckt.

PDSS (Patholinguistische Diagnostik bei Spracherwerbsstörungen, Kauschke & Siegmüller 2010)

Die PDSS ist ein Verfahren in drei Bänden, das die Sprachentwicklung auf allen linguistischen Ebenen sowohl produktiv als auch rezeptiv überprüft. Sie ist für Kinder im Alter von 2;0 bis 6;11 geeignet und deckt dadurch das ganze Kindergartenalter ab. Die Ebene der Semantik-Lexik wird durch Kategorisieren, Benennen, Verstehen und sprachfreies Zuordnen getestet. In den produktiven und rezeptiven Wortschatztests werden Nomen, Verben, Adjektive, Farbadjektive und Präpositionen mit insgesamt 66 Items getestet (7-20 Items/Wortart). Im rezeptiven Teil kommen nur semantischen Ablenker zum Einsatz, die in nahe und ferne eingeteilt werden. Die erste Auflage der PDSS war nicht ausreichend analysiert und nicht normiert. Seit der zweiten Auflage von 2010 erfüllt dieses Verfahren die Gütekriterien und 16 der 23 Untertests sind normiert.

Nachteil: Es sind keine Anpassungen an die Schweiz vorhanden.

SETK 3-5 (Sprachentwicklungstest für drei- bis fünfjährige Kinder, Grimm et al. 2010)

Mit diesem Verfahren wird die Sprachentwicklung in den Bereichen Sprachverständnis, Morphologische Regeln, Semantik und Gedächtnis überprüft. Das Sprachverständnis wird über das Verstehen von Sätzen abgeklärt und die Produktion über eine Bildbeschreibung. Der SETK 3-5 erfüllt die Gütekriterien und er ist normiert.

Nachteil: Der SETK 3-5 deckt die sprachlichen Fähigkeiten von Kindergartenkindern nur ungenügend ab. Das Verfahren enthält keine Wortschatztests und es wurden keine Anpassungen an die Schweiz vorgenommen.

SET 5-10 (Sprachstandserhebungstest für Kinder im Alter zwischen 5 und 10 Jahren, Petermann 2012)

Der SET 5-10 überprüft die sprachlichen Fähigkeiten in den Bereichen Sprachverständnis, Sprachproduktion, Grammatik und Gedächtnis. Der produktive Wortschatz wird durch Bildbenennung getestet und die Semantik über Kategorisieren. Auf der Satzebene wird das Sprachverständnis über das Ausagieren von Anweisungen, über Fragen zum Inhalt und über das Beurteilen von nicht korrekten Sätzen abgeklärt. Der SET 5-10 erfüllt die Gütekriterien und er ist normiert.

Nachteil: Dieses Verfahren enthält keinen rezeptiven Wortschatztest und es sind keine Anpassungen an die Schweiz erhältlich.

HSET (Heidelberger Sprachentwicklungstest, Grimm & Schöler 2001)

In diesem Verfahren für 4- bis 10-jährige Kinder werden in 13 Untertest vor allem morphosyntaktische Strukturen und Bedeutungen auf Satz- und Wortebenen überprüft. Das Verfahren ist standardisiert, es erfüllte weitestgehend die Gütekriterien und es war normiert.

Nachteil: Der Test beinhaltet keine Wortschatztests. Die Normierung stammt aus den 1970er Jahren und ihr kann deshalb auch gemäss Grimm und Aktas (2002) nicht mehr vertraut werden. Auch sei heute das der Entwicklung des HSET zugrunde liegende Sprachmodell nicht mehr aktuell.

HSET Schweizer Version (Grissemann, Baumberger und Hollenweger 1991)

Der HSET ist eines der ganz wenigen Verfahren, das auch eine Schweizer Version herausgegeben hat. Alle 13 Untertests des HSET wurden ins Schweizerdeutsche übersetzt, an die sprachlichen Unterschiede angepasst, analysiert und neu normiert. Die Schweizer Version verfügt also über eine aktuellere Normierung als die deutsche. Ansonsten ist er gleich aufgebaut wie die deutsche Version.

Nachteil: Auch diese Version beinhaltet keine Wortschatztests. Das Bildmaterial wirkt etwas veraltet.

Diese kurze Darstellung einiger bekannter Verfahren, die für die Überprüfung der semantisch-lexikalischen Fähigkeiten infrage kommen, zeigt, dass unter diesen nur drei (WWT 6-10, PDSS, Screening nach Rupp 2008) den Wortschatz rezeptiv und produktiv überprüfen, was für eine gute Diagnostik unbedingt notwendig wäre (Kauschke 2003, Rupp 2011). Schliesst man Screenings aus, so bleiben noch die PDSS und der WWT 6-10. Letzterer ist aber für einen grossen Teil der Kindergartenkinder aufgrund der Altersbegrenzung nicht geeignet. Es bleibt also nur noch die PDSS übrig, wenn man ein normiertes Verfahren für die produktive und rezeptive Überprüfung des Wortschatzes aller Altersklassen des Kindergartens wünscht. Angepasst an die Sprachsituation in der Schweiz ist nur eines der oben diskutierten Verfahren, die Schweizer Version des HSET, es enthält aber keine Wortschatztests.

2.2 Deutschschweizer Dialekte

Die in der Deutschschweiz gesprochene Umgangssprache wird vereinfachend meist als Schweizerdeutsch bezeichnet. Das Schweizerdeutsche gibt es aber nicht. Schweizerdeutsch ist ein Sammelbegriff für eine Vielfalt von alemannischen Dialekten in der deutschsprachigen Schweiz (Gallmann 2015). Diese Vielfalt ist so gross, dass die Autoren des Kleinen Sprachatlas der deutschen Schweiz (Christen, Glaser, Friedli & Renn 2013) vom Versuch die Varianten zu zählen sogar abraten. Es würden sich aber zumindest ganz automatisch Bezeichnungen für Dialekte finden, sobald grössere Dialektunterschiede zwischen den Regionen vorhanden sind. Dadurch besteht allerdings eine sehr ausgeprägte Kleinräumigkeit der Dialekte. In ländlichen Regionen hat jedes Dorf einen eigenen Dialekt und in grösseren Städten kann auch eine Dialektgrenze innerhalb der Stadt gebildet werden, z.B. durch einen Fluss. Dies war auch früher in Zürich der Fall, wo die Limmat Sprachgrenze bildete. Manchmal trennt ein Fluss oder See die Dialektregionen, manchmal auch Hügel und Berge.

Die Dialekte sind nicht nur zahlreich, sie lassen sich auch nicht klar gegeneinander abgrenzen. Man kann zwar die geografische Verbreitung eines einzelnen Dialekt-Wortes in Sprachatlanten ungefähr aufzeigen, für einen ganzen Dialekt-Wortschatz ist dies aber nur schwer möglich.

Will man trotzdem eine Ordnung in die Dialekte bringen, so kann man dies in einer sehr groben Weise tun indem man eine Nord-Süd-Grenze zwischen Hochalemannisch und Höchstalemannisch zieht. Eine weitere Unterteilung bietet die Brünig-Napf-Reuss-Linie, eine Ost-West-Grenze innerhalb des Hochalemannischen (Bickel, Schläpfer & Arquint 2000). Eine et-

was differenziertere und auch oft verwendete Einteilung ist die Gleichsetzung der Kantone mit Dialekträumen. Aber auch hier gilt wieder, es gibt nicht den Zürcherdialekt oder den St. Gallerdialekt. Im Zürcher Limmattal wird nicht der gleiche Dialekt gesprochen wie im Zürcher Oberland und der Dialekt des St. Galler Rheintals ist ein anderer als z.B. jener, der Stadt St. Gallen. Andererseits macht der Dialekt auch nicht an der Kantonsgrenze halt und so wird auf beiden Seiten einer Kantonsgrenze ein ähnlicherer Dialekt gesprochen als in weit auseinanderliegenden Gebieten innerhalb eines Kantons. Die Einteilung der Dialekte nach Kantonen greift also ebenfalls zu kurz und verletzt den Dialektstolz von vielen Regionen.

Eine weitere Schwierigkeit bei der Fassbarkeit der Dialekte ist, dass sie sich ständig verändern. Durch eine immer grössere Mobilität, häufigere Wohnortswechsel und eine Durchmischung der Bevölkerung verschwinden die Unterschiede zwischen den Dialekten. In früheren Zeiten verhinderten historische, kulturelle und konfessionelle Unterschiede und geografische Hindernisse eine Durchmischung der Dialekte. Heute ist dies durch eine gesteigerte Mobilität weniger der Fall. Verbindend bezüglich Sprache waren Verkehrswege hingegen immer schon. Die Schifffahrtswege auf dem Zürichsee oder die schon seit Jahrhunderten bestehende Verkehrswege auf dem Land führten zu Ähnlichkeiten entlang dieser Wege (Gallmann 2015). Die Veränderung der Dialekte über die Zeit führt dazu, dass Dialektwörterbücher und Sprachatlanten, die in mühsamer Kleinarbeit zusammengestellt werden, bei ihrem Erscheinen schon fast wieder veraltet sind. Das heisst, was in der heute vorhandenen Literatur beschrieben wird, gilt nur noch teilweise für die Sprache der heutigen Kinder.

2.3 Theorie zu Testentwicklung und Testanalyse

2.3.1 Einteilung diagnostischer Verfahren

Diagnostische Verfahren können einerseits in Tests und Screenings und andererseits in standardisierte und in informelle Verfahren eingeteilt werden.

Von einem **Test** wird nur dann gesprochen, wenn er psychometrisch untersucht wurde und bestimmte Gütekriterien erfüllt. Ein Test muss auch das ganze Leistungsspektrum der Untersuchten abbilden, d.h. er soll zwischen schwacher, durchschnittlicher und überdurchschnittlicher Leistung unterscheiden können. In einem **Screening** wird nur ein Cut-Off (Grenzwert) gesetzt, welcher die Untersuchten in Auffällige und Unauffällige oder in Personen mit oder

ohne Risiko einteilt. Ein Screening kann ein überprüftes, standardisiertes oder ein rein informelles Verfahren sein.

Bei **informellen Verfahren** wird keine Standardisierung oder Überprüfung durchgeführt. Die informellen Verfahren sind laut Kany und Schöler (2010) nur dann sinnvoll, wenn wenigstens gewisse Angaben zur Anwendung, Auswertung und Interpretation vorhanden sind. Rupp (2013) hingegen betrachtet informelle Verfahren als wichtige und notwendige Informationsquellen.

2.3.2 Itemverteilungen, Itemanalyse und Distraktorenanalyse

Bei der Entwicklung eines neuen Verfahrens müssen nach einer sorgfältigen Planung und der Datenerhebung im Feld die Items genauer untersucht werden. Dazu gehören die folgenden Methoden:

Itemrohwertverteilung:

Die in der Erprobung erhaltenen Daten des Tests müssen **normalverteilt** sein.

Itemanalyse und Itemselektion: Die Itemanalyse ist die Überprüfung der Güte der einzelnen Items. Sie beinhaltet die **Itemschwierigkeit** und die **Trennschärfe**. Die Itemschwierigkeit zeigt, wie viele Probanden das Item lösen konnten. Die Trennschärfe ist eine Korrelation der richtigen Itembeantwortung mit dem Gesamtergebnis des Tests. Werden Distraktoren (Ablenker) in einem Test eingesetzt, ist eine **Ratekorrektur** bei der Berechnung der Itemschwierigkeiten angezeigt. Mit einer steigenden Zahl von eingesetzten Distraktoren sinkt die Ratewahrscheinlichkeit. Items mit ungenügenden Analyseergebnissen werden aus dem Test entfernt.

Zusätzlich zur Bestimmung der Itemschwierigkeiten und Trennschärfen kann auch die **Homogenität** (Hierarchische Clusteranalyse) und die **Dimensionalität** (Faktorenanalyse) überprüft werden. Sie dienen der weiteren Reduktion der Items.

Distraktorenanalyse:

Eine Distraktoren-Analyse ist eine wichtige Voraussetzung für die Güte eines Mehrfach-Wahl-Aufgaben-Tests, d.h. eines Tests mit Distraktoren. Laut Lienert und Raatz (1998) sollten Testverfahren ohne Distraktoren-Analyse nicht als valide betrachtet werden. Die Distraktoren sollen gleich plausibel und gleich attraktiv für den nichtwissenden Probanden sein wie die richtige Antwort. Es muss auch verhindert werden, dass aus Informationen über die Dis-

traktoren in einem Ausschlussverfahren auf die richtige Antwort geschlossen werden kann. Mittring und Rost (2008) konnten zeigen, dass Probanden einen beachtlichen Anteil der Test-Items lösen konnten, ohne dass sie die Aufgaben dazu kannten, nur durch die Informationen, die sie aus den Distraktoren zogen.

Die Anzahl der Falschantworten sollen sich mehr oder weniger gleichmässig auf die vorhandenen Distraktoren verteilen. Werden z.B. bei einer Mehrfach-Wahl-Aufgabe mit 3 Distraktoren zwei von ihnen nie gewählt, gelten sie bei den Probanden als sehr unwahrscheinlich und die Rate-Wahrscheinlichkeit steigt auf 50 %. Schlechte Distraktoren müssen von einem diagnostischen Test ausgeschlossen werden (Lienert und Raatz 1998). Völlige Gleichverteilung bei der Besetzung der Distraktoren ist aber nicht zu erwarten und bei nur wenigen Falschantworten kann die Verteilung auch zufällig zustande kommen. In Lienert und Raatz (1998) findet man Tabellen, die eine untere und eine obere Grenzen der noch akzeptablen Distraktorenbelegung angeben.

Bei vielen Wortschatztests sind aber die Distraktoren nicht völlig neutral. Man wählt gezielt Ablenker die entweder eine semantische oder phonologische Ähnlichkeit mit dem Zielitem aufweisen. In diesem Falle können obige Regeln nicht exakt übernommen werden.

Wie die Items, können auch die Distraktoren auf Schwierigkeit und Trennschärfe untersucht werden. Beide Kennwerte sollten bei guten Distraktoren tief oder sogar negativ sein (Lienert & Raatz 1998). Die Reihenfolge der Distraktorenpräsentation kann ebenfalls eine Rolle spielen. Der erste Distraktor in einer Reihe wird tendenziell etwas häufiger gewählt als der letzte (Eheim 1977).

2.3.3 Messgütekriterien für diagnostische Verfahren

Der nächste Schritt in der Entwicklung neuer Verfahren ist die Überprüfung der Messgütekriterien. Folgende Kriterien müssen bei einem diagnostischen **Test** erfüllt sein:

Objektivität: Die Objektivität gibt den Grad der Unabhängigkeit des Testresultats von der Person, die den Test durchführt an. Unabhängigkeit oder Objektivität muss bezüglich Durchführung, Auswertung und Interpretation des Verfahrens bestehen. Objektivität entsteht, wenn der Test standardisiert durchgeführt wird. Das heisst es müssen Regeln bezüglich Durchführung, Auswertung und Interpretation aufgestellt werden. Nur so werden bei unterschiedlichen Untersuchern die gleichen Testresultate erhoben.

Reliabilität: Die Reliabilität sagt etwas über die Zuverlässigkeit einer Messung aus. Ein Test muss unabhängig von Situationseinflüssen zuverlässig die gleichen Resultate liefern. Die Testreliabilität gibt an, welcher Teil der gemessenen Varianz wahr ist und welcher Teil der Varianz durch Messfehler entstanden ist. Eine Methode die Reliabilität zu messen ist die Bestimmung von Cronbachs Alpha. Die Basis dieser Methode ist die Berechnung der Streuung der einzelnen Items bezogen auf den gesamten Test (= innere Konsistenz). Gemäss Bühner (2005) ist Cronbachs Alpha heute die Standardmethode für die Bestimmung der Reliabilität.

Validität: Die Validität ist ein Mass für die Gültigkeit eines Tests. Sie beschreibt, wie gut das Verfahren auch das testet, was es vorgibt zu testen. Die Validität wird bestimmt, indem man den Test mit anderen bereits bestehenden Tests vergleicht.

Objektivität, Reliabilität und Validität werden nur bei Tests angegeben. Bei einem **Screening** werden andere Gütekriterien erhoben. Es sind dies die Sensitivität und die Spezifität.

Sensitivität: Die Sensitivität ist das Mass für die richtigerweise als auffällig identifizierten Personen.

Spezifität: Die Spezifität ist ein Mass für den Anteil der als unauffällig Getesteten an den tatsächlich Unauffälligen. Eine Spezifität von 100 % würde also bedeuten, dass das Screening alle Unauffälligen richtig erkennt.

2.3.4 Normierung oder Cut-Off

Der letzte Schritt bei der Entwicklung eines diagnostischen Verfahrens ist die Normierung bei einem Test oder das Bestimmen eines Cut-Offs bei einem Screening.

Für eine Normierung müssen zuerst Daten anhand einer genügend grossen, repräsentativen Eichstichprobe erhoben werden. Für eine mittlere Güte soll ein $N=150-300$ angestrebt werden (Fisseni 1997). Eine weitere Voraussetzung ist die Normalverteilung der Eichstichproben-Resultate. Liegt diese vor, so werden die Rohwerte zuerst in z- und dann in T-Werte umgerechnet. T-Werte zwischen 40 und 60 werden einer durchschnittlichen Leistung zugeordnet, Werte darunter und darüber dementsprechend als unterdurchschnittlich und überdurchschnittlich bezeichnet. Liegt keine Normalverteilung vor, sollten nur Prozenträge angegeben werden.

Ein Cut-Off-Wert kann nicht berechnet werden. Er muss mit Hilfe von empirischen Erkenntnissen abgeschätzt und begründet werden.

3. Methodik

Eine Übersicht über die Vorgehensweise bei dieser Testentwicklung ist in Abb. 3 dargestellt. Im Rahmen dieser Arbeit konnte die Testentwicklung bis zur Beurteilung eines Teils der Gütekriterien vorangebracht werden. Die einzelnen Schritte sind unten näher ausgeführt.

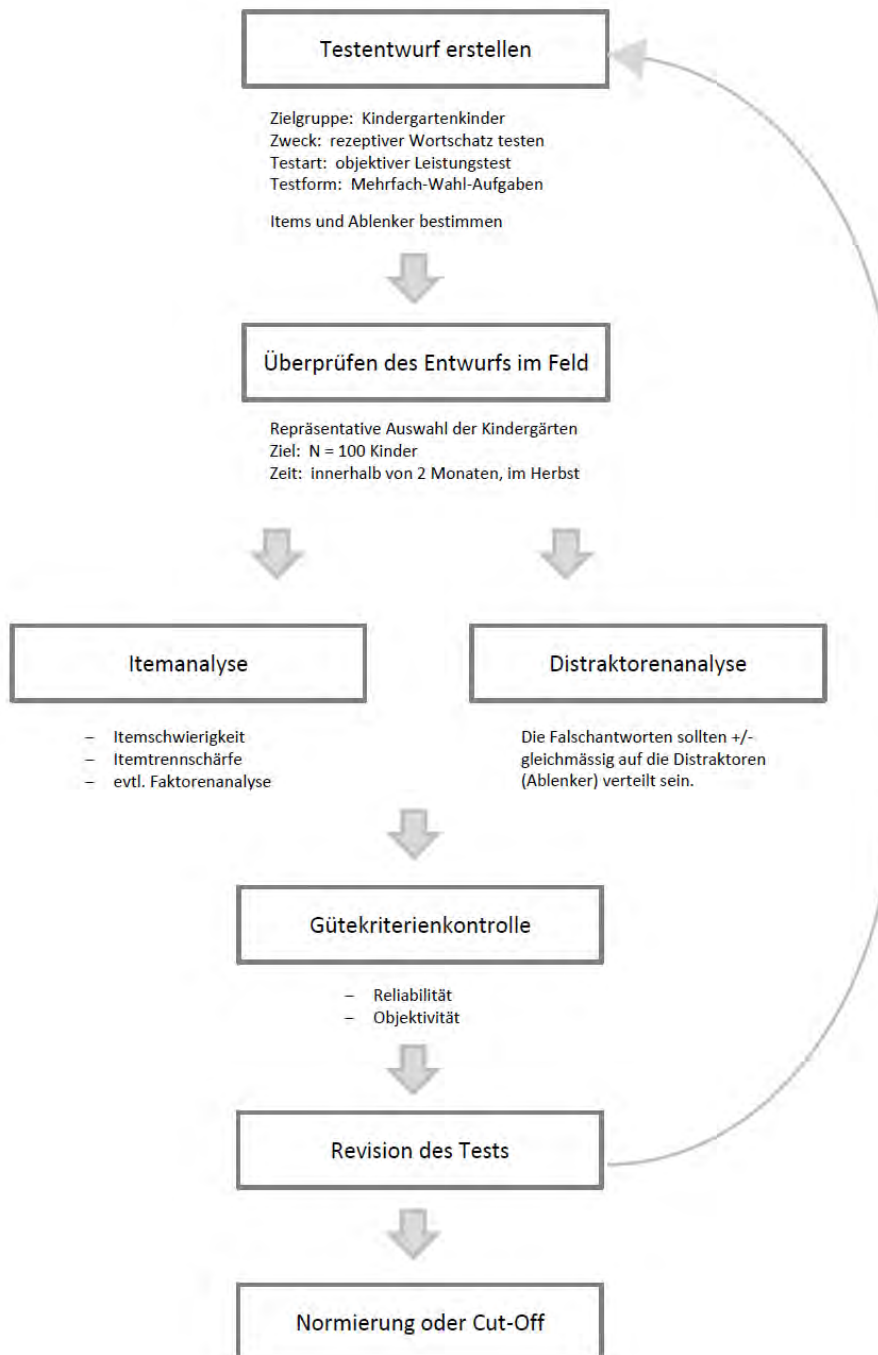


Abb. 3: Flussdiagramm der einzelnen Schritte bei der Entwicklung des rezeptiven Wortschatztests mit Mehrfach-Wahl-Aufgaben.

3.1 Auswahl und Anzahl der Test-Items

Die Testitems müssen diverse Bedingungen erfüllen. Als erstes sollen sie den rezeptiven Wortschatz der Zielgruppe der Kindergartenkinder widerspiegeln. Sie sollen aber auch unterschiedliche Schwierigkeitsgrade aufweisen, aus unterschiedlichen semantischen Feldern stammen und aus den Wortarten Nomen, Verben und Adjektive bestehen (siehe Tab. 1). Auch achtete ich darauf, dass die Items in der Erfahrungswelt von Jungen und Mädchen und von Stadt- und Landkindern vorkommen. Die Items müssen bildlich darstellbar sein und es sollen zu ihnen passende semantische und phonologische Ablenker zu finden sein. All diese Bedingungen zu erfüllen, erwies sich als eher schwierig. Aus einem ursprünglichen aufgestellten Pool von 200 Items erfüllten letztendlich nur 57 obige Bedingungen. Zum Teil mussten auch suboptimale Ablenker akzeptiert werden. Auf Kriterien wie Wortlänge und Verwendungshäufigkeit konnte bei der Auswahl nicht eingegangen werden.

Es wurde versucht, die Items aufgrund von Informationen aus der Literatur zu generieren. Da aber der rezeptive Wortschatz von Kindergartenkindern bereits sehr umfangreich ist, gibt es, im Gegensatz zum produktiven Wortschatz, keine Listen oder Datenbanken, die darüber Auskunft geben könnten. Für die Zusammenstellung des Itempools orientierte ich mich deshalb an den bereits bestehenden produktiven und rezeptiven Wortschatztests (AWST-R, WWT 6-10) und an Wortschatz fördernde Kinderbücher (z.B. Leue & Holthausen 2011). Es wurde darauf geachtet, eher Materialien für Siebenjährige zu berücksichtigen, damit der Schwierigkeitsgrad auch für die rezeptive Testung hoch genug ist.

Tab. 1: Nach semantischen Feldern geordnete Test-Items. Total 57 Items, 42 Nomen, 10 Verben, 5 Adjektive.

Semantisches Feld	Häufigkeit	Itemnummern
Natur	10	8, 19, 29, 37, 44, 46, 49, 50, 56, 57
Sonstiges	9	1, 4, 11, 12, 15, 31, 32, 35, 48
Gebäude, Bauten	8	2, 3, 9, 13, 16, 33, 38, 41
Küche, Haushalt	7	17, 20, 24, 39, 45, 53, 55
Fahrzeuge	7	7, 25, 36, 40, 51, 52, 54
Werkzeuge	5	6, 21, 33, 42, 47
im Haus	4	14, 22, 23, 30
Körper	3	5, 18, 28
Umgang mit Papier	2	26, 43
Möbel	1	10
Kleidung	1	27

Bei der Itemanzahl eines Tests muss ein Kompromiss zwischen Zeitaufwand, Belastbarkeit der Kinder und der Aussagekraft des Tests gefunden werden. Da zu erwarten ist, dass nach der Itemanalyse einige Items ausgeschlossen werden, muss die Anzahl in der Konstruktion deutlich höher sein als in der Endversion. Die für diesen Test verwendete Itemanzahl von 57 erachte ich als minimal für die Aussagekraft und als an der oberen Grenze bezüglich der Belastbarkeit von Kindergartenkindern.

3.2 Anzahl und Auswahl der Ablenker

Bei der Anzahl der Ablenker pro Item muss zwischen hoher Ratewahrscheinlichkeit und einer Überforderung der Probanden abgewogen werden. Ein einzelner Ablenker reicht nicht aus, denn dies würde eine zufällige Trefferrate von 50 % bedeuten. Meist werden deshalb zwei bis drei Ablenker empfohlen. Im Verfahren nach Rupp (2008) und in der PDSS entschieden sich die Autoren für zwei Ablenker und im WWT 6-10 wurden drei eingesetzt. Ich erachtete die Herabsetzung der Ratewahrscheinlichkeit als sehr wichtig und entschied mich deshalb für drei Ablenker.

Die Ablenker müssen in diesem Test vielen Anforderungen genügen. Da bei der späteren Testanwendung eine qualitative Auswertung der Fehler bezüglich Semantik und Wortform möglich sein soll, müssen die Ablenker einerseits phonologisch und semantische Bedingungen erfüllen und andererseits sollen sie für den nicht wissenden Probanden als Antwort grundsätzlich gleich plausibel sein wie die richtige Antwort. Idealerweise werden Ablenker durch Vortests ermittelt (Lienert und Raatz 1998). Eine Möglichkeit dazu wäre z.B. aus den in einem Benenntest gemachten Fehlern Ablenker für diesen rezeptiven Test zu generieren. Eine andere ist, Kinder gezielt nach Assoziationen zu den Items zu fragen. Rupp (2008) machte genau dies für die Bestimmung der Ablenker zu ihrem Semantik-Screening. Da die Kinder aber oft mit ganzen Geschichten und nicht mit einzelnen Wörtern antworteten, gestaltete sich die Auswertung als eher schwierig. Solche Vortests sind sehr aufwändig und übersteigen den Rahmen dieser Bachelorarbeit, deshalb verzichtete ich darauf und vertraute auf eine nachträgliche Distraktoren-Analyse.

Unten sind die Kriterien für die Ablenkerauswahl für diesen Test aufgelistet:

Für jeden Ablenker gilt:

1. er muss bildlich darstellbar sein
2. er soll ans Alter der Kinder angepasst sein
3. er darf nicht attraktiver als die Bestantwort (richtige Antwort) sein
4. er muss für einen Nichtbeantworter des Items als Antwort plausibel sein

Für den phonologischen Ablenker gilt

1. er kann sich in einem oder mehreren Lauten von der Bestantwort unterscheiden
2. er kann einen Reim zur Bestantwort bilden oder
3. er kann ein Minimalpaar dazu sein oder
4. er kann ähnliche Silben enthalten oder
5. er kann ähnliche Buchstabengruppen enthalten
6. er soll nicht gleichzeitig ein semantischer Ablenker sein

Dabei kommt auch der von Aitchison (1988) erwähnte „Badewanneneffekt“ zur Geltung, der besagt, dass ein Kind die Wortanfänge und Wortenden stärker beachtet, weil diese für das Kind aus dem Gesprochenen herausragen würden, wie die Zehen und der Kopf aus dem Wasser der Badewanne.

Für den semantischen Ablenker gilt:

1. er soll zum gleichen semantischen Feld wie die Bestantwort gehören oder
2. er soll eine mögliche Assoziation dazu sein
3. er soll nicht gleichzeitig ein phonologischer Ablenker sein

Für den dritten Ablenker gilt:

1. er soll optische und inhaltliche Ausgeglichenheit auf der Tafel schaffen
2. er kann auch ein weiterer semantischer oder phonologischer Ablenker sein

3.3 Darstellung des Tests und der Items

Für die Darstellung des Tests übernahm ich die Idee von Rupp (2008) und stellte die Items mit der Software Microsoft PowerPoint® dar. Dadurch konnte der Test den Kindern ganz einfach mittels Notebook präsentiert werden. Im Gegensatz zu ihr entschied ich mich aber für Fotos und nicht für Zeichnungen bei den Abbildungen. Fotos stellen die Begriffe in einer alltagsnahen, natürlichen Art und Weise dar. Das Zielitem und seine drei Ablenker bildeten eine Vierfeldertafel (siehe Abb. 4). Ich achtete darauf, dass die Items möglichst prototypisch

und einfach vor einem neutralen Hintergrund dargestellt waren. Falls das Item Teil eines Ganzen war, fügte ich einen Pfeil ein, der auf den beabsichtigten Bildausschnitt zeigte.

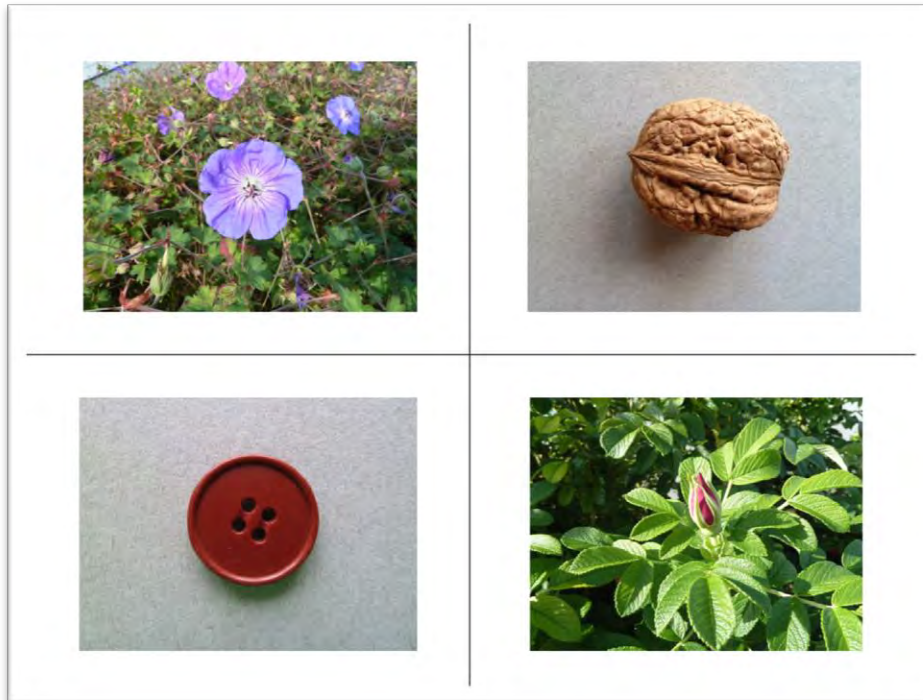


Abb. 4: Exemplarische Darstellung des Items *Chnospe*. Der Knopf ist der phonologische, die offene Blüte der semantische Ablenker. Die Baumnuss ist ein Ablenker, der die Vierfeldertafel ausgleicht.

Der Test startet mit ein paar eher leichten Items und endet mit einer Tafel, die nicht mehr zum Test gehört, aber vom Kind sicher gelöst werden kann. Auf diese Weise kann das Kind in jedem Fall den Test mit einem Erfolgserlebnis beenden. Bei der Abfolge wechseln sich leichtere Items mit schwierigeren ab. Die fünf Adjektive werden als Block präsentiert, so dass sich das Kind besser auf die leicht abgeänderte Aufgabe einstellen kann. Die zehn Verben hingegen werden zufällig eingestreut.

Die Position der Bestantwort innerhalb der Vierfeldertafel bestimmte ein Zufallsgenerator. Die Positionen der Distraktoren gestaltete ich so, dass die Vierfeldertafel optisch und inhaltlich ausgeglichen wirkt.

3.4 Auswahl der Dialektregionen für die Testkonstruktion

Der Test soll für eine möglichst grosse Region konstruiert und später normiert werden, in welcher aber die dialektalen Unterschiede nicht allzu gross sind. Das heisst, es soll eine eher bevölkerungsreiche Region sein ohne grössere geografische Hindernisse und mit verbind-

den Verkehrswegen. Der Entscheid fiel auf die Zürichseeregion im Kanton Zürich und die angrenzende Oberseeregion und das Linthgebiet im Kanton St. Gallen.

Die für die Konstruktionsstichprobe ausgesuchten Gemeinden sind die Stadt Zürich (ZH), die Gemeinde Erlenbach (ZH) am rechten Zürichseeufer, die Stadt Rapperswil-Jona (SG) am Oberen Zürichsee und die zwei Gemeinden Uznach (SG) und Benken (SG) in der Linthebene. Alle diese Gemeinden befinden sich am Gewässersystem Linth-Zürichsee, welches in früheren Zeiten, und zu einem kleinen Teil auch heute noch, als Wassertransportwegsystem genutzt wurde. Heute bilden diese Regionen einen Raum, in dem durch Pendlerströme, durch kulturellen Austausch und vermutlich auch durch Wohnortswechsel eine Durchmischung entsteht und sprachliche Unterschiede immer kleiner werden.



Abb. 5: Geografische Verteilung der Kindergärten für die Konstruktionsstichprobe. Die Gemeinden sind auf der Karte mit einem roten Punkt markiert. Von West nach Ost sind dies: Zürich, Erlenbach, Rapperswil-Jona, Uznach und Benken. (Bildquelle: Wikimedia Commons, Designer Tschubby, leicht verändert)

3.5 Die Rolle des Dialekts bei der Itemauswahl und der Itempräsentation

Bezüglich der für den Test ausgesuchten Items, sollten keine allzu grossen Unterschiede zwischen den Dialekten der Stichprobengemeinden bestehen, da keine speziell typischen Dialektwörter ausgesucht wurden. Sind mehrere Variationen eines Begriffes denkbar, so wurde die vermutlich gebräuchlichere oder die der Standardsprache nähere verwendet. Für das hochdeutsche Wort *bügeln* wurde z.B. nicht das Dialektwort *glette* eingesetzt, sondern das Dialektwort *büggle* und bei *Pinzette* wurde auf die in der Schweiz vielerorts gebräuchliche französische Aussprache verzichtet.

Die Wörter wurden den Kindern aus praktischen Gründen im Dialekt der Gemeinde Benken präsentiert, welcher aber auch durch die Dialekte von Rapperswil-Jona und Zürich beeinflusst wurde.

Dialekt-Wörterbücher konnten nur sehr beschränkt für die Itemauswahl und die Itempräsentation eingesetzt werden, denn es gibt sie nur für wenige Regionen, sie sind unvollständig und sie sind schnell veraltet. Momentan ist ein aktuelles Zürichdeutsches Wörterbuch (Gallmann 2015) erhältlich, aber keines für die ebenfalls ausgesuchten Regionen des Kantons St. Gallen.

3.6 Auswahl der Kindergärten für die Konstruktionsstichprobe

Für die Erprobung dieser ersten Testversion wurde eine weitgehend repräsentative Auswahl von Kindergärten im Untersuchungsgebiet getroffen. Es wurde auf ein ausgeglichenes Verhältnis bezüglich Gemeindegrösse, Ländlichkeit, ökonomischen Kennzahlen, Ausländeranteil und Bildungsniveau der Bevölkerung geachtet. Die Auswahl geschah aufgrund von Daten aus den Statistikdatenbanken der Kantone St. Gallen und Zürich (<http://stada2.sg.ch> und www.statistik.zh.ch [07.09.2016]).

Insgesamt wurden 15 Kindergärten besucht, die sich auf 5 Gemeinden verteilten. Die Gemeinden bestanden aus einer Grossstadt, einer mittelgrossen Stadt und aus drei kleineren Gemeinden. Bei den Städten wurde darauf geachtet, dass Kindergärten aus bevölkerungsmässig gut durchmischten Quartieren berücksichtigt wurden oder darauf, dass gleichzeitig Quartiere der gehobenen und der unteren Mittelklasse ausgewählt wurden. In der kleinsten Gemeinde konnten alle Kindergärten in die Studie einbezogen werden, d.h. alle dort wohnenden Bevölkerungsklassen waren vertreten.

Tab. 2: Statistische Kennzahlen der Gemeinden der Konstruktionsstichprobe.

Gemeinden für die Konstruktionsstichprobe	Anzahl Einwohner	Landwirtschaftliche Nutzfläche [%]	Ausländeranteil [%]	Steuerkraft pro Kopf [Fr.]	höchster Abschluss Tertiärstufe [%]	Schüler Tertiärstufe [Pers. / 1000 Einw.]
Benken	2900	72	11	1625	21	-
Uznach	6300	45	27	1875	22	-
Rapperswil-Jona	26'875	35	18	3100	32	-
Erlenbach	5470	21	21	10630	-	29
Zürich	396'027	10	32	5080	-	21

3.7 Durchführung der Erprobung

Die Testitems wurden den Kindern jeweils in einem separaten, ruhigen Raum mittels Notebook präsentiert. Die Aufgabe wurde jedem Kind kurz in zwei bis drei Sätzen erklärt und es wurde darauf hingewiesen, dass es auch nach einer Wiederholungen fragen darf, wenn es ein Wort nicht richtig verstanden hat. Das Item wurde in normaler Lautstärke und mit normaler Betonung vorgegeben. Die ersten Items wurden mit der Aufforderung „Zeig mir ...“ eingeleitet. Sobald das Kind die Aufgabenstellung sicher verstanden hatte, wurde nur noch das Item genannt. Bei den Adjektiven lautete die Frage „Was isch ...“, bei den Verben wurde nicht darauf eingegangen, dass nun nach einer Tätigkeit und nicht nach einem Objekt gefragt wurde.

Die Testung der 57 Items benötigte ca. 7-10 min pro Kind. Sehr schnelle Kinder schafften es auch in 5 min. Ein Kind benötigte 20 min, weil es bei jeder Tafel Assoziationen zu den Bildern erzählte. Für die Kinder des 2. Kindergartens und auch für die meisten Kinder des 1. Kindergartens war die hohe Anzahl der Items kein Problem. Sie zeigten keine Ermüdungserscheinungen. Dies mag aber auch daran gelegen haben, dass die Items eher zu leicht für sie waren.

Die Erprobung wurde innerhalb von 7 Wochen im November und Dezember 2016 von mir durchgeführt. Die Kinder befanden sich zu diesem Zeitpunkt bereits seit ca. 3 Monaten in der ersten oder zweiten Kindergartenstufe und hatten bereits einige Wochen Zeit um sich in ihrer Klassenzusammensetzung einzuleben

3.8 Grösse der Konstruktions-Stichprobe

Bühner (2005) empfiehlt für die Überprüfung eines Testentwurfs ein $N \geq 100$. Erst ab einer solchen Stichprobengrösse könne die Reliabilität zuverlässig geschätzt werden. In dieser Testentwicklung wurde ein $N=140$ erreicht. Die Stichprobengrösse ist also ausreichend für eine Itemanalyse und eine zuverlässige Abschätzung der Reliabilität.

3.9 Auswertung der Daten

Die Eigenschaften der Stichprobenpopulation für diese Testkonstruktion wurden mit beschreibender Statistik ausgewertet. Bei den Items wurden die Itemschwierigkeit und die Trennschärfe berechnet. Auf eine Faktorenanalyse wurde verzichtet, weil keine weitere Itemreduktion nötig war. Bei den Distraktoren wurden Häufigkeitstabellen erstellt und die Distraktorenbelegung bewertet. Da in dieser Erprobung die Distraktoren nicht sehr häufig gewählt wurden, wurde auf eine Analyse der Schwierigkeiten und Trennschärfen der Distraktoren verzichtet. Die Berechnungen wurden mit IBM SPSS Statistics® Version 24 durchgeführt. Für die Interpretation des SPSS-Outputs verwendete ich Bühner (2005).

4. Ergebnisse

4.1 Zusammensetzung der Konstruktionsstichprobe

Der Wortschatztest wurde zu einem grösseren Teil an Kindern des 2. Kindergartenjahres (108) und zu einem kleineren Teil an Kindern des 1. Kindergartenjahres (32) erprobt. Insgesamt wurden 140 Kinder getestet. Das Alter der Kinder lag zwischen 4;3 und 7;0 Jahren mit einem Durchschnitt von 5;6 Jahren (siehe Abb. 6). Das Geschlechterverhältnis war fast ausgeglichen (Tab. 3). In die Itemanalyse wurden alle befragten Kinder miteinbezogen, d.h. auch Kinder, die Logotherapie erhielten oder Kinder die mehrsprachig waren (Tab. 4). Mehrsprachige Kinder nahmen aber nur an der Befragung teil, wenn gemäss Kindergartenlehrperson ihre starke Sprache Schweizerdeutsch war. Bei einer strikten Auslegung von Mehrsprachigkeit wären alle Kinder, die Schweizerdeutsch und Standardsprache sprechen mehrsprachig. Auf diese strikte Auslegung wurde aber verzichtet. Die Tab. 5 bis 7 enthalten weitere Angaben über die Zusammensetzung der Konstruktionsstichprobe. Ich beurteile die Stichprobe insgesamt als repräsentativ. Für eine zweite Testversion ist es angezeigt, grössere N von Kindern im ersten Kindergartenjahr, Kindern aus Grosstädten und Kindern aus finanzstarken Gemeinden zu erreichen.

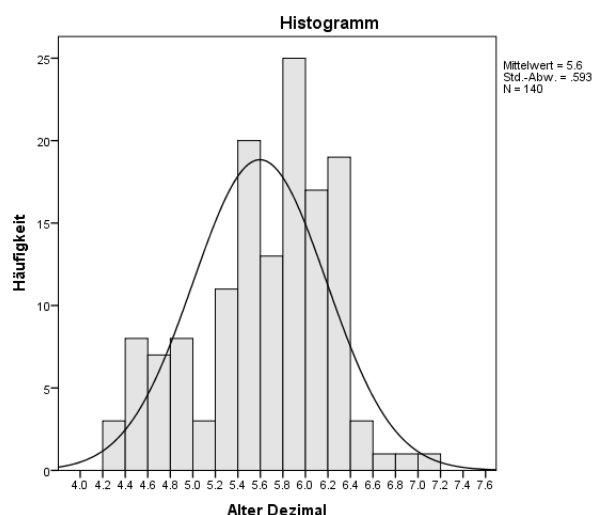


Abb. 6: Altersverteilung der Kinder der Konstruktionsstichprobe. Die zweigipflige Verteilung der Stichprobe ist auf die ungleiche Anzahl Kinder aus dem ersten und dem zweiten Kindergartenjahr zurückzuführen.

Tab. 3: Geschlechterverteilung in der Konstruktionsstichprobe.

	Geschlecht		Gesamt
	männlich	weiblich	
1. Kindergarten	15	17	32
2. Kindergarten	63	45	108
Gesamt	78	62	140

Tab. 4: Verteilung der Sprachkompetenzen in der Konstruktionsstichprobe. Mehrsprachige Kinder nahmen nur Teil, wenn gemäss Kindergartenlehrperson ihre starke Sprache Dialekt war. Unter Hochdeutsch fallen Kinder, die zu Hause mit mindestens einem Elternteil Hochdeutsch sprechen. Die 6 Kinder unter Logo erhielten Therapie v.a. wegen Aussprachestörungen.

	Sprache				Gesamt
	Dialekt	Hochdeutsch	Mehrsprachigkeit	Logo	
1. Kindergarten	25	1	6	0	32
2. Kindergarten	83	6	13	6	108
Gesamt	108	7	19	6	140

Tab. 5: Anzahl der Kinder nach Einwohnerzahl in ihrer Wohngemeinde.

	Gemeindegrösse			Gesamt
	>1000 Einwohner	> 10 000 Einwohner	> 100 000 Einwohner	
1. Kindergarten	6	21	5	32
2. Kindergarten	55	39	14	108
Gesamt	61	60	19	140

Tab. 6: Anzahl der Kinder nach Ländlichkeit ihrer Wohngemeinde. Die Ländlichkeit des Wohnortes der Kinder wird mittels landwirtschaftlicher Nutzfläche (LNF) bestimmt. Sie ist gross, wenn LNF>66 %, mittel bei LNF zwischen 33 und 66 % und klein bei LNF<33%.

	Ländlichkeit			Gesamt
	klein	mittel	gross	
1. Kindergarten	10	21	1	32
2. Kindergarten	22	58	28	108
Gesamt	32	79	29	140

Tab. 7: Anzahl der Kinder nach Finanzkraft ihrer Wohngemeinde. Die Finanzkraft wird durch die Steuerkraft pro Kopf (SPK) in der Wohnortgemeinde bestimmt. Sie ist gross bei SPK > 10 000 Fr., mittel bei SPK zwischen 3 000 Fr. und 10 000 Fr. und klein bei SPK < 3 000 Fr..

	Finanzkraft			Gesamt
	klein	mittel	gross	
1. Kindergarten	1	26	5	32
2. Kindergarten	47	53	8	108
Gesamt	48	79	13	140

4.2 Itemanalyse

4.2.1 Itemschwierigkeit

Die Analyse der Itemschwierigkeit ist unten in Tab. 8 aufgeführt. Die Werte in der vierten Spalte stellen die Itemschwierigkeiten nach der Ratekorrektur dar. Diese Korrektur wird nötig, weil bei einer Auswahl von vier Antwortmöglichkeiten ein Proband mit einer Wahrscheinlichkeit von 25 % zufällig die richtige Antwort nennt.

Je grösser der Wert der Itemschwierigkeit wird, desto einfacher ist das Item zu lösen. Eine Itemschwierigkeit von 0.80 bedeutet, dass 80 % der Probanden das Item richtig lösten. Items mit mittleren Schwierigkeiten differenzieren am besten zwischen Lösern und Nichtlösern von Items. Ein Test benötigt aber auch tiefe und hohe Schwierigkeiten um innerhalb der starken Lösern und innerhalb der schwachen Lösern noch Unterscheidungen treffen zu können. Extrem hohe und extrem tiefe Schwierigkeiten, die nur noch von sehr wenigen Probanden oder von allen gelöst werden, sind hingegen zu wenig informativ für einen Test. Laut Bortz und Döring (2006) liegen ideale Itemschwierigkeiten im mittleren Bereich, zwischen 0.20 und 0.80.

Aus Tab. 8 und Abb. 7 ist ersichtlich, dass der grösste Teil der Items hohe Schwierigkeitswerte aufweisen, also leicht zu lösen sind. Der Mittelwert aller Itemschwierigkeiten liegt gemäss Abb. 7 bei 0.81.

Drei Items wurden von der Berechnung automatisch ausgeschlossen, weil diese von allen 140 Probanden als richtig erkannt wurden. Es sind dies die Items *Chrono*, *Fuchs* und *truurig*. Diese Items eignen sich gut als Beispiel-Items oder als Eisbrecher für die Endversion des Tests.

Tab. 8: Itemschwierigkeiten. Die Beurteilung der Itemschwierigkeiten geschah aufgrund der Richtwerte von Bortz und Döring (2006). Die drei Items 1, 8 und 15 wurden bereits ausgeschlossen, weil sie von allen Probanden gelöst wurden.

	Standard- abweichung	Itemschwierigkeit	Itemschwierigkeit nach Ratekorrektur	N	Bewertung
Item_19	.401	.20	-0,07	140	hohe Schwierigkeit
Item_30	.476	.34	0,12	140	
Item_9	.494	.59	0,45	140	mittlere Schwierigkeit
Item_16	.487	.62	0,50	140	
Item_17	.450	.72	0,63	140	
Item_31	.443	.74	0,65	140	
Item_57	.439	.74	0,66	140	
Item_41	.435	.75	0,67	140	
Item_6	.401	.80	0,73	140	
Item_53	.401	.80	0,73	140	
Item_3	.396	.81	0,74	140	
Item_21	.396	.81	0,74	140	
Item_33	.396	.81	0,74	140	
Item_20	.384	.82	0,76	140	
Item_27	.384	.82	0,76	140	
Item_5	.378	.83	0,77	140	
Item_18	.372	.84	0,78	140	
Item_36	.372	.84	0,78	140	
Item_29	.358	.85	0,80	140	
Item_50	.358	.85	0,80	140	
Item_51	.358	.85	0,80	140	
Item_13	.344	.86	0,82	140	niedrige Schwierigkeit
Item_45	.344	.86	0,82	140	
Item_28	.336	.87	0,83	140	
Item_7	.328	.88	0,84	140	
Item_40	.328	.88	0,84	140	
Item_11	.319	.89	0,85	140	
Item_52	.310	.89	0,86	140	
Item_32	.310	.89	0,86	140	
Item_24	.291	.91	0,88	140	
Item_37	.281	.91	0,88	140	
Item_39	.291	.91	0,89	140	
Item_55	.281	.91	0,89	140	
Item_2	.270	.92	0,90	140	

Item_12	.270	.92	0,90	140	niedrige Schwierigkeit
Item_22	.270	.92	0,90	140	
Item_25	.258	.93	0,90	140	
Item_38	.258	.93	0,90	140	
Item_44	.233	.94	0,92	140	
Item_34	.219	.95	0,93	140	
Item_14	.203	.96	0,94	140	
Item_46	.203	.96	0,94	140	
Item_56	.186	.96	0,95	140	
Item_10	.167	.97	0,96	140	
Item_26	.167	.97	0,96	140	
Item_23	.145	.98	0,97	140	
Item_43	.145	.98	0,97	140	
Item_49	.145	.98	0,97	140	
Item_54	.145	.98	0,97	140	
Item_4	.119	.99	0,98	140	
Item_35	.119	.99	0,98	140	
Item_47	.119	.99	0,98	140	
Item_42	.085	.99	0,99	140	
Item_48	.085	.99	0,99	140	

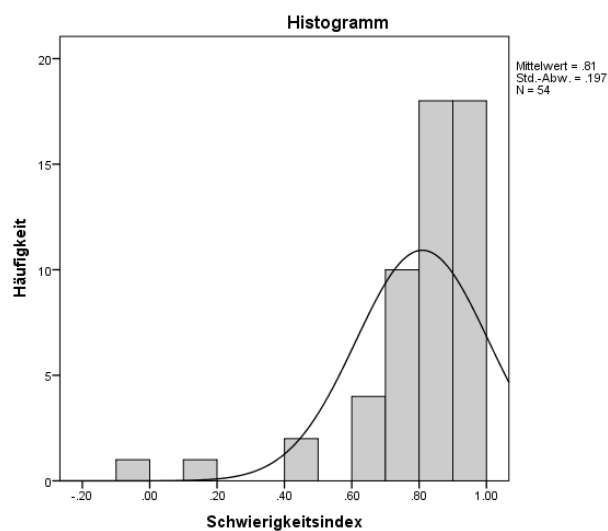


Abb. 7: Häufigkeiten der Schwierigkeitsindizes nach Ratekorrektur.
Die Itemschwierigkeiten liegen weit rechts, d.h. es liegt eine Häufung von sehr leichten Items vor.

4.2.2 Trennschärfe

Die Trennschärfe ist eine Korrelation, d.h. sie kann Werte von -1 bis +1 annehmen. Je höher die Trennschärfe ist, desto besser stimmt das einzelne Item mit der Leistung des ganzen Tests (Skala) überein. Items mit einer mittleren Trennschärfe sollten in einem Test häufig vertreten sein. Hohe Trennschärfen wären theoretisch noch besser geeignet, es müssen aber Kompromisse eingegangen werden, da die Trennschärfe und die Itemschwierigkeit voneinander abhängen. Sehr einfache und sehr schwierige Items generieren automatisch Einbusen bei der Trennschärfe. In einem Test werden aber auch sehr einfache und sehr schwierige Items benötigt, deshalb können nicht nur sehr hohen Trennschärfen ausgewählt werden. Items mit niedrigen oder gar negativen Trennschärfen sollten aus einem Test ausgeschlossen werden (Bortz & Döring 2006). Eine negative Trennschärfe bedeutet, dass die Probanden, die dieses Item richtig gelöst haben, die restlichen Items des Tests eher falsch lösten. Richtwerte für die Einteilung in niedrige, mittlere und hohe Trennschärfen findet sich z.B. in Bühner (2005) und in Bortz und Döring (2006). Dort werden Werte < 0.30 als niedrig, Werte von 0.30 bis 0.50 als mittel und Werte > 0.50 als hoch beurteilt. Laut Tab. 9 liegen die Trennschärfen von 21 Items dieses Tests in einem niedrigen Bereich, 26 in einem mittleren und 7 in einem hohen Bereich. Der Mittelwert aller Trennschärfen liegt gemäss Abb. 8 bei 0.31 .

In der fünften Spalte von Tab. 9 ist angegeben, wie Cronbachs Alpha sich verändern würde, falls das entsprechende Item aus dem Test ausgeschlossen wird. Das errechnete Alpha des ganzen Tests beträgt 0.856 . Übersteigt der Wert in der fünften Spalte 0.856 , kann ein Ausschluss des entsprechenden Items in Frage kommen.

Tab. 9: Trennschärfen. Die Beurteilung der Trennschärfen geschah aufgrund der Richtwerte von Bortz und Döring (2006). Die drei Items 1, 8 und 15 wurden bereits ausgeschlossen, weil sie von allen Probanden gelöst wurden. (Cronbachs Alpha = 0.856)

	Skalenmittelwert, wenn Item weg- gelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item- Skala-Korrelation = Trennschärfe	Cronbachs Alpha, wenn Item wegge- lassen	Beurteilung
Item_20	45.46	34.394	-.011	.861	niedrige Trennschärfe
Item_22	45.36	34.836	-.131	.861	
Item_35	45.29	34.410	.047	.857	
Item_14	45.32	34.335	.048	.857	
Item_25	45.35	34.200	.074	.857	
Item_11	45.39	33.938	.121	.857	
Item_30	45.94	33.571	.125	.859	
Item_41	45.53	33.647	.130	.858	
Item_24	45.37	33.861	.161	.856	
Item_19	46.08	33.569	.163	.857	
Item_44	45.34	33.922	.189	.856	
Item_48	45.29	34.292	.193	.856	
Item_6	45.48	33.374	.206	.856	
Item_16	45.66	33.061	.213	.857	
Item_54	45.30	34.096	.220	.856	
Item_50	45.43	33.441	.222	.856	
Item_51	45.43	33.427	.226	.856	
Item_9	45.69	32.905	.236	.857	
Item_26	45.31	33.941	.268	.855	
Item_13	45.41	33.295	.272	.855	
Item_53	45.48	33.000	.288	.854	
Item_46	45.32	33.730	.304	.854	mittlere Trennschärfe
Item_52	45.39	33.303	.304	.854	
Item_45	45.41	33.136	.312	.854	
Item_55	45.36	33.399	.312	.854	
Item_3	45.47	32.913	.313	.854	
Item_21	45.47	32.898	.316	.854	
Item_7	45.40	33.177	.319	.854	
Item_10	45.31	33.840	.320	.855	
Item_23	45.30	33.909	.331	.855	
Item_56	45.31	33.713	.343	.854	
Item_43	45.30	33.866	.356	.854	
Item_2	45.36	33.267	.369	.853	

Item_4	45.29	33.964	.369	.855	mittlere Trennschärfe
Item_31	45.54	32.423	.371	.853	
Item_39	45.37	33.127	.381	.853	
Item_57	45.54	32.394	.381	.852	
Item_5	45.45	32.638	.396	.852	
Item_42	45.29	34.090	.398	.855	
Item_12	45.36	33.166	.402	.853	
Item_47	45.29	33.892	.421	.854	
Item_32	45.39	32.857	.432	.852	
Item_28	45.41	32.704	.436	.851	
Item_29	45.43	32.549	.443	.851	
Item_38	45.35	33.078	.453	.852	
Item_37	45.36	32.910	.466	.851	
Item_49	45.30	33.679	.468	.854	
Item_34	45.33	33.129	.522	.852	hohe Trennschärfe
Item_18	45.44	32.133	.526	.849	
Item_36	45.44	32.133	.526	.849	
Item_27	45.46	32.049	.527	.849	
Item_17	45.56	31.558	.540	.848	
Item_40	45.40	32.213	.583	.849	
Item_33	45.47	31.632	.606	.847	

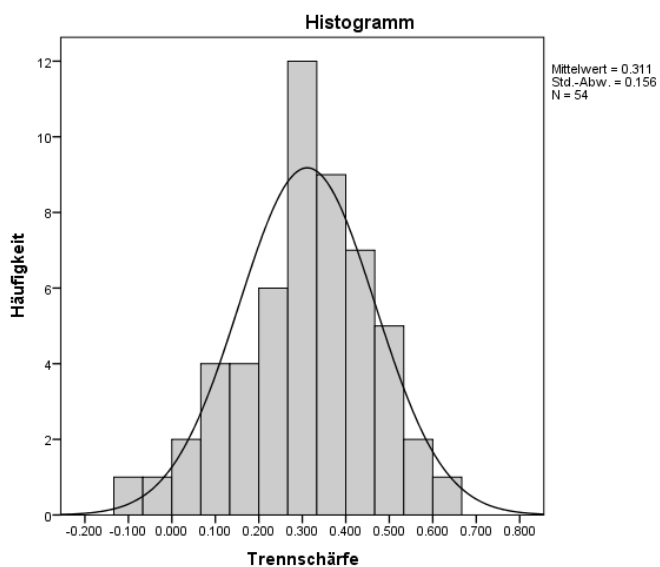


Abb. 8: Trennschärfen. Die Verteilung der Trennschärfen liegt nahe an einer Normalverteilung.

4.3 Itemausschluss

Bei einer nächsten, verbesserten Version des Tests können folgende 28 Items aufgrund zu tiefer Trennschärfen (<0.300) oder aufgrund eines Schwierigkeitsgrades, welcher im Item-Set übervertreten ist, ausgeschlossen werden. Item 9 *Gländer* wird trotz niedriger Trennschärfe beibehalten, da es eine hohe Schwierigkeit aufweist. Die Adjektive werden alle ausgeschlossen, weil ihre Darstellung nicht eindeutig war (siehe 5.4).

Item 1 Chronä	Item 16 Pyramide	Item 38 Rolltrappe
Item 5 Ellebogä	Item 19 Chnospe	Item 41 Drähtür
Item 6 Hammer	Item 20 abtröchne	Item 44 Chräbs
Item 7 Bagger	Item 22 Stecker	Item 48 chrüüche
Item 8 Fuchs	Item 24 schniide	Item 50 Heu
Item 11 troche	Item 25 Helikopter	Item 51 tanke
Item 12 suur	Item 26 rissä	Item 53 bügle
Item 13 alt	Item 30 Klinke	Item 54 Helm
Item 14 weich	Item 32 Pinzette	
Item 15 truurig	Item 35 stäche	

Durch den Ausschluss dieser Items steigt der Mittelwert der Trennschärfen von 0.31 auf 0.41, der Mittelwert der Itemschwierigkeiten steigt leicht von 0.81 auf 0.84 und Cronbachs Alpha steigt minim von 0.856 auf 0.858. Es verbleiben noch 29 Items im Test, davon sind nur noch 2 Items Verben.

4.4 Distraktoren-Analyse

Unten in Tab. 10 sind die Belegungen der Distraktoren dargestellt. Diese Belegungen wurden mit den *Tabellen zur Beurteilung der Verteilung von Distraktoren bei Merfach-Wahl-Aufgaben* aus Lienert und Raatz (1998) verglichen. Es zeigte sich dabei, dass gemäss diesen Tabellen bei 25 der 57 Items die Distraktorenbelegung nicht optimal war. Diese Tabellen gelten aber nur für untereinander, gleich attraktive Distraktoren, was aber bei diesem Test nicht der Fall ist, da der phonologische und der semantische Ablenker anziehender sind als der dritte Ablenker. Trotzdem wurde versucht anhand dieser Tabellen und subjektiven Eindrücken aus der Feldarbeit eine Einschätzung vorzunehmen. Diese Einschätzung folgt nach Tab. 10. Distraktoren, die gemäss den Tabellen in Lienert und Raatz (1998) zu häufig gewählt wurden, sind in Tab. 10 fett hervorgehoben. Zu beachten ist, dass unter den zusätzlichen Distraktoren weitere phonologische oder semantische Ablenker vorkommen können.

Tab. 10: Verteilung der Falschantworten auf die drei Distraktoren (Ablenker) . Insgesamt nahmen 140 Kinder an dieser Erprobung teil, aber durch Nullantworten reduzierte sich diese Zahl bei einigen Items. (best: richtige Antwort, phon: phonologischer Ablenker, sem: semantischer Ablenker, zus: zusätzlicher Ablenker)

Item 1 Chronä	best	phon	sem	zus	Gesamt
		Bohne	Huet	Rüebli	
Häufigkeit	140	0	0	0	140
Prozent	100,0	0,0	0,0	0,0	100

Item 2 Turm	best	phon	sem	zus	Gesamt
		Wurm	Sturm	Muulwurf	
Häufigkeit	129	6	0	4	139
Prozent	92,8	4,3	0,0	2,9	100

Item 3 Iglu	best	phon	sem	zus	Gesamt
		Igel	Zelt	Wiesel	
Häufigkeit	113	17	0	10	140
Prozent	80,7	12,1	0,0	7,1	100

Item 4 Schatz	best	phon	sem	zus	Gesamt
		Spatz	Pirat	Chatz	
Häufigkeit	138	0	1	1	140
Prozent	98,6	0,0	0,7	0,7	100

Item 5 Ellebogä	best	phon	sem	zus	Gesamt
		Rägebogä	Fuess	Pfilbogä	
Häufigkeit	116	2	1	20	139
Prozent	83,5	1,4	0,7	14,4	100

Item 6 Hammer	best	phon	sem	zus	Gesamt
		Hummer	Biäli	Fisch	
Häufigkeit	112	2	26	0	140
Prozent	80,0	1,4	18,6	0,0	100

Item 7 Bagger	best	phon	sem	zus	Gesamt
		Bagge	Chrane	Wade	
Häufigkeit	123	11	6	0	140
Prozent	87,9	7,9	4,3	0,0	100

Item 8 Fuchs	best	phon	sem	zus	Gesamt
		Büchs	roti Chatz	Fuess	
Häufigkeit	140	0	0	0	140
Prozent	100,0	0,0	0,0	0,0	100

Item 9 Gländer	best	phon	sem	zus	Gesamt
		Ständer	Haag	Drahtzuun	
Häufigkeit	82	12	21	16	131
Prozent	62,6	9,2	16,0	12,2	100

Item 10 Bank	best	phon	sem	zus	Gesamt
		Band	Stuehl	Gigampfi	
Häufigkeit	136	1	1	2	140
Prozent	97,1	0,7	0,7	1,4	100

Item 11 troche	best	phon	sem	zus	Gesamt
		trinke	nasse Hund	Socke	
Häufigkeit	54	2	11	70	137
Prozent	39,4	1,5	8,0	51,1	100

Item 12 suur	best	phon	sem	zus	Gesamt
		Muur	Schoggi	Uhr	
Häufigkeit	129	3	1	6	139
Prozent	92,8	2,2	0,7	4,3	100

Item 13 alt	best	phon	sem	zus	Gesamt
		chalt	nüüs Huus	Landschaft	
Häufigkeit	121	10	1	5	137
Prozent	88,3	7,3	0,7	3,6	100

Item 14 weich	best	phon	sem	zus	Gesamt
		Teich	Chopfstei	Holz	
Häufigkeit	134	6	0	0	140
Prozent	95,7	4,3	0,0	0,0	100

Item 15 truurig	best	phon	sem	zus	Gesamt
	Trube			Mandarinli	
Häufigkeit	140	0	0	0	140
Prozent	100,0	0,0	0,0	0,0	100

Item 16 Pyramide	best	phon	sem	zus	Gesamt
	Margarine	Tutench	Schoggi		
Häufigkeit	87	5	40	0	132
Prozent	65,9	3,8	30,3	0,0	100

Item 17 Gschirr	best	phon	sem	zus	Gesamt
	Gschmier	Töpf	Gschänk		
Häufigkeit	101	8	24	3	136
Prozent	74,3	5,9	17,6	2,2	100

Item 18 Ferse	best	phon	sem	zus	Gesamt
	Ferkel, Ferli	Elleboge	Fernseh		
Häufigkeit	117	2	10	9	138
Prozent	84,8	1,4	7,2	6,5	100

Item 19 Chnospe	best	phon	sem	zus	Gesamt
	Chnopf	offeni Blueme	Baumnuss		
Häufigkeit	28	21	2	79	130
Prozent	21,5	16,2	1,5	60,8	100

Item 20 abtröchne	best	phon	sem	zus	Gesamt
	abbreche	abwäsche	wüsche		
Häufigkeit	115	3	22	0	140
Prozent	82,1	2,1	15,7	0,0	100

Item 21 Schruube	best	phon	sem	zus	Gesamt
	Trube	Nagel	Duube		
Häufigkeit	113	4	19	2	138
Prozent	81,9	2,9	13,8	1,4	100

Item 22 Stecker	best	phon	sem	zus	Gesamt
	Wecker	Kabel	Bagger		
Häufigkeit	129	0	10	0	139
Prozent	92,8	0,0	7,2	0,0	100

Item 23 Kabel	best	phon	sem	zus	Gesamt
	Schnabel	Seil	Rüssel		
Häufigkeit	137	0	2	0	139
Prozent	98,6	0,0	1,4	0,0	100

Item 24 schiide	best	phon	sem	zus	Gesamt
	schniie, schnäie	raffe	schiibe		
Häufigkeit	127	0	13	0	140
Prozent	90,7	0,0	9,3	0,0	100

Item 25 Helikopter	best	phon	sem	zus	Gesamt
	Ägypter	Libelle	Flugzüg		
Häufigkeit	130	0	0	10	140
Prozent	92,9	0,0	0,0	7,1	100

Item 26 riissä	best	phon	sem	zus	Gesamt
	biissä	falte	trinke		
Häufigkeit	136	2	2	0	140
Prozent	97,1	1,4	1,4	0,0	100

Item 27 Gurt	best	phon	sem	zus	Gesamt
	Joghurt	Hoseträger	Gurke		
Häufigkeit	115	6	9	8	138
Prozent	83,3	4,3	6,5	5,8	100

Item 28 Chünni	best	phon	sem	zus	Gesamt
	Chämmi	Stirn	Füür		
Häufigkeit	122	9	3	3	137
Prozent	89,1	6,6	2,2	2,2	100

Item 29 Tanne	best	phon	sem	zus	Gesamt
	Pfanne	Laubbaum	Channe, Chante		
Häufigkeit	119	7	6	8	140
Prozent	85,0	5,0	4,3	5,7	100

Item 30 Klinke	best	phon	sem	zus	Gesamt
	Blinker	Knauf	Pedale		
Häufigkeit	48	28	20	31	127
Prozent	37,8	22,0	15,7	24,4	100

Item 31 Chranz	best	phon	sem	zus	Gesamt
	Chratz, Rettigsring		Zopf		
Häufigkeit	103	30	6	1	140
Prozent	73,6	21,4	4,3	0,7	100

Item 32 Pinzette	best	phon	sem	zus	Gesamt
	Prinzessin		Zange	Zettel	
Häufigkeit	125	8	5	1	139
Prozent	89,9	5,8	3,6	0,7	100

Item 33 Gitter	best	phon	sem	zus	Gesamt
	Gwitter		Netz	Zuun	
Häufigkeit	113	10	12	4	139
Prozent	81,3	7,2	8,6	2,9	100

Item 34 Nagel	best	phon	sem	zus	Gesamt
	Nabel		Hammer	Hagel	
Häufigkeit	133	3	2	0	138
Prozent	96,4	2,2	1,4	0,0	100

Item 35 stäche	best	phon	sem	zus	Gesamt
	bräche		schniide	streichle	
Häufigkeit	138	1	1	0	140
Prozent	98,6	0,7	0,7	0,0	100

Item 36 Sattel	best	phon	sem	zus	Gesamt
	Säbel		Gepäck-träger	Blockföte	
Häufigkeit	117	11	7	1	136
Prozent	86,0	8,1	5,1	0,7	100

Item 37 Zapfe	best	phon	sem	zus	Gesamt
	Zopf		Baumnuss	Eichhörnli	
Häufigkeit	128	9	3	0	140
Prozent	91,4	6,4	2,1	0,0	100

Item 38 Rolltrappe	best	phon	sem	zus	Gesamt
	Rolltäsche		Lift	Strickleiter	
Häufigkeit	130	3	0	7	140
Prozent	92,9	2,1	0,0	5,0	100

Item 39 raffle	best	phon	sem	zus	Gesamt
	Affe		schniide	Rabe	
Häufigkeit	127	1	3	6	137
Prozent	92,7	0,7	2,2	4,4	100

Item 40 Gleis	best	phon	sem	zus	Gesamt
	Geiss		Zug	Unterführig	
Häufigkeit	123	5	6	5	139
Prozent	88,5	3,6	4,3	3,6	100

Item 41 Drähtür	best	phon	sem	zus	Gesamt
	Dräht		Drähschalter	Huustür	
Häufigkeit	105	7	6	21	139
Prozent	75,5	5,0	4,3	15,1	100

Item 42 Pinsel	best	phon	sem	zus	Gesamt
	Insle		Farbchaste	Spachtel	
Häufigkeit	139	0	1	0	140
Prozent	99,3	0,0	0,7	0,0	100

Item 43 falte	best	phon	sem	zus	Gesamt
	Falter		schniide	Falle	
Häufigkeit	107	32	0	0	139
Prozent	77,0	23,0	0,0	0,0	100

Item 44 Chräbs	best	phon	sem	zus	Gesamt
	Chräbeli		Seestern	Chäfer	
Häufigkeit	132	0	4	3	139
Prozent	95,0	0,0	2,9	2,2	100

Item 45 Fadä	best	phon	sem	zus	Gesamt
		Fahnä	Seil	Flade	
Häufigkeit	121	8	7	2	138
Prozent	87,7	5,8	5,1	1,4	100

Item 46 Muschle	best	phon	sem	zus	Gesamt
		Masche	Schnägg	Nuss	
Häufigkeit	134	4	1	1	140
Prozent	95,7	2,9	0,7	0,7	100

Item 47 Leitere	best	phon	sem	zus	Gesamt
		Litermass	Stäge	Dirigent	
Häufigkeit	138	1	0	0	140
Prozent	98,6	0,7	0,0	0,0	99

Item 48 chrüuche	best	phon	sem	zus	Gesamt
		rüuche	ränne	Chueche	
Häufigkeit	139	1	0	0	140
Prozent	99,3	0,7	0,0	0,0	100

Item 49 Flamingo	best	phon	sem	zus	Gesamt
		Flamme	Storch	Füürlöscher	
Häufigkeit	137	1	1	1	140
Prozent	97,9	0,7	0,7	0,7	100

Item 50 Heu	best	phon	sem	zus	Gesamt
		Leu	grüne Wiese	Pneu	
Häufigkeit	119	17	3	0	139
Prozent	85,6	12,2	2,2	0,0	100

Item 51 tanke	best	phon	sem	zus	Gesamt
		Franke	fahrendes Auto	Trampolin	
Häufigkeit	119	11	0	7	137
Prozent	86,9	8,0	0,0	5,1	100

Item 52 Tram	best	phon	sem	zus	Gesamt
		Rahm	Bus	Joghurt	
Häufigkeit	125	6	4	1	136
Prozent	91,9	4,4	2,9	0,7	100

Item 53 bügle	best	phon	sem	zus	Gesamt
		zügle	lisme	bücke	
Häufigkeit	112	2	23	3	140
Prozent	80,0	1,4	16,4	2,1	100

Item 54 Helm	best	phon	sem	zus	Gesamt
		Halm	Huet	Töff	
Häufigkeit	137	0	3	0	140
Prozent	97,9	0,0	2,1	0,0	100

Item 55 schellä	best	phon	sem	zus	Gesamt
		Chellä	schniide	Schälle	
Häufigkeit	128	3	2	7	140
Prozent	91,4	2,1	1,4	5,0	100

Item 56 Horn	best	phon	sem	zus	Gesamt
		Dorn	Rüssel	Morchle	
Häufigkeit	135	0	3	2	140
Prozent	96,4	0,0	2,1	1,4	100

Item 57 Teich	best	phon	sem	zus	Gesamt
		Teig	Fluss	Zopf	
Häufigkeit	104	21	14	1	140
Prozent	74,3	15,0	10,0	0,7	100

Beim Item *Iglu* wurde der phonologische Ablenker *Igel* moderat überbelegt. Dies erachte ich als akzeptabel. Der semantische Ablenker *Zelt* hingegen, der nie gewählt wurde, sollte in einer nächsten Version ersetzt werden, denn es ist vermutlich ein zu bekanntes Wort für die Kinder.

Beim Item *Elleboge* ist der Ablenker *Pfiilboge* überbelegt. Er war für die Kinder sehr attraktiv, weil er in Form eines Spielzeuges dargestellt war. In einer nächsten Version würde ich ihn beibehalten aber nicht mehr als Spielzeug darstellen. Der Ablenker *Fuess* war eher wieder ein zu bekannter Begriff für die Kinder und wurde kaum gewählt. Ihn könnte man durch ein weniger bekanntes Körperteil ersetzen.

Beim Item *Hammer* war der semantische Ablenker *Biäli (Axt)* deutlich überbelegt. Hier würde ich die Axt ebenfalls beibehalten aber die Ablenker *Hummer* und *Fisch* durch weitere Werkzeuge ersetzen und nur noch Werkzeuge als Ablenker präsentieren. Dabei müsste leider auf einen phonologischen Ablenker verzichtet werden.

Beim Item *troche* wurden *Socke* viel zu oft gewählt. Hier ist die bildliche Darstellung verantwortlich für diese Bevorzugung. Die Socken hingen an einer Wäscheleine und sollten eigentlich nasse Socken darstellen. Für die Kinder waren sie aber bereits trocken oder sie trockneten gerade und wurden deshalb dem Wort *troche* zugeordnet. Die Bestantwort war das Bild eines ausgetrockneten Bodens, was aber für viele Kinder eine zu untypische Darstellung war. Dieses Item würde ich in einer nächsten Version ganz aus dem Test entfernen, wie alle weiteren Adjektive, da sie in dieser Testform kaum eindeutig dargestellt werden können.

Beim Item *alt* wurde der phonologische Ablenker *chalt* zu oft gewählt. Auch hier denke ich, dass die Darstellung nicht eindeutig war. *Chalt* wurde durch ein schneebedecktes Seeufer mit etwas Schilf dargestellt. Das Schilf war in dieser winterlichen Darstellung nicht mehr frisch und grün, sondern *alt* und trocken. So entstand eine zu grosse Konkurrenz zur Bestantwort, die als *altes* Haus dargestellt war.

Beim Item *weich* wurde der phonologische Ablenker *Teich* leicht zu häufig gewählt. Hier liegt vermutlich ein akustisches Problem vor. *Weich* und *Teich* klingen zu ähnlich. Auch möglich wäre, dass das Wasser sich für die Kinder weich anfühlt.

Beim Item *Pyramide* wurde die goldene Totenmaske des *Tutenchamun* deutlich zu häufig gewählt. Vermutlich war dieser glänzende Gegenstand visuell einfach zu attraktiv. Der Ablenker *Schoggi* war auch zu bekannt für die Kinder und kam deshalb für das unbekanntes Wort *Pyramide* nicht in Frage.

Beim Item *Gschirr* wurden die *Kochtöpfe* zu häufig gewählt. Dieser Ablenker ist tatsächlich semantisch sehr nah an der Bestantwort und benötigt in seiner Umgebung ebenfalls sehr nahe Ablenker, falls er nicht überbelegt werden soll. Ich würde hier wieder auf einen phonologischen Ablenker verzichten und nur Küchenutensilien als weitere semantische Ablenker anbieten.

Beim Item *Chnospe (Knospe)* wurde erstaunlicherweise extrem häufig die *Baumnuss* gewählt. Diese war einerseits für einige Kinder das unbekannteste Ding in der Auswahl und passte deshalb gut auf das für sie unbekannte Wort *Chnospe*, andererseits fragten auch Kinder nach, ob ich vielleicht *chnusprig* gemeint habe. Sie bildeten also eine Assoziation zu einer Eigenschaft der Nuss. Bei diesem Item würde ich in einer nächsten Version ein völlig neues Ablenkenset anbieten.

Beim Item *abtröchne* wurde *abwäsche* zu häufig als Ablenker gewählt. Dies vermutlich weil das Wasser auf der Abbildung zu wenig deutlich zu sehen war und weil die anderen beiden Ablenker zu wenig plausibel waren. Hier würde ich eine deutlichere Darstellung wählen. Und die Ablenker *wüsche* und *abbreche* durch unbekanntere ersetzen.

Beim Item *Rolltrappe* wurde die *Strickleiter* zu oft gewählt. Die dargestellte Strickleiter war zum Teil aufgerollt, was vermutlich den Ausschlag gab. In einer nächsten Version würde ich die Strickleiter völlig ausgerollt darstellen.

Beim Item *Drähtür* wurde der Ablenker *Huustür* zu häufig gewählt. Dies vermutlich weil das Wort *Tür* im Zielitem und im Ablenker vorkommt. Die *Haustüre* würde ich in einer nächsten Testversion durch ein Latten- oder Gittertor ersetzen.

Beim Item *tanke* wurde zu oft der phonologische Ablenker *Franke* gewählt. Da es schwierig ist einen anderen phonologischen Ablenker zu finden würde ich ihn beibehalten und versuchen einen attraktiveren semantischen Ablenker einzufügen.

Bei den folgenden Zielitem-Ablenker-Paaren wurde der semantische Ablenker zwar zu oft gewählt, ich würde ihn aber trotzdem beibehalten, weil er semantisch sehr passend ist. Allerdings müssten für die zwei weiteren Ablenker attraktivere, plausiblere Varianten gefunden werden.

Schruube – Nagel, Stecker – Kabel, schniide – raffle, Helikopter – Flugzüg, büggle (bügeln) - lisme (stricken), Helm - Huet

Bei den nächsten Zielitem-Ablenker-Paaren ist der phonologische Ablenker zu häufig belegt. Dies, weil er vermutlich akustisch sehr leicht mit dem Zielitem zu verwechseln ist. Diese Ablenker würde ich in einer nächsten Testversion ersetzen.

Chranz – Chratz, Heu – Leu, Teich - Teig

4.5 Gütekriterien

Objektivität

Da bei dieser Testform mit vorgegebenen Auswahlantworten nur eine Antwort richtig ist, ist der Test bezüglich der Auswertung objektiv. Bezüglich Interpretation wird er objektiv sein, sobald er normiert ist. Es wurden keine Vergleiche mit verschiedenen Untersuchern durchgeführt, trotzdem wird die Durchführung als objektiv betrachtet, denn die Items wurden den Kindern immer gleich präsentiert und die Aufgabe auf die gleiche Art und Weise erklärt.

Reliabilität

Für die Reliabilität in dieser ersten Testversion wurde ein Cronbach Alpha von 0.86 bestimmt. Der Test ist also reliabel.

Validität

Die Validität wurde noch nicht überprüft.

4.6 Normierung

Bei dieser ersten Testversion wurde noch keine Normierung durchgeführt. Zuerst soll in einer zweiten Testversion die Qualität der Items und die Verteilung der Probandenleistung verbessert werden. Die Probandenrohwerte sind nicht normalverteilt (siehe Abb. 9). Für eine Normierung ist eine Normalverteilung aber erforderlich. Eine Normalverteilung war aufgrund der berechneten Itemschwierigkeiten auch nicht zu erwarten.

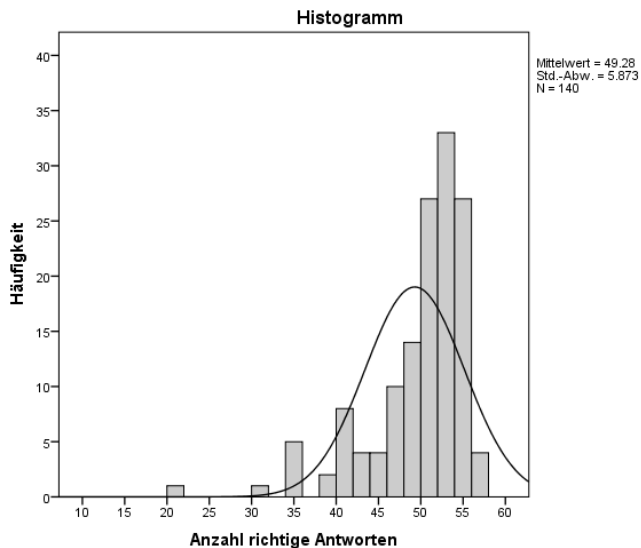


Abb. 9: Verteilung der Probandenrohwerte. Es liegt eindeutig keine Normalverteilung vor. Möglich waren 0 bis 57 richtige Antworten pro Proband.

4.7 Cut-Off

Das Verfahren kann zwar zum jetzigen Zeitpunkt noch nicht als Test eingesetzt werden, für ein Screening sollte meines Erachtens die Datenlage aber ausreichen. Ein Screening benötigt einen Cut-Off, der die auffälligen von den unauffälligen Kindern trennt. Da in der normalen Bevölkerung 5-8 % der Kinder eines Jahrgangs eine spezifische Spracherwerbsstörung aufweisen (Suchodoletz 2003), sollten auch in meiner Stichprobe 5-8 % der Kinder sprachauffällig sein. Die 5-8 % tiefsten Leistungen in diesem Verfahren sollten also unter dem zu setzenden Cut-Off liegen. Da es sich um ein Screening und nicht um einen Test handelt, wähle ich den höheren Wert und setze eine 8 %-Limite. Diese wende ich nur auf die Kinder des 2. Kindergartens und nur auf Kinder ohne Mehrsprachigkeit an. Auf diese Weise erhalte ich einen Cut-Off von 47 richtigen Antworten. Kinder des 2. Kindergartens, die 47 oder weniger der insgesamt 57 Items richtig beantworten, sollten also genauer abgeklärt werden.

Unter diesen 8 % der Kinder befinden sich auch alle Kinder die Logotherapie erhalten. Dies spricht für diesen Cut-Off und dafür, dass diese erste Version durchaus als Screening genutzt werden kann.

4.8 Wortliste

Da es sehr schwierig ist aus der Literatur Informationen über den rezeptiven Wortschatz von Kindergartenkindern zu erhalten, möchte ich hier zumindest einen kleinen Beitrag dazu leisten.

Unten aufgeführt ist eine aus den ursprünglich 57 Items erstellte Wortliste, die >95 % der Kinder im ersten Semester des 2. Kindergartenjahres rezeptiv beherrschten. Die 81 ausgewählten Kinder für diese Auswertung sind zwischen 5;0 und 6;9 Jahre alt, beide Elternteile sprechen Schweizerdeutsch und sie besuchen keine Logotherapie. Diese Liste kann für die Entwicklung zukünftiger rezeptiver Wortschatz-Tests hilfreich sein. Für das erste Kindergartenjahr wurde keine solche Wortliste erstellt, da die Stichprobengrösse zu klein war.

Chronä	rissä (reissen)	Pinsel
Turm	Gitter	Muschle
Schatz	Nagel	Leitere
Fuchs	stäche	chrüüche
(Sitz-)Bank	Zapfe	Flamingo
suur	Rolltrappe	Helm
weich	raffle	Horn
Kabel	Gleis	

5. Diskussion

5.1 Itemanalyse

5.1.1 Itemschwierigkeit

Bei dieser ersten Version des Tests lagen nur 19 der insgesamt 57 Items bezüglich Schwierigkeitsgrad in einem idealen Bereich, 2 sind zu schwierig, 33 sind eher zu leicht. Diese Verteilung wird sich allerdings noch etwas in Richtung höhere Schwierigkeiten verschieben, wenn die Stichprobe der Kinder des ersten Kindergartenjahres bis zu einem optimalen N aufgefüllt wird. Für eine nächste Testversion können aber trotzdem einige der 33 leichten Items ausgeschlossen werden um Platz für Items mit einer höheren Schwierigkeiten zu schaffen. Andererseits kann man diese Werte auch so akzeptieren, wie sie sind, denn der Mittelwert aller Schwierigkeiten liegt bei 0.81, was noch näher am idealen Bereich liegt als z.B. die Mittelwerte von 0.83 bis 0.97 für die verschiedenen Altersklassen vom rezeptiven WWT 6-10.

Ein Itemset mit einer optimalen Verteilung der Itemschwierigkeiten aufzustellen ist gerade bei rezeptiven Wortschatztests schwierig, weil das rezeptive Wissen der Kinder meist schwer einzuschätzen ist und oft auch unterschätzt wird. Dies war bei dieser Testentwicklung der Fall und auch bei weiteren. Elben und Lohaus (2000) mussten nach eigenen Angaben viele Verben und Adjektive aus ihrer ersten Version des MSVK ausschliessen, weil diese eine zu tiefe Schwierigkeit aufwiesen. Sie orientierten sich bei der Konstruktion ebenfalls an Informationen über den produktiven Wortschatz und unterschätzten die grosse Differenz zwischen Produktion und Rezeption. Rupp (2008) stiess auf die gleichen Probleme und auch Glück (2011) musste in seinem rezeptiven WWT 6-10 teilweise ungünstige Itemschwierigkeiten akzeptieren. Bei diesen beiden Verfahren wirkt sich aber erschwerend aus, dass sie itembasiert vorgehen und für produktive und rezeptive Untertests die gleichen Itemsets benutzen. Unter diesen Bedingungen ist es sehr schwierig genügend Schwierigkeitsgrade für beide Modalitäten und alle Altersklassen zu finden.

Kauschke und Siegmüller (2010) geben für ihre PDSS keine Itemschwierigkeiten an. Möglicherweise hatten aber auch sie Mühe eine optimale Verteilung der Itemschwierigkeiten zu erreichen, denn im Untertest Nomenverständnis müssen mit nur 20 Items alle Schwierigkeitsgrade für 2- bis fast 7-jährige Kinder abgedeckt werden. Bei den Normierungstabellen

der PDSS zeigt sich, dass im überdurchschnittlichen Leistungsbereich der höchsten Alterskategorie keine T-Werte mehr vorhanden sind, was ein Hinweis auf fehlende schwierige Items sein könnte. Die T-Werte von 40-60 für eine durchschnittliche Leistung sind aber immer vorhanden.

5.1.2 Trennschärfe

Die Trennschärfen dieses Tests liegen zwischen -0.01 und 0.61. Der Mittelwert aller Items beträgt 0.31 und liegt somit in einem mittleren Bereich. Nach Ausschluss der Items mit niederen Trennschärfen, liegen die Werte noch zwischen 0.30 und 0.61 und sie haben einen Mittelwert von 0.41, was für einen Test ideal wäre. Diese Werte sind vergleichbar mit jenen von bereits bestehenden Tests. Der WWT 6-10 weist in seinem rezeptiven Untertest Trennschärfen-Mittelwerte von 0.19 bis 0.38 für die verschiedenen Altersklassen auf. Fünf der neun Altersklassen weisen einen Trennschärfen-Mittelwert <0.30 auf und liegen somit im nicht optimalen Bereich. Rupp hat in ihrem Wortverständnis-Verfahren (2008) Trennschärfen von 0.00 - 0.66 und einen niedrigen Mittelwert von 0.24. Nur 13 der insgesamt 30 Items weisen Werte im mittleren und hohen Bereich auf.

Die PDSS weist in ihren rezeptiven Wortschatztests mittlere und hohe Trennschärfen-Mittelwerte von 0.39 - 0.62 in den verschiedenen Altersklassen auf.

5.2 Reliabilität

Die Reliabilität wurde in diesem Test durch die Berechnung der Kennzahl Cronbachs Alpha bestimmt. Sie liegt mit 0.856 gemäss Bühner (2005) in einem mittleren Bereich und gemäss Bortz und Döring (2006) in einem Bereich, den sie für einen guten Test fordern. Verglichen mit anderen Verfahren, ist dies ebenfalls ein guter Wert. Die Reliabilität des MSVK liegt beim rezeptiven Wortschatztest bei 0.82 und beim WWT 6-10 liegt sie für den rezeptiven Untertest bei den verschiedenen Altersklassen zwischen 0.75 und 0.92. Kauschke und Siegmüller (2010) schätzen die Reliabilität ihrer quantitativen Untertests der PDSS insgesamt als gut ein. Betrachtet man aber speziell die rezeptiven Wortschatztests, so liegen die Reliabilitäten in den verschiedenen Altersklassen in den meisten Fällen unter 0.80.

5.3 Distraktoren-Analyse

Es ist zwar fraglich, ob eine Distraktoren-Analyse Sinn macht, wenn phonologische und semantische Ablenker eingesetzt werden, die gewollt nicht neutral sind, aber ich habe es in Kap. 4.4 trotzdem versucht. Eine solche Analyse ist wohl eher subjektiv, aber dennoch wichtig, denn auch ein phonologischer oder ein semantischer Ablenker darf nicht extrem attraktiv sein. Geht man davon aus, dass auf gewisse Kinder der phonologische Ablenker stärker wirkt und auf andere der semantische, so sollte bei einem Item beide eine gewisse Anzahl von Falschantworten auf sich ziehen. Beide Ablenker sollten aber auch stärker belegt sein als der neutrale Ablenker.

Die Distraktoren-Analyse zeigt, dass es sehr schwierig ist ein ideales Distraktoren-Set zu einem Zielitem zu finden. Phonologische Ablenker können z.B. zu ähnlich sein, sodass das Hörvermögen und die Aufmerksamkeit des Kindes, die Aussprache und Lautstärke des Untersuchers oder allfälliger Umgebungslärm einen zu grossen Einfluss auf das Test-Ergebnis ausüben. Aufgrund der Distraktoren-Analyse und weiteren Beobachtungen komme ich zum Schluss, dass z.B. *Teich-Teig*, *Bagger-Bagge*, *Chranz-Chratz*, *Heu-Leu* und *falte-Falter* phonologisch zu ähnlich sind. Beim letzten Item waren die Kinder, die auf den Ablenker Falter zeigten, sogar stolz auf sich, weil sie ein so seltenes Wort vermeintlich kannten. Diesem Problem kann begegnet werden, indem man darauf achtet, dass der Unterschied nicht in einer unbetonten Endung liegt und bei Minimalpaaren der unterscheidende Laut für eine genügend grosse Differenz sorgt.

Ein weiteres Problem war, dass das Bildmaterial in einigen Fällen zu wenig eindeutig oder zu attraktiv war und es deshalb zu Überbelegungen von einzelnen Distraktoren kam. Dieses Problem kann aber schnell durch neue Fotos gelöst werden.

Die schwierigste Bedingung ist aber, dass die Ablenker untereinander den gleichen Schwierigkeitsgrad aufweisen und gleich attraktiv sein sollen. Sehr schnell ist einer der drei Ablenker ein zu bekannter Begriff für die Kinder, wird deshalb nicht gewählt und es kommt zu einer unerwünschten Konzentration auf die anderen Ablenker. Da das rezeptive Wortwissen von Kindergartenkindern nicht einfach einzuschätzen ist, ist diesem Problem oft nur durch Versuch und Irrtum zu begegnen.

Bei WWT 6-10, PDSS, MSVK und dem Verfahren nach Rupp (2008) sind keine Informationen über allfällige Distraktoren-Analysen erhältlich, obwohl dies von Lienert und Raatz (1998) für Testentwicklungen sehr empfohlen ist. Das Fehlen einer Distraktoren-Analyse ist allerdings

aufgrund der Schwierigkeiten und des Aufwands, die damit verbunden sind durchaus verständlich. Durch ungünstige Distraktoren wird die wahre Leistung eines Kindes möglicherweise nicht exakt gemessen, liegt aber eine Normierung vor, so ist der Vergleich mit ihr, aussagekräftiger als die absolute Leistung des Kindes.

5.4 Verschiedene Wortarten in einem Wortschatztest

Es ist sehr wichtig die Zusammensetzung des Wortschatzes auch bezüglich der verschiedenen Wortarten zu untersuchen. Obwohl in diesem Test das Hauptaugenmerk auf geeignete phonologische und semantische Ablenker gelegt wurde, wurde auch versucht Verben und Adjektive in den Test aufzunehmen.

Bei den Adjektiven bin ich nun der Meinung, dass diese für diese Testform nicht geeignet sind. In einer nächsten Test-Version würde ich alle Adjektive ausschliessen. Es ist schwierig eine Eigenschaft durch ein Foto eindeutig darzustellen. Auch war es nicht möglich für die Adjektive genügend phonologische Ablenker aus der gleichen Wortart zu generieren. So war es für die Kinder verwirrend, wenn innerhalb einer Vierfeldertafel gleichzeitig Objekte und Eigenschaften mit den Abbildungen gemeint waren. Adjektive müssten in einer anderen Testform, z.B. mit Abbildungen, die Gegensätze darstellen, überprüft werden.

Bei den Verben würde ich weiterhin versuchen, sie in diesen Test aufzunehmen. Man müsste sich aber von der Vorstellung lösen, zu jedem Verb auch einen phonologischen Ablenker anbieten zu können. Dies wird vermutlich nicht immer möglich sein. Die Anzahl der Verben in diesem Test war ursprünglich bei 10, durch die Itemselektion wurde sie allerdings auf 2 reduziert. Bei einer nächsten Testversion sollten wieder mindestens 10 oder mehr Verben angeboten werden, um auch wirklich eine Aussage über den Verbenwortschatz machen zu können.

5.5 Fazit und Ausblick

Die in dieser Arbeit entwickelte erste Version des geplanten diagnostischen Verfahrens kann aufgrund der Analysen und dem Vergleich mit bereits bestehenden Verfahren zwar noch nicht als Test, aber zumindest als Screening genutzt werden.

Werden einige der weiter oben diskutierten Verbesserungsmöglichkeiten umgesetzt, die Validität untersucht und eine Normierung durchgeführt, so kann aus diesem Verfahren ein

gültiger Test für die Untersuchung des rezeptiven Wortschatzes von Kindergartenkindern entstehen.

Für zukünftige Neuentwicklungen von rezeptiven Wortschatztests empfehle ich aufgrund von meinen Erfahrungen, ein paar kürzere Vortests ohne Anspruch auf standardisierte Bedingung durchzuführen und dabei mehrere Ablenkensets pro Item an verschiedenen Kindern zu erproben. In diesen Vortests können auch ab und zu die Überlegungen und Assoziationen der Kinder zu den Items und Ablenkern erfragt werden. So können wertvolle Informationen gewonnen und Fehleinschätzungen vermieden werden.

Tabellenverzeichnis

Tab. 1: Nach semantischen Feldern geordnete Test-Items.....	16
Tab. 2: Statistische Kennzahlen der Gemeinden der Konstruktionsstichprobe	22
Tab. 3: Geschlechterverteilung in der Konstruktionsstichprobe	25
Tab. 4: Verteilung der Sprachkompetenzen in der Konstruktionsstichprobe.....	25
Tab. 5: Anzahl der Kinder nach Einwohnerzahl in ihrer Wohngemeinde.....	25
Tab. 6: Anzahl der Kinder nach Ländlichkeit ihrer Wohngemeinde.....	25
Tab. 7: Anzahl der Kinder nach Finanzkraft ihrer Wohngemeinde.....	26
Tab. 8: Itemschwierigkeiten.....	27
Tab. 9: Trennschärfen.....	30
Tab. 10: Verteilung der Falschantworten auf die drei Distraktoren (Ablenker).....	33

Abbildungsverzeichnis

Abb. 1: Sprachproduktions-Modell nach Levelt (1989), vereinfacht.....	3
Abb. 2: Lexikon-Modell nach Dell (aus Dell et al. 1999).....	4
Abb. 3: Flussdiagramm der einzelnen Schritte bei der Entwicklung des rezeptiven Wortschatztests mit Mehrfach-Wahl-Aufgaben.....	15
Abb. 4: Exemplarische Darstellung des Items Chnospe	19
Abb. 5: Geografische Verteilung der Kindergärten für die Konstruktionsstichprobe.....	20
Bildquelle: https://commons.wikimedia.org/wiki/File:Karte_Zürichsee_V2.png [12.01.2017], verändert	
Abb. 6: Altersverteilung der Kinder der Konstruktionsstichprobe.....	24
Abb. 7: Häufigkeiten der Schwierigkeitsindizes nach Ratekorrektur.....	28
Abb. 8: Trennschärfen.....	31
Abb. 9: Verteilung der Probandenrohwerte.....	40

Literaturverzeichnis

- Aitchison, J. (1988). *Words in the mind. An introduction to the mental lexicon* (2. Aufl.). Oxford: Blackwell.
- Bickel, H., Schläpfer, R., & Arquint, J. C. (Hg.) (2000). *Die viersprachige Schweiz* (2., neubearb. Aufl.). Aarau: Sauerländer.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer Medizin Verlag.
- Bühner, M. (2005). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Christen, H., Glaser, E., Friedli, M. & Renn, M. (2013). *Kleiner Sprachatlas der deutschen Schweiz* (5. Aufl.). Frauenfeld: Huber.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist Models of Language Production: Lexical Access and Grammatical Encoding. *Cognitive Science*, 23 (4), 517-542.
- Eheim, W. P. (1977). Zur Beeinflussbarkeit der Schwierigkeit von Mehrfachwahl-Aufgaben. *Diagnostica* 23, 193-198.
- Elben, C. E. & Lohaus A. (2000). *MSVK: Marburger Sprachverständnistest für Kinder. Handanweisung*. Göttingen: Hogrefe.
- Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik. Mit Hinweisen zur Intervention* (2., überarb. und erw. Aufl.). Göttingen: Hogrefe.
- Gallmann, H. (2015). *Zürichdeutsches Wörterbuch* (3., überarb. und erw. Aufl.). Zürich: Verl. Neue Zürcher Zeitung.
- Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige. WWT 6-10* (2., überarb. Aufl.). München: Urban & Fischer.
- Grimm, H. & Aktas, M. (2002). Entwicklungstests im Vorschulalter: Beurteilung ihrer Nützlichkeit durch praktisch tätige Psychologen. *Frühförderung interdisziplinär*, 21, 163–177.
- Grimm, H., Aktas, M.; & Frevert, S. (2010). *SETK 3-5 Sprachentwicklungstest für drei - bis fünfjährige Kinder*. Göttingen: Hogrefe.
- Grimm, H. (Hg.) (2000). *Sprachentwicklung. Enzyklopädie der Psychologie. Themenbereich C, Serie III, Sprache*. Göttingen: Hogrefe.
- Grimm, H. & Schöler, H. (2001). *Heidelberger Sprachentwicklungstest (HSET)* (2. verbesserte Auflage). Göttingen: Hogrefe.
- Grissemann, H., Baumberger, W. & Hollenweger, J. (1991). *Heidelberger Sprachentwicklungstest (H-S-E-T) Schweizer Version*. Bern: Huber.


-
- Kany, W. & Schöler, H. (2010). *Fokus: Sprachdiagnostik. Leitfaden zur Sprachstandsbestimmung im Kindergarten* (2., erw. Aufl.). Berlin: Cornelsen Scriptor.
- Kauschke, C. (2003). Entwicklung, Störungen und Diagnostik lexikalischer Prozesse - Wortverständnis und Wortproduktion. *Sprache, Stimme und Gehör*, 27, 110–118.
- Kauschke, C. & Siegmüller, J. (2010). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS)* (2., völlig überarb. Aufl.). München: Urban & Fischer.
- Kiese-Himmel, C. (2005). *AWST-R Aktiver Wortschatztest für 3- bis 5-jährige Kinder. Revision*. Göttingen: Hogrefe.
- Kiese, C. & Kozielsky, P.-M. (1996). *Aktiver Wortschatztest (AWST 3-6)*. Göttingen: Beltz.
- Landert, K. (2007). *Hochdeutsch im Kindergarten? Eine empirische Studie zum frühen Hochdeutscherwerb in der Deutschschweiz*. Zugl.: Zürich, Univ., Diss., 2007. Bern: Lang.
- Leue, R. & Holthausen, L. (2011). *Duden, das Kindergarten-Wörterbuch* (2., aktualisierte Aufl.). Mannheim: Dudenverlag.
- Levelt, W. J. (1989). *Speaking: From Intention to Artikulation*. Cambridge MA: MIT Press.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Mayer, A. (2012). Semantisch-lexikalische Störungen - Basisartikel. *Sprachförderung und Sprachtherapie*, 2, 54-62.
- Mittring, G. & Rost, D. H. (2008). Die verflixten Distraktoren. *Diagnostica*, 54 (4), 193–201.
- Petermann, F. (2012). *Sprachstandserhebungstest für Kinder im Alter zwischen 5 und 10 Jahren (SET 5-10)*. (2., überarb. Aufl.). Göttingen: Hogrefe.
- Rupp, S. (2008). *Modellgeleitete Diagnostik bei kindlichen lexikalischen Störungen* (1. Aufl.). Idstein: Schulz-Kirchner.
- Rupp, S. (2011). Diagnostik semantisch-lexikalischer Spracherwerbsstörungen. Vorschlag einer modellgeleiteten Vorgehensweise. *SAL-Bulletin*, 142, 5-18.
- Rupp, S. (2013). *Semantisch-lexikalische Störungen bei Kindern. Sprachentwicklung: Blickrichtung Wortschatz*. Berlin: Springer.
- Schulz, P. & Grimm, A. (2012). Spracherwerb. In H. Drügh, S. Komfort-Hein, A. Krass, C. Meier, G. Rohowski, R. Seidel & H. Weiss (Hrsg.), *Germanistik. Sprachwissenschaft – Literaturwissenschaft – Schlüsselkompetenzen* (S. 155-172). Stuttgart: J. B. Metzler.
- Suchodoletz, W. (2003). Umschriebene Sprachentwicklungsstörungen. *Monatsschrift Kinderheilkunde* 151, 31–37.

Anhang

Anhang 1: Elternbrief.....I

Anhang 2: Protokollbogen.....II

Anhang 1: Elternbrief



Esther Glaus
Absender

An die Eltern der Kindergartenkinder von Rapperswil-Jona

14.10.2016

Besuch im Kindergarten

Sehr geehrte Eltern


Für meine Abschlussarbeit im Logopädie-Studium besuche ich mehrere Kindergärten im Raum Zürichsee und im Linthgebiet. Im Dezember darf ich auch im Kindergarten Busskirch vorbeikommen.

Gerne informiere ich Sie über meine Arbeit und den Besuch. In den letzten Monaten entwickelte ich ein Verfahren um den Wortschatz von Kindern zu überprüfen. Neu daran ist, dass es in Mundart ist. Nun soll dieses Verfahren erprobt werden. Ich möchte dabei herausfinden, ob die Wörter, die ich zusammengestellt habe, wirklich für Kindergartenkinder geeignet sind oder ob sie zu leicht oder zu schwierig sind. Dazu sage ich einem Kind ein Wort vor und es soll mir aus vier Bildern das passende herausuchen. Es geht dabei nicht darum, die Kinder zu testen, sondern darum mit Hilfe der Kinder mein entwickeltes Verfahren zu testen. Ich bin also auf die Mitarbeit von Kindergärten und Kindergartenkindern angewiesen und freue mich sehr, dass der Kindergarten Busskirch bei meinem Projekt mitmacht.

Falls Sie noch mehr über meine Abschlussarbeit wissen möchten, dürfen Sie sich gerne an mich oder die Kindergartenlehrpersonen wenden.

Vielen Dank für Ihre Unterstützung und freundliche Grüsse

Esther Glaus



Anhang 2: Protokollbogen

Ort:		Datum/Zeit:	/	
Kind:		1./ 2. Kiga:	1	2
Geb.:		Geschlecht:	♀	♂

Nr.	Ziel-Item	Ablenker phonologisch	Ablenker semantisch	Ablenker divers
1	Chronä	Bohnä	Huet	Rüebli
2	Turm	Wurm	Sturm	Muulwurf
3	Iglu	Igel	Zält	Wieseli
4	Schatz	Spatz	Pirat	Chatz
5	Ellebogä	Rägebogä	Fuess	Pfiilbogä
6	Hammer	Hummer	Bieli	Fisch
7	Bagger	Bagge	Chrane	Wade
8	Fuchs	Büchs	Chatz, rot	Fuess
9	Gländer	Ständer	Haag	Drahtzuun
10	Bank	Band	Stuehl	Gigampfi
11	troche	trinke	nasse Hund	Socke
12	suur	Muur	Schoggi	Uhr
13	alt	chalt	nüüs Huus	Frühlingslandschaft
14	weich	Teich	härts Chopfsteipflaster	Holz
15	truurig	Truube	fröhlich	Mandarine
16	Pyramide	Margarine	Tutenchamun	Schokolade
17	Gschirr	Gschmier	Kochtöpfe	Gschänk
18	Ferse	Ferli, Ferkel (Säuli)	Ellbogen	Fernseeh
19	Chnospe	Chnopf	Blueme, offe	Nuss
20	abtröchne	abbreche	abwäsche	wüsche
21	Schruube	Truube	Nagel	Tuube
22	Stecker	Wecker, Wegger	Kabel	Bagger
23	Kabel	Schnabel	Seil	Rüssel
24	schniide	schniie, schneie	raffle	schriibe
25	Helikopter	Ägypter	Libelle	Flugzüg
26	rissä	bissä	falte	trinke
27	Gurt	Joghurt	Hoseträger	Gurke
28	Chünni	Chämmi	Stirn	Füür
29	Tanne	Pfanne	Laubbaum	Channe, Chante
30	Klinke	Blinker	Knauf	Pedale

31	Chranz	Chratz, Chräbel	Rettigsring	Zopf
32	Pinzette	Prinzessin	Zange	Zäddel
33	Gitter	Gwitter	Fischernetz	Zuun
33	Nagel	Nabel	Hammer	Hagel
35	stäche	bräche	schniide	streichle
36	Sattel	Säbel	Gepäckträger	Blockflöte
37	Zapfe	Zopf	Nuss	Eichhörnli
38	Rolltrappe	Rolltäsche	Lift	Strickleitere
39	raffle	Affe	schniide	Rabe
40	Gleis	Geiss	Zug	Unterfüehrig
41	Drähtür	Dräht	Drähschalter	Huustür
42	Pinzel	Insle	Farbchaste	Spachtel
43	falte	Falter	schniide	Falle
44	Chräbs	Chräbeli	Seestern	Chäfer
45	Fadä	Fahnä	Seil	Flade
46	Muschle	Masche	Schnäg	Nuss
47	Leitere	Liter	Stäge	Dirigent
48	chrüüche	rüüche	ränne	Chueche
49	Flamingo	Flamme	Storch	Füürlöscher
50	Heu	Leu	Wiese, grün	Pneu, Pnö
51	tanke	Franke	Auto, fahrend	Trampolin springe
52	Tram	Rahm	Bus	Joghurt
53	büggle	zügle	lisme	bücke
54	Helm	Halm	Huet	Töff
55	schellä	Chellä	schniide	Schälle
56	Horn	Dorn	Rüssel	Morchle
57	Teich	Teig	Fluss	Zopf
xx	Zebra	Hund	Chatz	Schwan