



**Kognitive Biopsychologie
und Methoden**

Departement für Psychologie

Leitfaden zur Erstellung von Experimentalpraktikums-/Bachelor-/ und Masterarbeiten in der Abteilung Kognitive Biopsychologie und Methoden

Departement für Psychologie

Verfasser:

PD Dr. Siegfried Macho (siegfried.macho@unifr.ch)
Lic. Phil. Michael Munz (michael.munz@unifr.ch)

Fribourg, August 2013
Update, Herbst 2017

Inhaltsverzeichnis

1. Gliederung des Berichts	1
2. Titelseite	1
2.1 Titel	1
2.2 Autorennamen und Affiliation	2
2.3 Author Note	2
3. Abstract	2
4. Einleitung	3
4.1 Theoretische Erklärungen	5
4.2 Hypothese(n)	6
5. Methode	6
5.1 Versuchspersonen (Vpn)	7
5.2 Durchführung	7
5.3 Versuchsmaterial	9
5.4 Unabhängige Variablen, abhängige Variablen und Versuchsplan (Experimentelles Design)	11
5.5 Statistische Analyse	13
5.5.1 Darstellung von statistischen Prozeduren	13
6. Ergebnisse	14
6.1 Darstellung von Variablen, Statistiken und Zahlen	15
6.2 Inferenzstatistische Ergebnisse	15
6.2.1 Darstellung von Ergebnissen statistischer Analysen	16
6.2.2 Konventionen bezüglich der Darstellung von Ergebnissen der wichtigsten Statistischen Verfahren	17
6.2.3 Beschreibung von absolut insignifikanten Ergebnissen	19
6.2.4 Beschreibung von knapp nicht signifikanten Ergebnissen	20
6.2.5 Intuitive Prüfung von inferenzstatistischen Ergebnissen	20
6.3 Deskriptive Statistik	21
6.3.1 Verwendete Anzahl Dezimalstellen in Tabellen und Text	23
6.3.2 Allgemeine Richtlinien zur Gestaltung von Tabellen	24
6.3.3 Allgemeine Richtlinien zur Gestaltung von Abbildungen	26
7. Diskussion und Ausblick	28
8. Literaturverzeichnis	29
8.1 Zitieren IM Text	30
8.1.1 Werke mit zwei AutorInnen:	30
8.1.2 Bei mehr als zwei Autoren : bei 1. Nennung ausgeschreiben:	30
8.1.3 Werke mit mehr als fünf Autoren:	30
8.1.4 Mehrere Angaben in einem Klammerausdruck:	30
8.1.5 Autoren mit gleichem Familiennamen:	31

8.1.6 Wörtliche Zitate	31
8.1.7 Längere Zitate (mehr als 40 Wörter)	31
9. Anhang	31
10. Style	32
10.1 Zahlen ausgedrückt als Zahl:	32
10.2 Zahlen ausgedrückt als Wort:	32
10.3 Dezimalzahlen:	32
10.4 Schriftgrösse und -art	32
Anhang A: Boxplots	33
Anhang B: Interpretation und Darstellung von Interaktionen	36
11. Eigenständigkeitserklärung	39

Vorwort

Der vorliegende Leitfaden soll den Studierenden helfen, möglichst verständliche und sachgerechte Berichte zu verfassen, welche den internationalen Standards entsprechen. Er enthält die folgenden vier Arten von Inhalten:



Prinzipien:

Sie erklären in abstrakter Weise den Zweck und die Struktur der einzelnen Abschnitte eines Experimentalberichts.



Positive Beispiele:

Diese dienen der Illustration eines zuvor präsentierten Prinzips. Beispielsweise zeigt Versuchsplan:

Bsp. , wie ein Versuchsplan konkret beschrieben werden kann. Positive Beispiele sind in gewöhnlicher Schrift ausgeführt und besitzen einen Titel, welcher sich in Klammern hinter der Beispielsnummer befindet.



Negative Beispiele:

Hierbei handelt es sich um Auszüge aus früheren studentischen Experimentalberichten. Diese wurden ohne Korrektur übernommen und sind kursiv und in leicht verkleinerter Schrift ausgeführt. Zusätzlich werden unsere Kritikpunkte präsentiert (in gewöhnlicher Schrift).



Bemerkungen:

Sie dienen der näheren Erläuterung von Sachverhalten.

Neben der Befolgung der in diesem Leitfaden dargestellten Prinzipien erwarten wir von einem guten Experimentalbericht, dass er in korrektem Deutsch verfasst und von der Form her ansprechend ist (z.B. sollte der Text keine unnötigen Leerzeichen enthalten und bei Verwendung von Blocksatz sollten Silbentrennungen durchgeführt werden um grössere „Löcher im Text“ zu vermeiden).

Wir haben uns bemüht, diesen Leitfaden möglichst fehlerfrei zu gestalten. Dennoch kann nicht garantiert werden, dass er keine Fehler, Ungenauigkeiten oder Unklarheiten enthält. Wir sind daher für jede Art von Rückmeldung dankbar, welche helfen kann, diesen Leitfaden weiter zu verbessern (die Email Adressen der Ansprechpartner finden sich auf dem Deckblatt).

Weitere hilfreiche Links:

<http://www.apastyle.org/learn/index.aspx>, retrieved on 27.3.2017

<http://www.discoveringstatistics.com/docs/writinglabreports.pdf>, retrieved on 22.5.2017

1. Gliederung des Berichts



Prinzip:

- ◆ Der Bericht soll ein Deckblatt besitzen, welches den Titel, die Liste der Autoren (mit Email Adresse), und das Datum der letzten Änderung enthält (vergleiche das Deckblatt zu diesem Leitfaden).
- ◆ Der Bericht sollte in die folgenden Abschnitte gegliedert sein:
 1. Titelseite
 2. Inhaltsverzeichnis
 3. Abstract
 4. Einleitung
 5. Methode
 - 5.1 Versuchspersonen
 - 5.2 Durchführung
 - 5.3 Versuchsmaterial
 - 5.4 Versuchsplan und abhängige Variablen
 - 5.5 Statistische Analyse
 6. Ergebnisse
 7. Diskussion und Ausblick
 8. Literaturverzeichnis
 9. Anhang
 10. Eigenständigkeitserklärung



Bemerkungen zur vorgeschlagenen Gliederung:

- (i) Die präsentierte Gliederung entspricht in etwa dem Aufbau von wissenschaftlichen Berichten in internationalen experimentalpsychologischen Journalen (Vergleiche *Publication Manual of the American Psychological Association* (6th edition, Chapter 1).
- (ii) Die dargestellte Gliederung betrifft Berichte mit nur einem Experiment, wie dies im Rahmen der experimentalpsychologischen Übungen der Fall ist. In Berichten mit mehreren Experimenten werden die Absätze 1-4 für jedes Experiment wiederholt. Zusätzlich wird dem Bericht eine einführende Einleitung voran gestellt und den Abschluss bildet eine allgemeine Diskussion.
- (iii) Im Folgenden wenden wir uns der Diskussion der einzelnen Abschnitte zu und beschreiben für die einzelnen Kapitel die zentralen Prinzipien, Richtlinien und Konventionen, welche bei deren Erstellung zu beachten sind.

2. Titelseite

2.1 Titel

Empfohlen wird ein Titel nicht länger als 12 Worte. Darin sollten keine Abkürzungen benutzt werden. Das Format enthält Gross- und Kleinschreibung, mittige Ausrichtung, obere Seitenhälfte.

Auf den folgenden Seiten steht ein „running head“ in der Kopfzeile, sowie die Seitenzahl. Der Running Head ist ein kurzer Arbeitstitel (z.B. die ersten 2 – 3 Worte des Titels oder eine Kurzfassung dessen).

2.2 Autorennamen und Affiliation

Vor- und Nachname des Verfassers, ohne Titel (Dr., M.Sc. etc), Angabe des Ortes an dem die Forschung durchgeführt wurde (d.h. Universität, Stadt, Land)

2.3 Author Note

Vollständige Korrespondenzadresse des Autors inkl. Emailadresse

3. Abstract



Prinzip:

Das Abstract sollte die Kernbotschaft der Untersuchung wiedergeben. Es sollte auf möglichst konzise Art und Weise den Inhalt in 150 - 250 Worten zusammenfassen. Es sollte folgende Punkte beinhalten:

- ◆ Überblick über die theoretischen Fragestellung bzw. die Zielsetzung der Untersuchung
- ◆ Die zentralen experimentellen Manipulationen
- ◆ Die zentralen Hypothesen
- ◆ Die wichtigsten Ergebnisse
- ◆ Die Schlussfolgerung oder Anwendung

Das Abstract beginnt auf einer neuen Seite. Das Wort „Abstract“ wird in Gross- und Kleinbuchstaben mittig platziert. Es enthält keine Absätze. Ein wichtiger Aspekt betrifft die Verständlichkeit: Das Abstract sollte keine komplexen, unerklärten Begriffe beinhalten. Das Abstract muss in sich verständlich sein, es darf keine Verweise auf den späteren Text enthalten.

Hier ein negatives Beispiel, welches verschiedene Schwächen aufweist.



Bsp. 3.1:

(...) Die 48 Versuchspersonen (Vpn), davon 32 weiblich, werden in vier Gruppen geteilt. Jede Gruppe bekommt vom Versuchsleiter (VL) ein Konsonanten-Item vorgesprochen, gefolgt von einer Zahl. In der ersten und zweiten Gruppe werden unähnliche Konsonanten-Items verwendet, in der dritten und vierten ähnliche. Die Vp muss die Zahl, welche sie vom VL gehört hat, wiederholen und darauf möglichst schnell in Dreierschritten korrekt rückwärtszählen (Distraktoraufgabe), bis der VL „Stop“ sagt. Die Vpn der ersten und der dritten Gruppe müssen nun das zuvor gehörte Konsonanten-Item aussprechen. Die Vpn der zweiten und der vierten Gruppe müssen nach dem „Stop“ des VL aktiv im Gedächtnis nach dem Konsonanten-Item suchen, bis nochmals ein „Stop“ erklingt. Erst darauf dürfen sie das Konsonanten-Item aussprechen.

Im vorliegenden Experiment sind fünf Hypothesen untersucht worden: Die erste Hypothese - „Je länger das Behaltensintervall ist, desto kleiner wird der Prozentsatz der korrekten Antworten“ - wird durch das Experiment bestätigt.

(...)

Die vierte Hypothese, „die Konsonanten-Items werden im Mittel besser rekonstruiert, wenn die Vpn aktiv während der Latenzzeit in ihrem Gedächtnis danach suchen, als bei unmittelbarer Wiedergabe“, wird durch die vorliegenden signifikanten Ergebnissen bestätigt.

Kritikpunkte:

- (i) Details, wie die Anzahl Vpn bzw. die Anzahl männlicher und weiblicher Vpn gehört nicht ins Abstract (ausser es hat eine spezielle Bedeutung).
- (ii) Das Abstract enthält zu viele Details der Prozedur.
- (iii) Ein weiteres Problem betrifft die Verwendung von Begriffen, deren Bedeutung nicht spezifiziert wird, z.B. *Konsonanten-Item*, *Behaltensintervall* und *Latenzzeit*.
- (iv) Das Abstract soll keine Absätze enthalten.

Das folgende Beispiel demonstriert, wie in einem Abstract die zentralen Aspekte eines Experiments knapp und verständlich dargestellt werden können.

**Bsp. 3.2 (Abstract):**

In einem Experiment wurde der Einfluss der proaktiven Hemmung auf die Behaltensleistung von Konsonantentripel untersucht. Hierbei wurden die folgenden Variablen manipuliert: Erstens, die Dauer des Behaltensintervalls (BI), d.h. des Zeitintervalls zwischen der Präsentation eines Konsonantentripels plus einer dreistelligen Zahl und dem Signal zur Wiedergabe. In dieser Zeit mussten die Vpn eine Ablenkungsaufgabe (Rückwärtszählen in Dreierschritten, beginnend mit der präsentierten dreistelligen Zahl) durchführen. Zweitens, die Ähnlichkeit zwischen den präsentierten Konsonantentripel; drittens, das Vorhandensein bzw. die Abwesenheit einer Latenzzeit. Hierbei handelte es sich um eine kurze Zeitspanne von 6 Sekunden, welche unmittelbar auf das BI folgte und in welcher die Vpn versuchen sollten, das präsentierte Konsonantentripel aus dem Gedächtnis leise für sich abzurufen. Die Ergebnisse des Experiments bestätigen die zentralen Hypothesen des Experiments: (1) Die Wiedergabe nahm mit der Länge des Behaltensintervalls ab; (2) Das Vorhandensein der Latenzzeit führte zu einer verbesserten Abrufleistung.

4. Einleitung

Das Wort „Einleitung“ wird, anders als die anderen Teile des Artikels, nicht explizit über die Einleitung geschrieben, da die Position im Text den Inhalt des Abschnitts bereits deutlich macht.

**Prinzip:**

Der Einleitungsteil erfüllt die folgenden wesentlichen Zwecke:

- ◆ Darstellung der Fragestellung und des Zieles der aktuellen Untersuchung
- ◆ Hinweise auf früher vorgeschlagene Ansätze zur Behandlung der Problematik
- ◆ Explikation des theoretischen Hintergrundes, auf dem die Hypothesen beruhen
- ◆ Die Darstellung der Sachhypothesen am Schluss der Einleitung
- ◆ Am Ende der Einleitung wird eine kurze Übersicht zur Struktur der in dem Bericht dargestellten Inhalte gegeben (Was folgt im Anschluss an die Einleitung, wie ist der Text gegliedert).

Die Einleitung sollte sich nicht in irgendwelchen Details der Durchführung oder anderen Aspekten der Methode verstricken. Der Text wird so geschrieben, als würde die Arbeit mit der Einleitung beginnen, es darf keine Verweise auf das Abstract geben.

**Bsp. 4.1:**

Diese Silben bestehen je aus drei nicht ähnlichen Konsonanten (zum Beispiel: WHF) und werden im Folgenden unähnliche Konsonanten-Items genannt. Im Experiment, das im vorliegenden Bericht beschrieben wird, sind bei einer Bedingung nur solche unähnliche Items verwendet worden. Bei einer zweiten Bedingung sind zwei unähnliche Konsonanten-Items als Initial-Items eingesetzt worden. Alle weiteren Konsonanten-Items entsprangen der Permu-

tation der Konsonanten dieser Initial-Items. Die Konsonanten-Items der zweiten Bedingung werden ähnliche Konsonanten-Items genannt.

Kritikpunkt:

Die gezeigte Passage erfüllt keine der oben genannten Kriterien für eine gute Einleitung. Sie schildert Details des Versuchsmaterials, die für die Darstellung des Zieles des Experiments irrelevant sind.

Das folgende Beispiel illustriert den Aufbau einer Einleitung.



Bsp. 4.2 (Struktur einer Einleitung):

Das Arbeitsgedächtnis (AG) stellt eine zentrale Komponente der kognitiven Ausstattung des Menschen dar, welches für das Lernen komplexer Inhalte von zentraler Bedeutung ist. Das heute allgemein anerkannte Grundmodell des Arbeitsgedächtnisses stammt von Baddeley & Hitch (1974). Es gibt jedoch zahlreiche frühere Untersuchungen verschiedener Aspekte des AG.

Eine klassische Untersuchung zum Vergessen im AG wurde von Peterson & Peterson (1959) durchgeführt.

[...]

Das Ziel des aktuellen Experiments bestand in einer Replikation der zentralen Ergebnisse des ersten Experiments von Peterson & Peterson (1959).

[...]

Aufgrund dieser Überlegungen ergeben sich die zentralen Hypothesen des Experiments: (1) Je länger die Zeit zwischen Präsentation und Wiedergabe der präsentierten Konsonanten desto geringer ist die Wahrscheinlichkeit, dass diese korrekt wiedergegeben werden, sofern während der Behaltensdauer eine verbale Ablenkungsaufgabe erfolgt. Im Modell von Baddeley & Hitch (1974) lässt sich diese Hypothese wie folgt begründen: Die Ablenkungsaufgabe (Rückzählen von einer vorgegebenen Zahl in Dreierschritten) verhindert das Behalten der Konsonanten in der phonologischen Schleife. Je länger nun der Prozess des Zurückzählens andauert, desto geringer ist die Wahrscheinlichkeit, dass sich das zu behaltende Item noch in der phonologischen Schleife befindet.

[...]



Bemerkung zum Aufbau der Einleitung

- ◆ Die Einleitung beginnt mit einer allgemeinen Bemerkung, welche die Bedeutung des Untersuchungsgegenstandes (im aktuellen Fall des Arbeitsgedächtnisses) für die psychologische Forschung klar macht: Warum ist das Thema relevant? Hierauf folgt eine Passage, welche relevante theoretische und empirische Forschung zum Thema darstellt, wobei eine systematische Fokussierung auf den Gegenstand der aktuellen Untersuchung erfolgen sollte. Im aktuellen Fall wird kurz das gängige Modell des AG von Baddeley & Hitch (1974) erwähnt und sodann auf die Arbeit von Peterson & Peterson (1959) eingegangen, welche in enger Beziehung zum aktuellen Experiment steht.
- ◆ Anschliessend erfolgt die Identifikation der wissenschaftlichen Lücke, die das aktuelle Experiment füllen wird. Diese „Lücke“ kann auch in der Notwendigkeit einer Replikation bestehen. Es folgt die Darstellung der Zielsetzung bzw. der zugrunde liegenden Idee des aktuellen Experiments.
- ◆ Zum Schluss werden die zentralen Hypothesen dargestellt, mit einer theoretischen Begründung der einzelnen Hypothesen. Im aktuellen Fall wird hierzu das Modell von Baddeley & Hitch (1974) herangezogen



Bemerkung zur theoretischen Argumentation

Die theoretische Argumentation sollte – unter anderem – den folgenden Kriterien genügen:

- ◆ Die Darstellung muss verständlich sein. Es sollten daher keine Begriffe mit unklarer Bedeutung verwendet werden ohne diese näher zu erläutern.
- ◆ Es sollte ein „roter Faden“ in der Argumentation erkennbar sein. Dies impliziert im Wesentlichen zwei Dinge:
 - (i) Ein Gedankengang sollte möglichst nicht durch unwichtige Nebenaspekte unterbrochen werden.
 - (ii) Die Beziehung zum Vorangegangenen sollte klar sein, d.h. abrupte Gedankensprünge sind zu vermeiden. Dies wird erreicht, indem Übergänge explizit gemacht werden.
- ◆ Begriffe und Abkürzungen sind so zu wählen, dass sie leicht zu merken sind, und im gesamten Text konsistent verwendet werden.
- ◆ Abkürzungen sind von Zeit zu Zeit wieder zu erklären.
- ◆ Allgemein gilt: Die Autoren sollten als Zielperson des Berichts von einer Leserin ausgehen, die zwar das methodische Verständnis mitbringt (Kenntnis der statistischen Verfahren und experimentellen Methodik), jedoch keinerlei Wissen über die behandelte Thematik besitzt, d.h. sie kennt weder Theorien noch empirische Befunde zur aktuellen Fragestellung.

4.1 Theoretische Erklärungen



Prinzip:

Theoretische Erklärungen haben den folgenden Kriterien zu genügen:

- ◆ Sie sollten möglichst einfach sein: Unter zwei möglichen Erklärungen ist jene zu wählen, welche einfacher ist, d.h. weniger theoretische Annahmen macht (=Prinzip des *Occamschen Rasiermessers*).
- ◆ Die Erklärung muss entweder auf empirischen Daten oder auf bestätigten Mechanismen basieren.
- ◆ Man hüte sich vor unbewiesenen Aussagen und Spekulationen!



Bsp. 4.1.1:

Jede Informationsverarbeitung, wie auch das Aufnehmen von gehörten Konsonanten-Items, hinterlässt eine Spur im Gedächtnis (Pawlik, 2006). Diese erleichtert für eine gewisse Zeit das Verarbeiten derselben Information (Priming). Die zuerst dargebotenen Konsonanten-Items werden Prime genannt (Städler, 1998). Ähnliche Konsonanten-Items können mit dem entsprechenden Prime verknüpft und so besser gemerkt werden.

Die zweite Hypothese besagt, dass die ähnlichen Konsonanten-Items besser erinnert werden als die unähnlichen.

Kritikpunkt:

Diese Erklärung ist unnötig komplex. Um die Tatsache zu erklären, dass ähnliche Items besser erinnert werden, ist es nicht notwendig, das Phänomen des Priming zu bemühen. Eine einfachere Erklärung ergibt sich dadurch, dass die Stimulusmenge bei ähnlichen Items weniger komplex ist und daher besser gelernt werden.



Bsp. 4.1.2 (Theoretische Erklärung):

Die Hypothese einer verbesserten Abrufleistung nach einer kurzen Latenzzeit, in welcher die Versuchsperson versucht, die zu behaltenden Konsonanten abzurufen, lässt sich durch das Zusammenwirken von zwei Faktoren erklären: Der erste Faktor betrifft das Metagedächtnis, d.h. das Wissen um das eigene Gedächtnis. Im

aktuellen Fall handelt es sich um die Einschätzung, ob ein weiterer Abrufversuch die Wahrscheinlichkeit eines korrekten Abrufs erhöht. Fällt diese Einschätzung negativ aus, so wird kein weiterer Abrufversuch unternommen und eine Antwort erfolgt. Der zweite Faktor betrifft das Phänomen der Hypermnésie, wonach wiederholter Abruf aus dem Gedächtnis zu einer besseren Erinnerungsleistung führen kann. Der verbesserte Abruf nach einer Latenzzeit ergibt sich demnach, indem die Versuchsperson ohne Instruktion ihre Versuche, die Information aus dem Gedächtnis abzurufen zu schnell abbrechen, was zu einer nicht optimalen Leistung führt. Die Einführung der Latenzzeit mit Abrufinstruktion beseitigt dieses Defizit und resultiert daher in einer verbesserten Leistung.

4.2 Hypothese(n)

Prinzip:



- ♦ Am Schluss der Einleitung werden die Fragestellung und die Hypothesen beschrieben. Die Hypothesen werden hierbei abstrakt im Kontext allgemeiner Theorien diskutiert, d.h. ohne Hinweis auf Details der Operationalisierungen.
- ♦ Die Diskussion der Hypothesen beinhaltet die beiden folgenden Aspekte:
 - (i) Eine Darstellung der Theorien und Annahmen, aus denen sich die Hypothesen ergeben. Diese Darstellung muss möglichst klar ausführen, wie sich die Hypothese aus der Theorie und möglichen Zusatzannahmen ergibt.
 - (ii) Der Inhalt der Hypothese, d.h. der vorausgesagte Effekt ist genau zu spezifizieren: Welches Ergebnis wird unter welchen Bedingungen erwartet und welche Richtung hat der Effekt: Erhöht oder verringert sich der Wert der abhängigen Variable?



Bsp. 4.2.1 (Darstellung eines theoretischen Ansatzes mit abgeleiteter Hypothese):

Gemäss der Konzeption von Nairne (2002) folgt das Vergessen im Arbeits- und im Langzeitgedächtnis den gleichen Prinzipien. Demgemäss ist die Ursache des Informationsverlustes nicht – wie im Modell von Baddeley angenommen – der Zerfall von Information. Vielmehr beruht der Verlust von Inhalten auf Interferenz mit konkurrierenden Inhalten. Dies bedeutet, dass die Information nach Ablauf des Behaltensintervalls nicht verloren ist, auch wenn sie nicht unmittelbar abgerufen werden kann. Es sollte daher möglich sein, diese Information wiederzugewinnen.

Hieraus ergibt sich die zentrale Hypothese der Untersuchung: Personen, welche angewiesen wurden, die zu erinnernde Information nicht unmittelbar nach dem Behaltensintervall wiederzugeben, sondern erst nach einer kurzen Latenzzeit, in der sie versuchen sollten, den Inhalt still für sich zu rekonstruieren, sollten eine erhöhte Wiedergabeleistung gegenüber jenen Personen zeigen, welche sofort nach Ablauf des Behaltensintervalls ihre Antwort geben.

5. Methode

Prinzip:



Im Methodenteil werden die Details des Experiments beschrieben. Die Darstellung muss die Leserin in die Lage versetzen, alle Details des Experiments nachvollziehen zu können, so dass sie es replizieren kann. Der Methodenteil gliedert sich in vier Abschnitte:

- ♦ Versuchspersonen (Kapitel 5.1)
- ♦ Durchführung (Kapitel 5.2)

- ♦ Versuchsmaterial (Kapitel 5.3)
- ♦ Experimentelles Design (Kapitel 5.4)
- ♦ Statistische Analyse (Kapitel 5.5)

Diese Abschnitte werden nun im Detail beschrieben.

5.1 Versuchspersonen (Vpn)



Prinzip:

Die Beschreibung der Vpn umfasst die folgende Information:

- ♦ Grundlegende Daten: Anzahl, Alter (Bereich, d.h. Minimum, Maximum und Mittelwert), die Anzahl Frauen (und Männer) und der Beruf.
- ♦ Optional, spezifische Daten, welche für die Untersuchung von Bedeutung sind (Werden z.B. gesunde mit kranken Personen verglichen, so sind Informationen bezüglich der Selektionskriterien der letzteren zu spezifizieren).
- ♦ Die Beschreibung, wie viele Vpn den einzelnen Versuchsgruppen zugeordnet wurden
- ♦ Die eventuelle Elimination oder Ersetzung von Daten aufgrund von Problemen, die während des Experiments auftraten, wird berichtet – z.B. defektes Gerät, fehlende Compliance der Vp, oder Fehler innerhalb der Items (siehe Bsp. 5.1). Dies betrifft jedoch nicht Daten, die nach inhaltlichen Kriterien verworfen werden, wie z.B. Ausreisser oder Daten, die nicht zu einem vorgegebenen Modell passen (Letztere werden im Rahmen der Beschreibung der AV behandelt).
- ♦ Angabe dass die Einwilligungserklärung eingeholt wurde, sowie über Art und Höhe der Aufwandsentschädigung.



Bsp. 5.1 (Bericht von Ausfällen):

Die Ergebnisse von 2 Vpn wurden vollständig durch die Daten zweier zusätzlich getesteter Vpn ersetzt, da erstere Vpn offensichtlich die Instruktion nicht befolgt und nur die extremsten Kategorien verwendet hatten.

5.2 Durchführung



Prinzip:

- ♦ Die Beschreibung der Durchführung sollte der Leserin ein klares Bild über den Ablauf des Experiments vermitteln. Hierzu bietet sich eine hierarchische Darstellung an:
 - (i) Zuerst wird die allgemeine Ablaufstruktur des Experiments beschrieben. Diese beinhaltet:
 - (a) Eine Einleitung, welche den Ablauf bis inklusive Präsentation der (ersten) Instruktion schildert: Hierher gehören Aspekte wie die relative Position von Vp und VI, sitzen sich Vp und VI z.B. gegenüber, oder sitzt die Vp vor dem Computer und die VI schräg links dahinter, damit sie die Vp genau überwachen kann?
 - (b) Die wesentlichen Punkte der Instruktion (Die Instruktion sollte sich im Anhang zum Bericht befinden. Auf diese wird im Text verwiesen).
 - (c) Einen Überblick über die verschiedenen Experimentalblöcke: Wie viele verschiedene Experimentalblöcke umfasst das Experiment? Besitzen sie alle die gleiche Struktur (siehe Bsp. 5.2.1)?

- (d) Einen Schlussteil: Dieser enthält Information darüber, ob die Vp über den Zweck des Experiments aufgeklärt wurde, ob sie Rückmeldung über ihre Leistung erhalten hat, sowie über die Gesamtdauer des Experiments.
 - (ii) Auf der nächsten Stufe erfolgt die Beschreibung des Aufbaus der einzelnen Experimentalblöcke: Gibt es am Beginn des Blocks eine Instruktion, gibt es Aufwärmdurchgänge? Sind diese als solche für die Vp gekennzeichnet? Wie viele gleichartige Durchgänge umfasst der Block (siehe Bsp. 5.2.2)?
 - (iii) Auf der untersten Stufe erfolgt die detaillierte Darstellung des Ablaufs eines einzelnen Durchgangs.
- ◆ Der hier dargestellte Aufbau betrifft komplexe Experimente, welche aus mehreren Blöcken mit komplexen Durchgängen bestehen. Oft gibt es nur einen einzigen Experimentalblock. In diesem Fall entfällt die Erwähnung bzw. Beschreibung des Blocks weg und wird ersetzt durch die Darstellung des Durchgangs. Manchmal gibt es zwar mehrere Experimentalblöcke, aber der Aufbau der Durchgänge ist sehr einfach, sodass deren Darstellung in die Beschreibung eines Experimentalblocks integriert werden kann (siehe Bsp 5.2.2).
 - ◆ Ein weiterer Aspekt der Durchführung betrifft die Beschreibung unterschiedlicher experimenteller Bedingungen. Hier ergibt sich die Frage, ob jede Bedingung separat beschrieben werden soll oder nicht. Hier gilt als Daumenregel, dass bei geringen Unterschieden zwischen den Bedingungen eine einzelne Beschreibung ausreichend ist, welche den Ablauf innerhalb der verschiedenen Bedingungen integriert (siehe Bsp 5.2.3). Sind hingegen die prozeduralen Differenzen zwischen den Bedingungen beträchtlich, so sollten diese separat beschrieben werden.

**Bsp. 5.2.1 (Überblick über die experimentelle Blockstruktur):**

Das Experiment umfasste vier Lern-/Testblöcke, welche alle gleich aufgebaut waren und sich nur durch die verwendeten Lern- und Testlisten unterschieden. Zwischen dem zweiten und dritten Block wurde eine Pause von ca. 5 Minuten eingeschaltet.

**Bsp. 5.2.2 (Darstellung eines Experimentierblocks, mit Beschreibung des Aufbaus der Durchgänge):**

Ein Lern-/Testblock war wie folgt aufgebaut: Am Beginn des Blocks wurde die Instruktion für das Lernen präsentiert. Sie betonte (...)

Nach Drücken der Starttaste durch die Vp erfolgte die Präsentation der Lernliste. Diese bestand aus 48 Wortpaaren, welche sequentiell im 3-Sekunden-Takt ohne Zutun der Vp präsentiert wurden. Die Wortpaare wurden in der Bildschirmmitte in weisser Farbe auf blauem Hintergrund in der Grösse 42 und in der Schriftart Arial präsentiert. Die ersten 4 Wortpaare der Liste fungierten als Primacy und die letzten 4 als Recency Items, welche nicht ausgewertet wurden. Die Unterscheidung zwischen den Primacy / Recency Items und den 40 regulären Wortpaaren war für die Vp nicht erkennbar.

Unmittelbar nach Präsentation des letzten Lernitems erschien die Instruktion für den Test. Diese erklärte (...) Nach Drücken der Starttaste wurden die 160 Testitems dargeboten, bestehend aus den 80 unmittelbar zuvor in der Lernliste präsentierten Worten und 80 neuen. Es wurde jeweils ein Wort in der Bildschirmmitte präsentiert (in gleicher Grösse, Farbe und Schriftart wie in den Lerndurchgängen). Unterhalb des Wortes erschien eine 4-Punkte Skala, auf welcher die Vp ihr Urteil abgab. Die Vp musste eine der 4 Optionen mit der Maus anklicken und dann die OK-Taste drücken. Wurde keine Option gewählt, so erschien eine Fehlermeldung, welche der Vp mitteilte, dass keine der verfügbaren Optionen gewählt wurde. Sobald eine korrekte Eingabe erfolgt war, erschienen automatisch das nächste Wort und die Skala.

Nach Beendigung der 160 Testdurchgänge erschien auf dem Bildschirm die Instruktion für den nächsten Lern-/Test-Block.



Bsp.5.2.3.:

5.2.3.1 Verlauf des einzelnen Durchgangs ohne Wartezeit

Ein einzelner Durchgang des Experiments war wie folgt aufgebaut: Die VI sprach die Buchstaben gefolgt von einer dreistelligen Zahl laut aus (...)

5.2.3.2 Verlauf des einzelnen Durchgangs mit Wartezeit

Ein einzelner Durchgang des Experiments war wie folgt aufgebaut: Die VI sprach die Buchstaben gefolgt von einer dreistelligen Zahl laut aus (...)

Kritikpunkte:

Die beiden Bedingungen (mit und ohne Latenzzeit) unterscheiden sich nur dadurch, dass in der Bedingung mit Latenzzeit ein Zeitintervall von 6 Sekunden nach dem Behaltensintervall eingeschoben wurde. Es macht wenig Sinn aufgrund dieses kleinen Unterschiedes den Aufbau eines Durchgangs für die beiden Bedingungen separat zu beschreiben.

- (i) Als kleiner Kritikpunkt ist noch anzumerken, dass der Ausdruck *Wartezeit* anstelle von *Latenzzeit* verwendet wird, obwohl letzterer Begriff zur Bezeichnung der UV verwendet wurde (Vergleiche hierzu obige Bemerkung zur Bezeichnung von UV und AV).

5.3 Versuchsmaterial



Prinzip:

Versuchsmaterial umfasst die folgenden Komponenten:

- ♦ Die experimentellen Stimuli. Diese Beschreibung sollte klar und verständlich sein, sodass sie für eine Leserin nachvollziehbar ist. Es bietet sich an, den Hergang der Konstruktion des Materials in seiner chronologischen Reihenfolge zu schildern:
 - (i) Prinzipien nach denen das Rohmaterial erstellt wurde.
 - (ii) Das Prinzip der Auswahl des tatsächlich verwendeten Materials.
 - (iii) Ausbalancierung und Randomisierung und die genauen Zwecke, welche damit verfolgt wurden und welche Störvariablen damit ausgeschaltet bzw. kontrolliert werden sollten.
- ♦ Mittel zur Erfassung der Antworten, wie z.B. Fragebögen und Skalen (Referenzen jeweils nicht vergessen).
- ♦ Verwendete Geräte: Computer zur Steuerung, Stoppuhren, etc. Dieser Punkt sollte kurz gehalten werden, falls er nicht zentral für die Untersuchung ist (z.B. falls EEG-Aufnahmen gemacht wurden, ist die Darstellung des verwendeten Geräts relevant). Für EEG Geräte, Analysesoftware u.ä. sollten Softwarename, Versionsnummer und Hersteller angegeben werden.

Es ist hilfreich, die verschiedenen Arten von Stimulusmaterial im Text anhand einzelner, konkreter Beispiele zu illustrieren und – falls möglich – das vollständige Material im Anhang des Berichts zu präsentieren (siehe Abschnitt 9) und im Text darauf zu verweisen.



Bsp. 5.3.1 (Beschreibung des Versuchsmaterials):

Das im Experiment verwendete Material umfasste 34 Silben bestehend aus je 3 Konsonanten, sowie 34 dreistelligen Zahlen. Für die Erstellung der Silben wurde wie

folgt verfahren: Zuerst wurden 40 unterschiedliche Silben erzeugt (z.B. *JPM*). Diese sollten einen möglichst geringen assoziativen Gehalt aufweisen. Weiter wurde darauf geachtet, dass innerhalb einer Silbe keine gleich- bzw. ähnlich lautende Konsonanten auftraten (wie *M* und *N*, oder *P* und *B*). Der Buchstabe *Y* wurde aufgrund seiner grossen Auffälligkeit nicht verwendet. Von den 40 erzeugten Silben wurden schliesslich 34 ausgewählt, wobei 2 Silben für die Probe- und die restlichen 32 für die eigentlichen Testdurchgänge verwendet wurden. Alle Versuchspersonen (Vpn) erhielten daher die gleichen 34 Silben.

Die 34 dreistelligen Zahlen (z.B. 751) wurden nach folgenden Prinzipien erzeugt: Erstens, die Zahl musste ausreichend gross sein, damit beim Herunterzählen in Dreierschritten niemals der dreistellige Bereich unterschritten werden konnte. Zweitens, sollte keine der Zahlen auf den Ziffern 0, 3, 6, oder 9 enden, da dies die Aufgabe des Rückzählens am Beginn erleichtern würde. Wiederum wurden allen Vpn die gleichen 34 Zahlen präsentiert, wobei 2 der Zahlen in den Probe- und die restlichen 32 in den eigentlichen Testdurchgängen verwendet wurden.



Bemerkung zur Vermengung von Stimulusmaterial und Durchführung:

Gemäss unserer Erfahrung kommt es häufig zur Vermengung der Darstellung des Stimulusmaterials mit jener der Durchführung. Hierbei fliessen Details der Beschreibung des Ablaufs in die Beschreibung des Materials ein, welche für das Verständnis des Letzteren unwichtig sind. Sie beeinträchtigen lediglich die Klarheit und Kompaktheit der Beschreibung des Stimulusmaterials.



Bsp. 5.3.2:

Die Reihenfolge der Behaltensintervalle wurde mit Hilfe eines 6x6 sequentiell ausbalancierten lateinischen Quadrats festgelegt. Nachdem so die Reihenfolge der Behaltensintervalle für die erste Liste bestimmt worden war, wurde für jede weitere Liste der jeweils erste Block (die ersten sechs Intervalle) nach hinten verschoben. So wurden sechs unterschiedliche Listen erstellt. Dadurch konnte der Positionseffekt (in diesem Fall der Übungseffekt) kontrolliert werden.

Kritikpunkte:

Diese Darstellung weist folgende Probleme auf:

- (i) Die gegebene Darstellung ist weder schlüssig, noch kann sie nachvollzogen werden, da sie zu unpräzise ist:
Begründung: Eine Liste besitzt 48 Durchgänge (d.h. insgesamt 48 Behaltensintervalle). Wie passt hierzu das 6x6 sequentiell ausbalancierte Lateinische Quadrat (LQ)?
 Wie soll eine Person aus der Darstellung erkennen, wie das LQ angewendet wurde?
- (ii) Die Darstellung erklärt *nicht detailliert genug*, was genau mit Hilfe dieser Prozedur erreicht wird, nämlich:
 - (a) Über alle Personen hinweg sollte jedes Behaltensintervall (BI) an jeder Stelle innerhalb der Sequenz gleich oft vorkommen.
 - (b) Die Vp sollte nicht voraussehen können, welches BI an welcher Stelle der Sequenz auftaucht.
- (iii) Der Hinweis, dass der Positionseffekt sich vor allem als ein Übungseffekt bemerkbar machen könnte, ist zwar korrekt aber unverständlich, sofern nicht explizit auf die Ergebnisse von Peterson und Peterson (1959) verwiesen wird.

**Bsp. 5.3.3:**

Die Reihenfolge der Intervalle wurde durch ein lateinisches Quadrat ausbalanciert, indem die verschiedenen Anreihungen die dadurch entstanden, sowohl horizontal wie auch vertikal zu Blöcken zusammengefasst und den Vpn zugeteilt wurden.

Kritikpunkte:

Diese Darstellung hat folgende Probleme:

- (i) Was bedeutet „vertikal“ und „horizontal“? Als Leser kann ich mir nichts darunter vorstellen.
- (ii) Welches lateinische Quadrat liegt vor und wie kommt dieses zur Anwendung?

5.4 Unabhängige Variablen, abhängige Variablen und Versuchsplan (Experimentelles Design)

UV:**Prinzip:**

Die Beschreibung des Versuchsplans beinhaltet die folgenden Spezifikationen:

- ♦ Die Beschreibung der unabhängigen Variablen (UVn) enthält für jede unabhängige Variable (UV) die folgenden Spezifikationen:
 - (i) Name und Operationalisierung der UV: Es sollte ein prägnanter und leicht zu merkender Name gewählt und dieser in der Folge konsistent verwendet werden (z.B. bei der Beschreibung der Ergebnisse).
 - (ii) Die einzelnen Stufen der UV und deren Operationalisierung. Zur Benennung der Stufen der UVn sollten ebenfalls möglichst markante Namen verwendet werden.
- ♦ Die Art und Weise, wie die Stufen variierten: *Zwischen* den Vpn oder *innerhalb* der Vpn. Die Beschreibung, auf welche Art und Weise (zufällig, oder aufgrund bestimmter Kriterien?) Versuchspersonen zu den Gruppen zugeordnet wurden.
- ♦ Die Kombination der UVn: Wurden die UVn vollständig gekreuzt (was dem Idealfall entspricht) oder liegt ein (teil-) hierarchischer Plan vor, indem nur bestimmte Stufen von UVn kombiniert wurden.

AV:**Prinzip:**

- ♦ Die Beschreibung der abhängigen Variablen (AVn) erfolgt nach einem ähnlichen Prinzip wie jene der UVn: Jede abhängige Variable (AV) erhält einen Namen, der leicht im Gedächtnis haftet. Sodann wird dargestellt, wie die AV operationalisiert wurde.
- ♦ Bei komplexen Experimenten mit mehreren abhängigen Variablen (AVn) besteht die Möglichkeit, dass für unterschiedliche AVn verschiedene Kombinationen von UVn relevant sind. In diesem Falle werden die einzelnen Versuchspläne zu den einzelnen AVn separat beschrieben.
- ♦ Die Beschreibung der Abhängigen Variablen (AVn) dient zum besseren Verständnis, speziell bei Vorliegen mehrerer AVn. Die Beschreibung der AV umfasst die folgenden Aspekte:
 - ♦ Eine genaue Spezifikation der Messgrösse, auf der die nachfolgenden Analysen beruhen.
 - ♦ Das eventuelle Vorliegen und die Behandlung von Ausreissern.

- ◆ Das eventuelle Vorliegen und die Behandlung fehlender Werte: Sind die Ausfälle zufälliger oder systematischer Natur, oder waren sie sogar geplant? Wie wurde mit den Ausfällen verfahren?
- ◆ Möglichen Transformationen oder Zusammenfassungen der Rohwerte: Es muss genau dargestellt werden, wie die abgeleiteten bzw. transformierten Werte zustande kamen. Zusätzlich sind diese angewendeten Prozeduren zu begründen.
- ◆ Komplexere Analysen sollen in einem separaten Abschnitt zusätzlich ausgeführt werden (z.B. EEG Analysen, Modellierungen, etc.).

**Bsp. 5.4.1:**

Es liegt eine abhängige Variable vor, welche die Behaltensleistung der Buchstabenfolgen beschreibt.

Kritikpunkte:

- (i) Es wird die abstrakte abhängige Variable *Behaltensleistung* dargestellt. An dieser Stelle ist jedoch jene Grösse zu spezifizieren, welche Gegenstand der nachfolgenden Auswertung war. Im aktuellen Beispiel war dies die Anzahl korrekt wiedergegebener Buchstabenfolgen (und nicht – was ebenfalls denkbar wäre – der Prozentsatz korrekt reproduzierter Items).
- (ii) Die Ausdrucksweise ist leicht irreführend. Der Satz legt nämlich nahe, dass es die Buchstabenfolgen sind, welche eine Behaltensleistung besitzen (und nicht die Versuchspersonen).

Eine verbesserte Darstellung der abhängigen Variable könnte wie folgt lauten:

**Bsp. 5.4.2 (Beschreibung einer abhängigen Variable):**

Als abhängige Variable für die nachfolgende Auswertung diene die Anzahl korrekt reproduzierter Buchstabenfolgen. Die maximal mögliche Anzahl korrekt wiederzugebender Folgen betrug für jede Kombination von Stufen der beiden unabhängigen Variablen (Block und Behaltensintervall) 8.

**Bsp. 5.4.3 (Entfernen von Ausreissern und Transformation der Rohwerte):**

Für jede der 4 Testlisten ergaben sich pro Vpn die folgenden 3 Schätzungen: die geschätzte Sensitivität d' , die geschätzte Standardabweichung der ALT-Verteilung σ und das geschätzte Kriterium c . Damit wurden für jede der 120 Vpn 12 Messwerte ermittelt.

Von diesen 120 Datensätzen wurden jene 18 entfernt, für die mindestens eine der drei ermittelten Sensitivitäten einen Wert $d' > 3.0$ aufwies. Die Elimination dieser Daten lässt sich wie folgt begründen: Die erhöhten d' -Werte sind von geringer diagnostischer Bedeutung (Metz & Kronman, 1980). Zusätzlich verursachten sie eine signifikante Abweichung der Daten von der Normalverteilung (gemäss Kolmogorov-Smirnov-Test mit Lilliefors Anpassung). Die verbleibenden 102 Datensätze, welche für die weitere Analyse verwendet wurden, zeigten keine signifikante Abweichung von der Normalverteilung.

Da die Standardabweichungen σ nur positive Werte annehmen können, wurden sie einer logarithmischen Transformation (auf der Basis des natürlichen Logarithmus) unterzogen. Die transformierten Standardabweichungen konnten somit Werte im Bereich von $-\infty$ und $+\infty$ annehmen. Die Transformation eliminierte auch die für die untransformierten Daten (mittels Kolmogorov-Smirnov-Test) festgestellten Abweichungen von der Normalverteilung.



Bemerkung zur Bezeichnung von UV und AV:

Mehrfachbezeichnungen für UVn, AVn, sowie von Stufen der UVn sind zu vermeiden. Sie stiften nur Verwirrung und erhöhen die kognitive Belastung an die Leserin, da diese immer für sich abklären muss, welcher Begriff mit welchem anderen korrespondiert.

Versuchsplan:



Bsp. 5.4.4 (Beschreibung eines Versuchsplans):

Das Experiment umfasste drei unabhängige Variablen:

Erstens, das *Behaltensintervall (BI)* entsprach der Zeitdauer zwischen der Nennung der Zahl durch den VI und dem Signal zur Beendigung des Rückzählens. Es umfasste 6 Stufen: 3, 6, 9, 12, 15 und 18 Sekunden. Das Behaltensintervall variierte innerhalb der Vpn.

Zweitens, die *Ähnlichkeit des Stimulusmaterials* betraf die Anzahl unterschiedlicher Buchstaben, welche zur Konstruktion der Menge der Silben verwendet wurden und beinhaltete 2 Stufen. Die Menge der *ähnlichen Silben* umfasste Buchstabenkombinationen, die aus einer Grundmenge von 6 verschiedenen Konsonanten konstruiert wurden. Für die Menge der *unähnlichen Silben* wurden alle Konsonanten des Alphabets mit der oben genannten Einschränkung verwendet (Details, wie die Silben konstruiert wurden, finden sich in Beschreibung des Stimulusmaterials). Die Ähnlichkeit des Stimulusmaterials variierte zwischen den Vpn.

Drittens, die *Latenzzeit* entsprach der Zeitdauer zwischen der Beendigung des Rückzählens und dem Signal zur Wiedergabe der Silbe. Sie umfasste ebenfalls 2 Stufen: In der Bedingung *mit Latenzzeit* erhielt die Vp 6 Sekunden nach Ende des Zurückzählens das Signal zur Wiedergabe der Silbe. In dieser Zeit sollte sie versuchen, die wiederzugebende Silbe still für sich zu rekonstruieren. In der Bedingung *ohne Latenzzeit* entsprach das Signal zur Beendigung des Rückzählens der Aufforderung zur unmittelbaren Wiedergabe der Silbe. Die Latenzzeit wurde ebenfalls zwischen den Vpn variiert.

Die drei unabhängigen Variablen wurden vollständig gekreuzt. Somit ergab sich ein $6 \text{ (BI)} \times 2 \text{ (Ähnlichkeit)} \times 2 \text{ (Latenzzeit)}$ Versuchsplan, mit dem Wiederholungsfaktor BI und den beiden Gruppierungsfaktoren Ähnlichkeit und Latenzzeit.

5.5 Statistische Analyse

Hier sollte dargestellt werden, mit welchem statistischen Test welche Faktoren mit ihren Stufen angewendet werden. Hier soll auch angegeben werden, ob die Faktoren Innersubjekt- oder Zwischensubjektfaktoren sind.

Ebenso wird beschrieben, welche Korrekturen verwendet werden, falls Testvoraussetzungen nicht erfüllt werden. Auch das Alphaniveau wird in diesem Abschnitt definiert.

5.5.1 Darstellung von statistischen Prozeduren



Prinzip:

Die verwendeten Prozeduren sind so zu schildern, dass es möglich wird, das Verfahren genau nachzuvollziehen. Das Programm, welches für die Auswertung verwendet wurde, wird nicht erwähnt.

**Bsp. 5.5.1.1:**

Die Daten wurden mit einer dreifaktoriellen Varianzanalyse mit Messwiederholung ausgewertet. Dieses Verfahren hat ergeben, dass es einen signifikanten Unterschied...

Kritikpunkt:

Die Varianzanalyse ist zu ungenau spezifiziert. Es fehlt:

- (i) Welche UVn wurden einbezogen?
- (ii) Aus wie vielen und welchen Stufen bestanden die einzelnen UVn?
- (iii) Die Verständlichkeit wird erhöht, wenn anstelle von „Daten“ die AV konkret genannt wird.

**Bsp. 5.5.1.2:**

Zur Prüfung der Hypothesen wurden unsere Daten mit dem „General linear Model – repeated (sic) Measures“ ausgewertet.

Kritikpunkte:

- (i) Das General Linear Model ist kein konkretes Verfahren, sondern spezifiziert jene Klasse von Modellen, welche alle mit Hilfe einer multiplen multivariaten linearen Regression berechnet werden können. Es macht daher wenig Sinn, dieses Modell zu erwähnen (wobei die Autoren der obigen Passage vermutlich nicht einmal wissen, was der Begriff genau bedeutet).
- (ii) Wie in Bsp 5.5.1.1 wird aufgrund der Darstellung nicht klar, welche Prozedur konkret angewendet wurde.

**Bsp. 5.5.1.3 (Schilderung der Anwendung einer vierfaktoriellen Varianzanalyse):**

Die Anzahl korrekter Antworten wurde einer 2 (Stimulustyp) × 2 (Latenzzeit) × 6 (Behaltensintervall) × 4 (Block) – Varianzanalyse unterzogen. Hierbei gingen die Faktoren Stimulustyp und Latenzzeit als Gruppierungsvariablen in die Analyse ein. Das Behaltensintervall und der Block bildeten Wiederholungsfaktoren.

6. Ergebnisse

**Prinzip:**

Hier gibt es zwei mögliche Wege:

1. Die Deskriptive Statistik vor der Inferenzstatistik berichten um das Muster der Resultate zu beschreiben. Die Inferenzstatistik erfolgt danach und prüft zusätzlich, ob das Muster reliabel ist.
2. Ergebnisse der Inferenzstatistik und die deskriptive Statistik werden integriert präsentiert.

Beschrieben wird im Folgenden der zweitgenannte Weg:

Präsentation der inferenz-statistischen Analysen und Ergebnisse. Anschliessend deskriptiv und mit Post hoc Tests die Richtung der Effekte beschreiben.

- ♦ Es gilt im Allgemeinen, dass *statistische Hypothesen* (d.h. Aussagen über Populationsparameter, wie z.B. $\mu_1 = \mu_2$) innerhalb des experimentellen Berichtes nicht erwähnt werden sollten.

Begründung: Es gibt im Wesentlichen 2 Gründe für dieses Prinzip:

- (i) Die statistischen Analysen dienen der Beurteilung von *Sachhypothesen*. Die Erwähnung von statistischen Hypothesen ist daher eher störend.

- (ii) Die genaue Form der statistischen Hypothese hängt vom gewählten statistischen Verfahren ab: So ist die statistische Hypothese $\mu_1 = \mu_2$ im Kontext des t -Test äquivalent zur statistischen Hypothese $\alpha = 0$ im Kontext der Varianzanalyse (falls nicht das Mittelwertsmodell verwendet wird).

6.1 Darstellung von Variablen, Statistiken und Zahlen



Prinzip:

Bezüglich der Darstellung von Variablen, Stichprobenstatistiken, Zahlen etc. gibt es eine Reihe von Konventionen. Diese Konventionen betreffen das Auftreten innerhalb des gesamten Berichts. Sie werden aber in den meisten Fällen erst im Zuge der Darstellung der Ergebnisse relevant. Im Folgenden werden einige der wichtigsten Konventionen präsentiert:

- ◆ Größen, die Variablen bezeichnen – meist X , Y , etc. – werden in lateinischen Buchstaben und *kursiv* gesetzt. Das gleiche gilt für Stichprobenstatistiken – wie z.B. Stichprobengröße N , Mittelwert \bar{x} , Varianz s^2 etc. – sowie für Teststatistiken – wie z.B. z , t , F , χ^2 , p .
- ◆ Größen, die Populationsparameter symbolisieren, werden mit griechischen Buchstaben bezeichnet, z.B. Mittelwert μ und Varianz σ^2 innerhalb der Population (nicht in der Stichprobe!).
- ◆ In wissenschaftlicher Notation werden Dezimalstellen durch einen Punkt abgetrennt (und nicht durch ein Komma).
- ◆ Falls eine Zahl nur Werte im Bereich von -1 und 1 annehmen kann, so wird die Null vor dem Komma weggelassen – z.B. $p = .043$, $r = .30$.



Bemerkung zur Unterscheidung zwischen Stichprobenstatistiken und Populationsparameter

Stichprobenstatistiken sind Funktionen der Stichproben, d.h. jede Art von Grösse, die sich aus der Stichprobe ermitteln lässt. *Teststatistiken* sind besondere Arten von Stichprobenstatistiken, die zur Prüfung von statistischen Hypothesen verwendet werden.

Populationsparameter sind Größen, welche sich auf die Population beziehen. Ihre Werte sind meist nicht bekannt und können daher nur geschätzt oder – wie in statistischen Hypothesen – vermutet werden.

6.2 Inferenzstatistische Ergebnisse



Prinzip:

Der Zweck der statistischen Analyse der Daten besteht in der Prüfung bzw. Entscheidung, ob die im Rahmen der deskriptiven Statistik gefundenen Muster rein zufällig oder eher systematischer Natur sind. Die Darstellung dieser Ergebnisse folgt dem folgenden Schema: Für jede der analysierten AVn wird berichtet:

- ◆ Eine Darstellung der Ergebnisse unter Verwendung des im Methodenteil (statistische Analyse) ausführlicher beschriebenen statistischen Verfahrens.
- ◆ Eine Folgeanalyse der signifikanten Haupteffekte bzw Interaktionen mit Hilfe von Post Hoc Tests und deskriptiver Statistik.

Im Folgenden werden diese beiden Aspekte näher behandelt.

6.2.1 Darstellung von Ergebnissen statistischer Analysen

Bezüglich der Darstellung von Ergebnissen aus statistischer Analysen gilt:

Prinzip:



- ◆ Es sind immer alle Ergebnisse eines Verfahrens zu berichten, seien sie nun signifikant oder nicht. So sind bei Verwendung einer dreifaktoriellen Varianzanalyse (vollständig gekreuzt und ohne leere Zellen) folgende Ergebnisse zu berichten:
 - (i) Die Haupteffekte.
 - (ii) Die 2-weg Interaktionen für alle Paare von Faktoren.
 - (iii) Die 3-weg Interaktion.
- ◆ Im Falle signifikanter Haupteffekte sind Lage und Richtung der signifikanten Unterschiede zu spezifizieren, d.h. es muss angegeben werden, welche Mittelwerte oder Gruppen von Mittelwerten sich unterscheiden und welche grösser bzw. kleiner sind. Die genaue Lokalisierung von signifikanten Mittelwertsunterschieden erfolgt bei mehr als zwei Stufen der betreffenden unabhängigen Variablen mit Hilfe von Einzelvergleichen (siehe Bsp 6.2.1.1).
- ◆ Bei signifikanten Haupteffekten oder Interaktionen sind jeweils die Effektgrößen mit anzugeben (d bei t-Tests, η^2 (eta^2) bei ANOVAs).
- ◆ Bei Vorliegen signifikanter Interaktionen ist die genaue Konfiguration der in der Interaktion involvierten Mittelwerte zu spezifizieren. Die genaue Lokalisierung von signifikanten Unterschieden zwischen diesen Mittelwerten erfolgt wiederum mit Hilfe von Einzelvergleichen = Post Hoc Tests (siehe Bsp 6.2.1.2).
Zur Verdeutlichung des genauen Muster der Interaktion kann es sinnvoll sein, die relevanten Mittelwerte in einer Interaktionsgraphik darzustellen.

Die Darstellung der Ergebnisse bei Vorliegen von nicht signifikanten Ergebnissen lässt sich oft vereinfachen (siehe Abschnitt 6.2.3 und im Speziellen Bsp 6.2.3.1, sowie 6.2.3.2).



Bsp. 6.2.1.1 (Detaillierte Analyse eines Haupteffekts):

Aus der univariaten ANOVA mit dem 4-stufigen Faktor Lernliste ergab sich ein signifikanter Effekt auf die Wiedererkennungslleistung: $F(3, 303) = 29.27, p < .001, \eta^2 = .56$. Dieser Effekt beruhte auf einer erhöhten Wiedererkennungsrates für Liste E1 gegenüber den anderen 3 Lernlisten: $F(1, 303) = 85.52, p < .001, \eta^2 = .45$. Für die anderen drei Listen ergab sich kein Unterschied hinsichtlich der Wiedererkennungslleistung: $F(2, 303) = 1.28, p = .28$ (vergleiche Abb.2).



Bsp. 6.2.1.2 (Detaillierte Analyse einer 2-weg Interaktion):

Die Varianzanalyse erbrachte auch eine signifikante Latenzzeit \times Stimulus - Interaktion: $F(1, 44) = 15.19, p < .001, \eta^2 = .23$. Die signifikante Interaktion beruht auf der Tatsache, dass für die unähnlichen Silben die Latenzzeit einen hochsignifikanten Effekt aufwies ($F(1, 44) = 312.62, p < .001, \eta^2 = .33$), während für die ähnlichen Silben der Effekt der Latenzzeit nicht signifikant war: $F(1, 44) = 3.61, p = .064$ (vergleiche Abb-B. 2).

Bemerkung:



Eine Anleitung zur Interpretation von Interaktionen und zur Verwendung von Interaktionsgraphiken findet man in Anhang B.

Die Darstellung der Ergebnisse verschiedener statistischer Verfahren folgt ganz bestimmten Konventionen. Diese sind für die grundlegenden statistischen Verfahren im Abschnitt 6.2.3 zusammengestellt. Die nachfolgenden Abschnitte behandeln Spezialfälle.

6.2.2 Konventionen bezüglich der Darstellung von Ergebnissen der wichtigsten Statistischen Verfahren

Prinzip:



Die Präsentation der Ergebnisse zu den fünf gebräuchlichsten statistischen Tests hat den folgenden Richtlinien zu entsprechen:

1. Ergebnisse von **t-Tests** werden wie folgt berichtet:
 - (i) Das Symbol t zur Bezeichnung der t -Statistik wird geschrieben.
 - (ii) Es folgt in Klammern die Anzahl Freiheitsgrade.
 - (iii) Der errechnete Wert der t -Statistik wird präsentiert.
 - (iv) Am Ende wird die Wahrscheinlichkeit der beobachteten Teststatistik unter der Nullhypothese angegeben. Diese Angabe erfolgt im Allgemeinen in der Form: „ $p = \dots$ “. Falls der Wert von p kleiner als .001 ist, erfolgt die Angabe in der Form: „ $p < .001$ “.



Bsp. 6.2.2.1 (Bericht von p Werten bei t -Tests):

- ♦ Es ergab sich ein signifikanter Geschlechtseffekt: $t(54) = 2.01$, $p = .049$ (zweiseitig). Hierbei wiesen die Männer höhere Werte auf als die Frauen.
- ♦ Die Wiedergabeleistung war im ersten Lern-Test-Block signifikant höher als im zweiten: $t(64) = 5.48$, $p < .001$.

2. Ergebnisse von **F-Tests** werden wie folgt berichtet:
 - (i) Das Symbol F zur Bezeichnung der F -Statistik wird geschrieben.
 - (ii) Es folgt in Klammern durch ein Komma separiert die Anzahl der Zähler- und Nenner-Freiheitsgrade.
 - (iii) Der errechnete Wert der F -Statistik wird präsentiert.
 - (iv) Am Ende wird die Wahrscheinlichkeit der beobachteten Teststatistik unter der Nullhypothese angegeben. Diese Angabe erfolgt im Allgemeinen in der Form: „ $p = \dots$ “. Falls der Wert von p kleiner als .001 ist, erfolgt die Angabe in der Form: „ $p < .001$ “.



Bsp. 6.2.2.2 (Bericht von p Werten bei F -Tests):

- ♦ Es gab einen signifikanten Effekt der Latenzzeit: $F(1, 145) = 12.45$, $p < .001$.
- ♦ Der Haupteffekt des Stimulustyps war hingegen nicht signifikant: $F < 1$.
- ♦ Die Interaktion zwischen Latenzzeit und Stimulustyp erreichte jedoch Signifikanz: $F(2, 145) = 3.43$, $p = .035$.



Bemerkung zur Darstellung von Mittelwertsvergleichen:

Mittelwertsvergleiche (bzw. Kontraste im Allgemeinen) werden entweder mittels F -Test geprüft, wobei die Anzahl der Zähler-Freiheitsgrade gleich 1 ist, oder – äquivalent – mittels t -Test (Ein F -Test mit einem Zähler-Freiheitsgrad entspricht einem zweiseitigen t -Test, wobei der t -Wert gleich der Wurzel aus dem F -Wert ist). Je nach Wahl der Teststatistik folgt demgemäß die Darstellung der Ergebnisse von Mittelwertsvergleichen jener eines t - bzw. F -Tests.



Bemerkung zur Darstellung von Verfahren zur α -Fehler-Korrektur:

Unter bestimmten Bedingungen ist es sinnvoll eine Korrektur des α -Fehlers vorzunehmen, um einer Inflation des α -Fehlers aufgrund der gehäuften Durchführung von Signifikanztests vorzubeugen. Die gebräuchlichsten Verfahren zur α -Fehler-Korrektur sind hierbei die Methoden von Bonferroni, Tukey und Scheffé. Falls eine Methode zur Korrektur des α -Fehlers verwendet wurde, so ist diese zu berichten.

Falls eine solche verwendet wird, muss diese auch im Methodenteil Abschnitt statistische Analyse erwähnt werden.



Bsp. 6.2.2.3 (Bericht von α -Fehler Korrekturen):

Es wurde ein Vergleich der Mittelwerte der jeweils benachbarten Stufen der UV *Behaltensintervall* durchgeführt. Zur Korrektur des α -Fehlers wurde die Methode von Bonferroni verwendet.

3. Ergebnisse zu **Korrelationen** werden wie folgt berichtet:

- (i) Das Symbol r zur Bezeichnung der Stichprobenstatistik der Korrelation wird geschrieben.
- (ii) Es folgt in Klammern die Stichprobengröße N .
- (iii) Am Ende wird die Wahrscheinlichkeit der berechneten Teststatistik unter der Nullhypothese angegeben. Diese Angabe erfolgt im Allgemeinen in der Form: „ $p = \dots$ “. Falls der Wert von p kleiner als .001 ist, erfolgt die Angabe in der Form: „ $p < .001$ “.



Bsp. 6.2.2.4 (Bericht von p -Werten für Korrelationskoeffizienten):

Die Geschwindigkeit des Rückzählens und die Behaltensleistung wiesen eine positive Korrelation auf: $r(N = 44) = .32, p = .017$.

4. Ergebnisse einer (**einfachen**) **linearen Regression** werden wie folgt berichtet:

- (i) Das Symbol b zur Bezeichnung der Stichprobenstatistik des Regressionskoeffizienten wird geschrieben.
- (ii) Es folgt der Wert des Regressionskoeffizienten.
- (iii) Das Ergebnis des t -Tests zur Prüfung des Koeffizienten wird – wie oben dargestellt – präsentiert.



Bsp. 6.2.2.5 (Bericht von p -Werten für Regressionskoeffizienten):

Der Grad der Depression nahm mit der Zunahme der sozialen Unterstützung ab: $b = -.34, t(225) = 2.04, p = .033$.

5. Ergebnisse von **Chi-Quadrat Tests** werden wie folgt berichtet:

- (i) Das Symbol X^2 zur Bezeichnung der Chi-Quadrat Statistik (d.h. der Chi-Quadrat verteilten Statistik) wird geschrieben.
- (ii) Es folgen in Klammern die Anzahl Freiheitsgrade sowie die Stichprobengröße. Letztere wird durch das Symbol N bezeichnet.
- (iii) Der errechnete Wert der Chi-Quadrat Statistik wird präsentiert.
- (iv) Am Ende wird die Wahrscheinlichkeit der beobachteten Teststatistik unter der Nullhypothese angegeben. Diese Angabe erfolgt im Allgemeinen in der Form: „ $p = \dots$ “. Falls der Wert von p kleiner als .001 ist, erfolgt die Angabe in der Form: „ $p < .001$ “.



Bsp. 6.2.2.6 (Bericht von p -Werten für Chi-Quadrat Tests):

Der Anteil der verheirateten Teilnehmer war für beide Geschlechter identisch: $\chi^2(1, N = 90) = 0.89, p = .35$.



Bemerkung:

Für den Chi-Quadrat Test verwendet man oft als Symbol für die Teststatistik anstelle von X^2 das griechische Zeichen χ^2 (nicht kursiv). Dies hat sich so eingebürgert, obwohl es nicht den allgemeinen Richtlinien für die Verwendung von Symbolen entspricht (Siehe Abschnitt 6.1).

6.2.3 Beschreibung von absolut insignifikanten Ergebnissen



Prinzip:

1. Falls der F -Wert (im Rahmen einer Varianzanalyse) für einen Effekt oder eine Interaktion kleiner als 1 ist, so wird einfach $F < 1$ geschrieben, ohne Angabe von Freiheitsgraden oder p -Wert (Vgl. Bsp 6.2.2.2).
2. Falls ein Grossteil der Ergebnisse einer Varianzanalyse nicht signifikant ist, so wird nicht jedes insignifikante Ergebnis einzeln berichtet, sondern es wird erwähnt, dass die restlichen Ergebnisse nicht signifikant sind und entweder der kleinste p -Wert angegeben (siehe Bsp 6.2.3.1), oder aber – für die zutreffenden Fälle – $F < 1$ (siehe Bsp 6.2.3.2).

Hierbei sind jedoch nicht signifikante Ergebnisse, welche von direkter Relevanz für die Sachhypothesen sind, ausgenommen: Ergebnisse, welche eine Sachhypothese direkt betreffen, sind immer einzeln aufzuführen, seien sie nun signifikant oder nicht (siehe Bsp 6.2.4.1).



Bsp.6.2.3.1:

Die Ergebnisse für den kubischen ($F(1,55)=1.796$, $p=.186$) und quartischen ($F(1,55)=1.675$, $p=.201$) Trend waren nicht signifikant.

Kommentar:

Falls alle Ergebnisse – wie im aktuellen Fall – weit von einem signifikanten Resultat entfernt sind, so fasst man die Ergebnisse – auf ungefähr die folgende Weise zusammen: „Die restlichen Trends waren nicht signifikant (Alle $p>.18$).“ Man wählt also den kleinsten gemeinsamen p -Wert und berichtet diesen.



Bsp.6.2.3.2:

Alle anderen Behaltensintervalle unterschieden sich nicht signifikant. Für das 6 und 9 Sekunden Behaltensintervall lautete das Resultat ($F(1,55)=0.092$, $p=.763$), für das 9 und 12 Sekunden Behaltensintervall ($F(1,55)=0.573$, $p=.452$), für das 12 und 15 Sekunden Behaltensintervall ($F(1,55)=0.00$, $p=1$) und das Resultat für das 15 und 18 Sekunden Behaltensintervall belief sich auf ($F(1,55)=0.206$, $p=.652$).

Kommentar:

Falls alle $F < 1$ sind, so lässt sich dies abkürzen, indem man z.B. schreibt: „Alle anderen Behaltensintervalle unterschieden sich nicht signifikant (Alle $F < 1$).“



Bsp.6.2.3.3:

Alle anderen Werte wurden nicht signifikant (Alle $p > .055$).

Kommentar:

Der Wert $p > .055$ weist darauf hin, dass mindestens ein Ergebnis knapp nicht signifikant ist. Derartige Ergebnisse fallen nicht unter die Kategorie *absolut insignifikantes Ergebnis* sondern unter die Kategorie *Trend* (Zur Darstellung von Trends, vergleiche Abschnitt 6.2.4).



Bsp. 6.2.3.4 (Bericht von nicht signifikanten Ergebnissen, welche für die Sachhypothese relevant sind):

Gegeben sei die folgende Situation:

- ◆ Es wurde eine Varianzanalyse mit den Faktoren *Block*, *Präsentation* und *Behaltensintervall* durchgeführt.
- ◆ Keine der drei 2-weg Interaktionen war signifikant. Die 3-weg Interaktion war hingegen signifikant.

- ♦ Gemäss Sachhypothese sollte der Faktor *Präsentation* den Effekt des Behaltensintervalls moderieren (indem bei wiederholter Präsentation der Effekt des Behaltensintervalls reduziert ist). Die Sachhypothese sagt daher eine Interaktion zwischen Behaltensintervall und Präsentation vorher.

Darstellung der nicht signifikanten Ergebnisse:

Die Interaktion zwischen dem Behaltensintervall und der Präsentation war nicht signifikant: $F(2, 34) = 2.14$, $p = .13$. Unsere Hypothese wurde somit nicht bestätigt. Die beiden anderen 2-weg Interaktionen waren ebenfalls nicht signifikant ($p > .23$).

Bemerkung:

Die Behaltensintervall \times Präsentation Interaktion ist von direkter Bedeutung für die Sachhypothese. Daher wird das nicht signifikante Ergebnis separat und vollständig präsentiert. Die Ergebnisse, welche die anderen, nicht signifikanten Interaktionen betreffen, werden nur zusammenfassend dargestellt, da letztere für die Sachypothesen irrelevant sind.

6.2.4 Beschreibung von knapp nicht signifikanten Ergebnissen



Prinzip:

- ♦ Falls ein signifikanter Wert nur knapp verfehlt wurde, so ist dies deutlich zu machen.
- ♦ Falls $p < .10$ so kann man von einem *Trend* in eine bestimmte Richtung sprechen.



Bsp.6.2.4.1:

Der Test der Within-Subjekt Kontraste ergab weiter, dass zwischen der ersten und zweiten Stufe keine signifikante Gedächtnisverschlechterung existierte ($F(1,11)=4$, $p>.05$). Das gleiche galt für die zweite und dritte ($F(1,11)=4.231$, $p>.05$) (...)

Die Signifikanz wurde in beiden Fällen nur knapp verfehlt. Es ist wichtig, dies deutlich zu machen.



Bsp. 6.2.4.2 (Bericht von knapp nicht signifikanten Ergebnissen):

Der Unterschied in der Behaltensleistung zwischen der zweiten und dritten Stufe des Behaltensintervalls verfehlt nur knapp einen signifikanten Wert: $F(1, 11) = 4.23$, $p = .064$.

6.2.5 Intuitive Prüfung von inferenzstatistischen Ergebnissen



Prinzip:

Es ist sinnvoll, zu prüfen, ob ein gefundenes Ergebnis plausibel ist. Dies gilt vor allem im Falle von Berechnungen, die nicht mit einem Standardstatistikprogramm wie SPSS durchgeführt wurden. Man prüfe daher, ob die Ergebnisse der statistischen Analyse zum qualitativen Muster der Ergebnisse passen.



Bsp.6.2.5.1:

Es zeigte sich, dass drei von fünf Kontrasten Signifikanzen aufwiesen: (...) zwischen 12 und 15 Sekunden ($F(1,55)=27.19$, $p<.01$). Die anderen beiden waren nicht signifikant (...) und zwischen 15 und 18 Sekunden: $F(1,55)=0.34$, $p>.05$).

Kommentar:

Die Mittelwerte für die Behaltensintervalle von 12, 15 und 18 Sekunden betrugen (nach Angaben der Autoren): 2, 2.33 und 2. Die Anzahl Beobachtungen war für alle drei Behaltensintervalle identisch ($N = 12$).

Die Differenz der Mittelwerte zwischen 12 und 15 Sekunden und zwischen 15 und 18 Sekunden war daher identisch (0.33). Da mit gepoolten Kontrasten gerechnet wurde

(d.h. es wurde der gleiche Fehlerterm für die Berechnung aller Kontraste verwendet), muss sich daher für beide Kontraste *exakt* das gleiche Resultat ergeben. Somit es ist es unmöglich, dass die Mittelwertsdifferenz zwischen 12 und 15 Sekunden hoch signifikant war, während jene zwischen 15 und 18 Sekunden weit von der Signifikanz entfernt war.

6.3 Deskriptive Statistik



Prinzip:

Das Ziel der deskriptiven Statistik besteht darin, der Leserin das *zentrale Ergebnismuster* im Hinblick auf die Hypothese bzw. das Untersuchungsziel zu enthüllen. Für diese Darstellung bieten sich als Mittel Tabellen und Graphiken an. Zusätzlich zu diesen ist das *qualitative* Muster der Ergebnisse im Text zu beschreiben. Dies bedeutet, dass die Ergebnisse durch Verwendung von Ausdrücken wie *grösser/kleiner*, *häufiger/weniger* häufig etc. beschrieben werden. Zusätzlich wird berichtet, ob und inwieweit die qualitativen Resultate den Vorhersagen entsprechen. Die Darstellung sollte die folgenden Fehler vermeiden:

- ◆ Es ist wichtig, sich nicht in Details zu verlieren, damit das Wesentliche vom Unwesentlichen unterschieden werden kann. Im Speziellen sollten keine Unterschiede, welche sich in den vorangegangenen Tests als nicht signifikant erweisen, interpretiert werden.
- ◆ Das Berichten der Standardabweichungen in Tabellen ist fast immer uninformativ. Viel mehr Sinn macht es, einen oder zwei Standardfehler für die Mittelwerte (oder Konfidenzintervalle) zu präsentieren. In Graphiken können Standardfehler(SEM)/Konfidenzintervalle mit Hilfe von Fehlerbalken dargestellt werden (siehe Bsp 6.3.1).



Bsp. 6.3.1 (Darstellung von Konfidenzintervallen mittels Fehlerbalken):

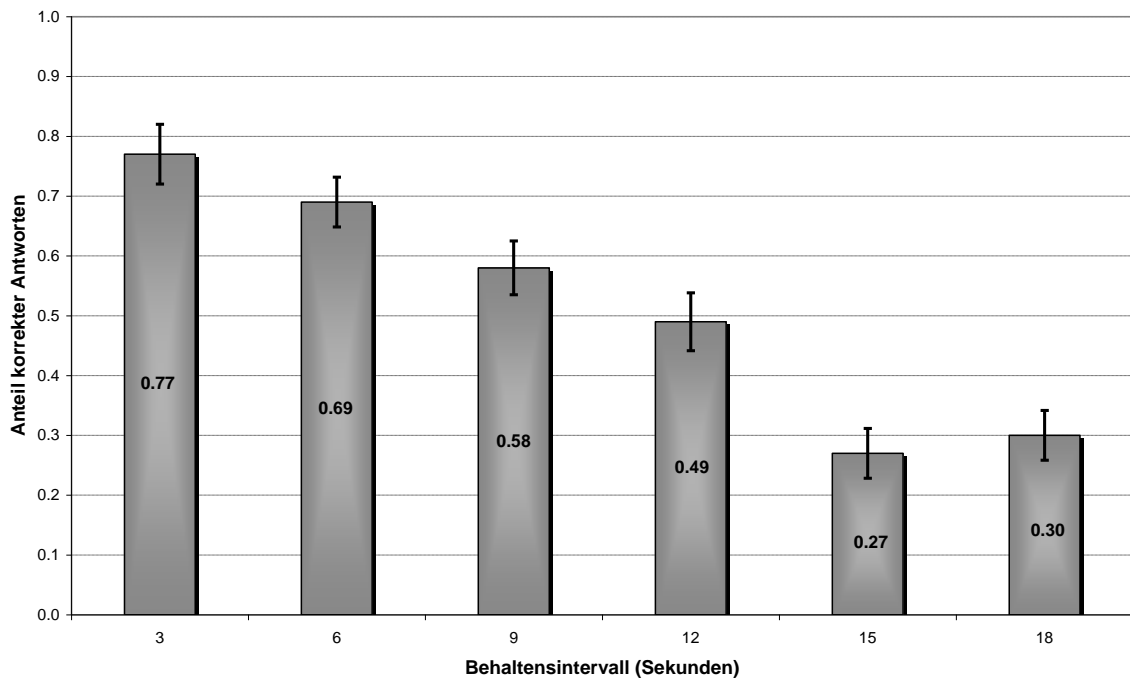


Abb. 1: Anteil korrekter Antworten in Abhängigkeit von der Länge des Behaltensintervalls. Die Fehlerbalken repräsentieren zwei Standardfehler der Mittelwerte.

Die Fehlerbalken in Abb. 1 repräsentieren zwei Standardfehler des Mittelwerts (d.h. im aktuellen Fall des Prozentsatzes korrekter Antworten pro Behaltensintervall). Mit Hilfe der Fehlerbalken erkennt man auf einen Blick, wo sich signifikante Unterschiede befinden: Falls sich die Fehlerbalken nicht überlappen, so ist der Unterschied zwischen den Mittelwerten signifikant (bei einem SEM als Fehlerbalken: vermutlich signifikant). Beispielsweise legt Abb. 1 nahe, dass zwischen den Behaltensintervallen 3 und 6 Sekunden kein signifikanter Unterschied besteht, wohl aber zwischen dem 6- und 9-Sekunden Behaltensintervall.

Für nähere Informationen zur Interpretation der Abstände von Fehlerbalken, siehe Cumming & Finch, 2005.



Bemerkungen:

- (i) Da es sich um Anteile handelt, wurde ein Balkendiagramm verwendet. Fehlerbalken lassen sich auch für andere Graphiken erstellen (siehe Bsp 6.3.2).
- (ii) In die Balken wurden die genauen Werte eingeschrieben. Dies erspart die Erstellung einer Mittelwertstabelle (zusätzlich zur Graphik).
- (iii) *Zur Erinnerung:* Der Standardfehler des Mittelwerts entspricht der geschätzten Standardabweichung des Mittelwerts und ergibt sich durch:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

Die Symbole haben hierbei die folgende Bedeutung:

$\sigma_{\bar{x}}$ bezeichnet den Standardfehler des Mittelwerts.

- σ symbolisiert die Standardabweichung der Rohwerte, aus der der Mittelwert berechnet wurde.
 N ist die Stichprobengrösse.



Bsp. 6.3.2 (Vergleich von Excel und SPSS Fehlerbalken-Diagrammen):

Balkendiagramme lassen sich sowohl mit Excel oder auch mit SPSS erstellen. Abb. 2 zeigt einen Vergleich der beiden Graphiken. Die Excel-Graphik (rechte Graphik) ist (vielleicht) optisch etwas ansprechender als jene von SPSS. Die Erstellung der SPSS-Graphik erfordert jedoch einen geringeren Aufwand und ist weniger fehlerbehaftet.

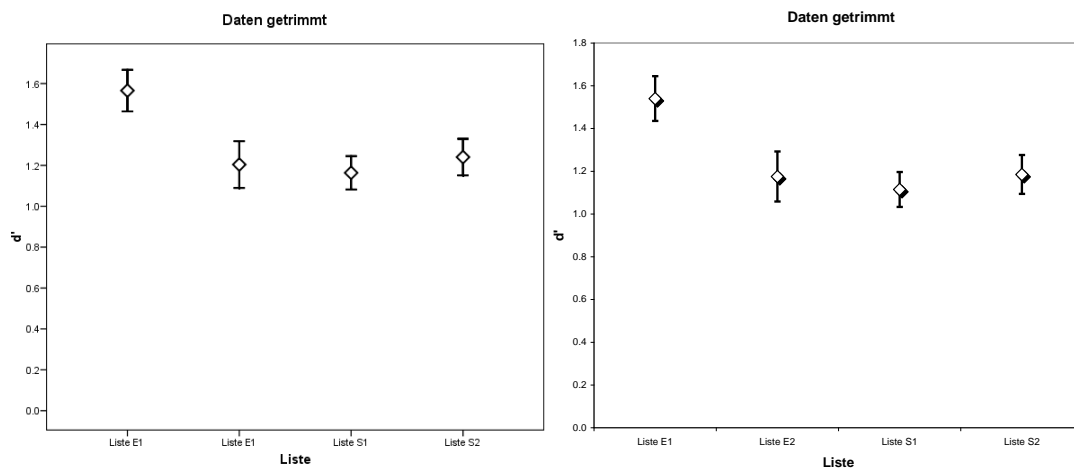


Abb. 2: Vergleich eines SPSS (linke Graphik) und Excel Diagramms mit Fehlerbalken.



Bemerkung:

In vielen Fällen ist es sinnvoll die Struktur der Rohdaten darzustellen um festzustellen ob diese annähernd symmetrisch verteilt sind und ob es eine grössere Anzahl von Ausreissern und Extremwerten gibt. Zu diesem Zweck eignen sich so genannte Boxplots. Eine Darstellung und Erklärung von Boxplots findet sich im Anhang A.

6.3.1 Verwendete Anzahl Dezimalstellen in Tabellen und Text



Prinzip:

- ◆ Es ist sinnvoll, die Anzahl Nachkommastellen einheitlich festzulegen bzw. zu verwenden (sowohl im Text, wie in Tabellen).
- ◆ Es ist sinnvoll, wenige Dezimalstellen zu verwenden, da dies das Verständnis erleichtert. Allerdings dürfen durch die Unterdrückung von Dezimalstellen keine bedeutsame Unterschiede verloren gehen.



Bsp.6.3.1.1:

Mittelwert /	Behaltensintervall					
Anzahl Vpn	3	6	9	12	15	18
\bar{x}	6.5	4.5	4.083	2	2.33	2
N	12	12	12	12	12	12

Kritikpunkte:

Die gezeigte Tabelle weist mehrere Probleme auf:

- (i) Die Anzahl Dezimalstellen sind uneinheitlich.
- (ii) Es wurden relevante Dezimalstellen unterdrückt. Dadurch ergibt sich der Eindruck, dass der Abstand der Mittelwerte zwischen Behaltensintervall 12 und 15 exakt dem zwischen 15 und 18 entspricht. Dies ist jedoch nicht der Fall.
- (iii) Tabellennummer und –überschrift fehlen.

Korrekt sieht die Tabelle wie folgt aus:



Bsp. 6.3.1.2 (Korrektter Aufbau einer Tabelle):

Tabelle 1. *Mittlere Erinnerungsleistung je Behaltensintervall*

Mittelwert /	Behaltensintervall					
Anzahl Vpn	3	6	9	12	15	18
\bar{x}	6.50	4.50	4.08	2.08	2.33	2.00
<i>SEM</i>	0.01	0.02	0.07	0.04	0.04	0.06
<i>N</i>	12	12	12	12	12	12

Wie bereits erwähnt sind Tabellen und Graphiken wichtige Mittel um Ergebnisse im Überblick dazustellen. Hierbei sind bestimmte Richtlinien zu beachten, von denen einige im Folgenden besprochen werden.

6.3.2 Allgemeine Richtlinien zur Gestaltung von Tabellen



Prinzip:

Tabellen umfassen vier Komponenten:

- (i) Die Tabellennummerierung: Bezeichnung der Tabelle mit fortlaufender Nummerierung (z.B. Tab.1, Tab.2, ... oder Tabelle 1, Tabelle 2, ...). Im Text wird mit Hilfe dieser Bezeichnung auf die Tabelle verwiesen.
- (ii) In der folgenden Zeile ein Titel/ eine Kurzbeschreibung der Tabelle. Dieser Text befindet sich **oberhalb** der Tabelle, wird kursiv geschrieben und endet ohne Satzzeichen. Es ist darauf zu achten, dass Tabellentext und Tabelle nicht durch einen Seitenumbruch getrennt werden.
- (iii) Unter der Tabelle können Notes. angegeben werden. Hier werden Abkürzungen und Symbole erklärt (z.B. * kennzeichnet Werte $p < .05$), welche innerhalb der Tabelle verwendet werden.
- (iv) Die eigentliche Tabelle ist durch 2 horizontale Linien (oberhalb und unterhalb der Tabelle) begrenzt und besteht aus 2 Teilen:
 - (a) Der *Kopf der Tabelle* dient zur Beschreibung der Spalten der Tabelle. Grundsätzlich ist es wünschenswert, jede Spalte zu bezeichnen. In Ausnahmefällen kann es vorkommen, dass dies nicht möglich ist, aufgrund des heterogenen Inhalts einer Spalte. In diesem Fall entfällt die Spaltenbezeichnung. Zusammengehörige Spalten erhalten eine übergeordnete Überschrift. Diese wird durch eine horizontale Linie, welche sich oberhalb der zu gruppierenden Spalten befindet und diese genau überdeckt, von den untergeordneten Spaltenüberschriften getrennt. Auf diese Weise können

Hierarchien von Gruppen von Spalten gebildet werden (siehe Bsp 6.3.2.1).

Der Tabellenkopf ist durch eine horizontale Linie vom Tabelleninhalt getrennt.

- (b) Der *Tabellenkörper* enthält einerseits Spalten mit Bezeichnungen der einzelnen Zeilen der Tabelle und andererseits die eigentlichen Daten. Innerhalb des Tabellenkörpers ist jegliche Art von Trennstrichen (horizontal oder vertikal) zu vermeiden. Zur Verdeutlichung von Gruppierungen von Zeilen oder Spalten verwendet man unterschiedliche Abstände.
- (c) Innerhalb des Tabellenkörpers können Subüberschriften verwendet werden (siehe Bsp 6.3.2.1).

Die Tabelle wird gewöhnlich zentriert dargestellt. Es sollte auch – wenn möglich – darauf geachtet werden, dass die Tabelle nicht durch einen Seitenumbruch auseinander gerissen wird.



Bsp. 6.3.2.1 (Aufbau von Tabellen):

Tabelle 1

Sensitivität (d') und Standardabweichung der Alt-Verteilung (σ) in Abhängigkeit von der Studierzeit, der Testmethode, der verwendeten Ratingskala und der Reihenfolge

		Skala-6		Skala-7	
		Ranking	Rating	Ranking	Rating
Studierzeit	Kennwert				
Ranking zuerst					
1 sec	d'	1.00 ± 0.22	0.89 ± 0.20	0.77 ± 0.19	0.61 ± 0.16
	σ	1.52 ± 0.28	1.39 ± 0.27	1.23 ± 0.12	1.13 ± 0.22
3 sec	d'	0.89 ± 0.22	0.85 ± 0.26	1.56 ± 0.34	1.40 ± 0.25
	σ	1.28 ± 0.33	1.24 ± 0.20	1.54 ± 0.36	1.40 ± 0.21
Rating zuerst					
1 sec	d'	0.70 ± 0.22	0.45 ± 0.62	0.62 ± 0.29	0.33 ± 0.15
	σ	1.08 ± 0.28	1.09 ± 0.22	1.22 ± 0.32	1.09 ± 0.31
3 sec	d'	1.03 ± 0.33	0.52 ± 0.23	1.23 ± 0.13	0.68 ± 0.37
	σ	1.41 ± 0.19	1.00 ± 0.23	1.47 ± 0.35	1.15 ± 0.39

Notes. d' = Sensitivität, σ = Standardabweichung. Es werden die Mittelwerte \pm Standardfehler des Mittelwerts (SEM) angegeben.

**Bemerkung:**

Gemäss unserer Erfahrung eignet sich das Tabellenprogramm von Word hervorragend zur Gestaltung von Tabellen. Wir empfehlen daher dessen Verwendung.

6.3.3 Allgemeine Richtlinien zur Gestaltung von Abbildungen

Prinzip:

- ♦ Analog zu den Tabellen besitzt eine Abbildung einen Text, welcher alle Komponenten der Abbildung erklärt und dem eine Bezeichnung der Abbildung mit fortlaufender Nummerierung (z.B. Abb.1, Abb.2, etc.) vorangestellt ist. Auch dieser Text befindet sich unterhalb der Abbildung. Der Inhalt einer Abbildung ist aber auch – mehr oder weniger detailliert – im Text zu erwähnen.
- ♦ Ein wichtiger Aspekt (neben anderen, wie z.B. die Wahl des Typs einer Graphik, oder dem Problem der Komplexität, welche hier nicht behandelt werden) stellt die Wahl des geeigneten Achsenbereiches dar: Es ist wichtig, den Bereich für die Achsen derart zu wählen, dass Effekte nicht verzerrt dargestellt werden (siehe hierzu Bsp 6.3.3.1).
- ♦ Werden mehrere ähnliche Grafiken erstellt, sollte darauf geachtet werden, dass die Achsenskalierung nach Möglichkeit gleich gewählt wird (gleicher Range).
- ♦ Schliesslich ist darauf zu achten, dass Abbildungstext und Abbildung nicht durch einen Seitenumbruch getrennt werden, oder dass die Abbildung nur teilweise sichtbar ist (wenn sie z.B. am Ende einer Seite platziert wird).

**Bsp.6.3.3.1 (Verzerrte Darstellung von Effekten):**

Abbildung 3 und

zeigen die gleichen Daten. In Abbildung 3 wird eine eingeschränkte Skala (y-Achse) verwendet, welche den Eindruck einer Interaktion zwischen der *Ähnlichkeit der Items* und dem *Experimentalblock* vermittelt: Für ähnliche Items nimmt demnach die Leistung im Laufe des Experiments (von Block 1 zu Block 4) zu, während für unähnliche Items die Leistung ungefähr gleich bleibt.

Abbildung 3. Darstellung des Blockeffekts auf ähnliche und unähnliche Items

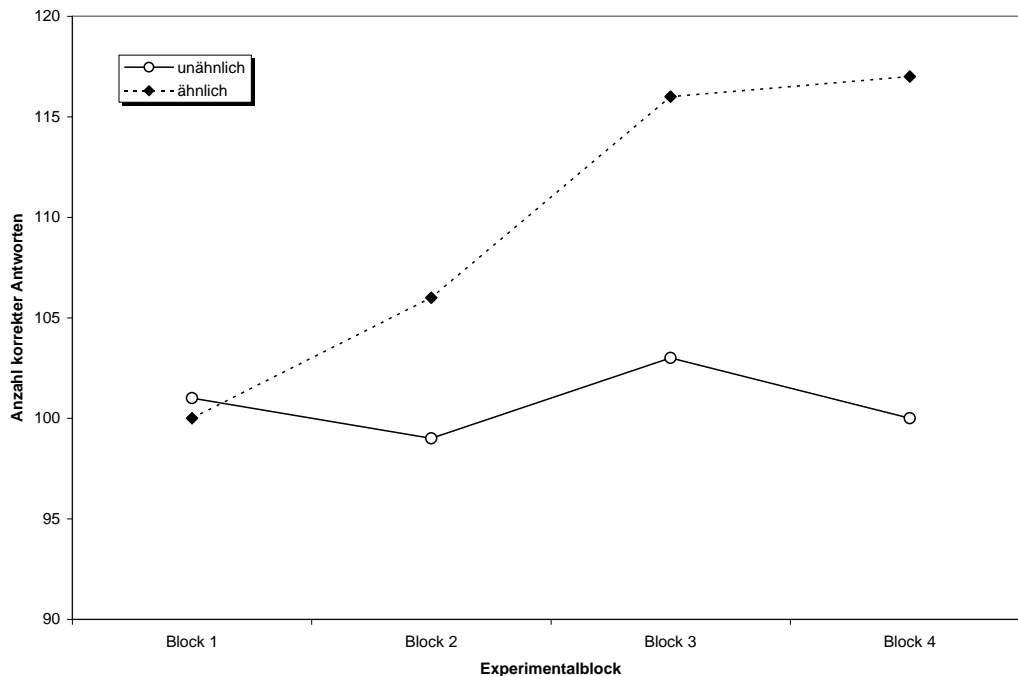


Abbildung 3. Stellt den differentiellen Blockeffekt auf ähnliche (gestrichelte Linie) und unähnliche (durchgezogene Linie) Items dar.

Während Abbildung 3 die Darstellung durch die gewählte Skalierung verzerrt, zeigt eine ehrlichere Darstellung: Auch hier zeigt sich eine leichte Verbesserung für ähnliche Items, jedoch erscheint dieser Effekt äusserst gering. Dies wird durch das Ergebnis aus der Varianzanalyse bestätigt, wonach die *Ähnlichkeit* \times *Experimentalblock* Interaktion weit davon entfernt ist, signifikant zu sein ($F < 1$). Der irrige Eindruck in Abbildung 3 kann durch die Verwendung von Fehlerbalken vermieden werden (Zur Bedeutung von Fehlerbalken, siehe Abschnitt 6.3, sowie Bsp 6.3.1 und Bsp 6.3.2).

Abbildung 4. Darstellung des Blockeffekts auf ähnliche und unähnliche Items

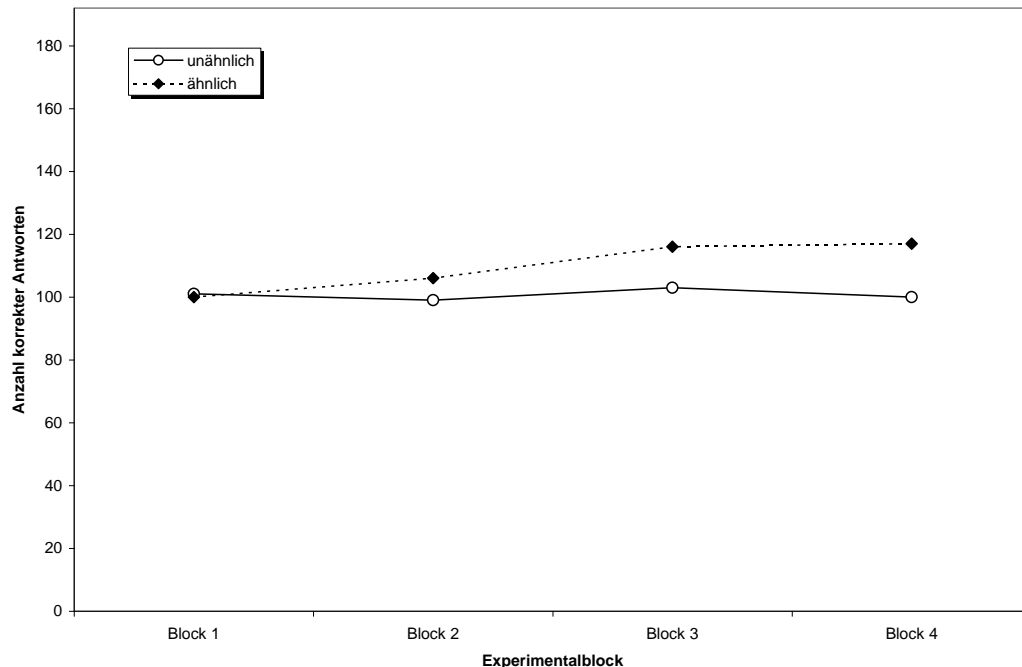


Abbildung 4. Korrekte Darstellung des differentiellen Blockeffekts auf ähnliche (getrichelte Linie) und unähnliche (durchgezogene Linie) Items.

7. Diskussion und Ausblick



Prinzip:

Das Ziel der Diskussion besteht darin, die Bedeutung der Ergebnisse im Hinblick auf die zu prüfende Theorie und die aufgestellten Hypothesen darzustellen. Der Ausblick sollte auf mögliche neue untersuchenswerte Fragestellungen hinweisen.

Der Abschnitt sollte wie folgt gegliedert sein:

- ◆ Zuerst wird ein kurzer Überblick das Ziel der Untersuchung betreffend gegeben.
- ◆ Hierauf folgen eine Zusammenfassung der zentralen Ergebnisse (ohne Zahlen), sowie eine Interpretation, inwiefern die Ergebnisse die Hypothesen stützen. Falls die Ergebnisse die Hypothesen nicht stützen, oder die Ergebnisse bekannten Resultaten aus der Literatur widersprechen, so sollten die Gründe hierfür diskutiert werden: Ist die Theorie falsch oder war die Methode fehlerhaft (z.B. zu kleine Stichprobe und der daraus resultierenden geringen Teststärke, oder unbeachtete Störvariablen).
- ◆ Der Ausblick sollte schliesslich Hinweise auf mögliche Nachfolgeuntersuchungen geben.



Bsp. 7.1 (Struktur des Diskussionsteils):

Das gegenwärtige Experiment untersuchte den Einfluss wiederholter Abrufversuche auf die Behaltensleistung von Konsonantentripel. Zusätzlich wurde der Effekt der Länge des Behaltensintervalls auf die Behaltensleistung getestet.

[...]

Die Ergebnisse bestätigen die zentrale Hypothese des Experiments, wonach wiederholte Versuche, das zu Erinnernde aus dem Gedächtnis abzurufen, die Behaltensleistung erhöhen: Vpn, welche instruiert wurden, nach Ende des Behaltensintervalls (BI) für 6 sec die Konsonanten verdeckt abzurufen, erzielten eine höhere Leistung gegenüber Vpn, die unmittelbar nach Ende des BI die zu erinnernden Konsonanten wiedergeben sollten.

[...]

Die Ergebnisse weisen darauf hin, dass Personen beim Versuch Information aus dem Gedächtnis abzurufen, oft vorzeitig den Informationsabruf abbrechen und daher eine nicht optimale Wiedergabeleistung aufweisen.

[...]

Die nicht optimale Abrufleistung durch vorzeitigen Abbruch der Informationssuche im Gedächtnis könnte durch einen Zielkonflikt erklärt werden, indem im realen Leben Information, welche zu spät verfügbar ist, oft an Wert verliert. Es ist daher evolutionär durchaus sinnvoll, den Informationsabruf zu begrenzen. Im Falle eines gut funktionierenden Gedächtnisses wird dies zu keinen nennenswerten Verlusten führen und gleichzeitig den Aufwand der Suche im Gedächtnis verringern. Ob sich diese Erklärung als empirisch tragbar erweist, muss die zukünftige Forschung zeigen.

Im Zuge der Diskussion werden oft unbewiesene Spekulationen und fadenscheinige Argumente (der Form „Es könnte sein, dass ...“) präsentiert. Derartiges ist zu vermeiden (Siehe hierzu auch Abschnitt 4.1).



Bsp. 7.1:

Allerdings fällt in unserem Experiment die 6. Stufe aus dem Rahmen, weil sich in der deskriptiven Auswertung eine Zunahme von der 5. zur 6. Stufe herausstellte. Das kann aber unter Umständen durch Statistische Ausreisser oder Zufälle erklärt werden. Und da unsere Stichprobe relativ klein war, sind unsere Ergebnisse auch nicht repräsentativ genug, um genaue Aussagen zu treffen.

Kritikpunkte:

- (i) Der Unterschied zwischen der 5. und 6. Stufe war statistisch nicht signifikant. Da zusätzlich – wie die Autoren ausführen – die Stichprobe klein war, macht es wenig Sinn ein derartiges Ergebnis zu diskutieren.
- (ii) Die Autoren führen als Gründe statistische Ausreisser an. Diese Behauptung muss begründet werden (z.B. indem gezeigt wird, dass nach Entfernen der Ausreisser der beobachtete Unterschied verschwindet).

8. Literaturverzeichnis



Prinzip:

Es müssen alle im Text angeführten Referenzen (mit Ausnahme der Sekundärzitate) in die Literaturliste aufgenommen werden. Die Literaturangaben haben den Konventionen der American Psychological Association zu entsprechen (Vergleiche Publication manual of the American Psychological Association, 2010, 6th edition). Am einfachsten und wenigsten fehleranfällig ist die Nutzung eines Zitierprogramms (Mendeley, Citavi, EndNote, etc) und daher sehr empfohlen (vor allem für längere Arbeiten).



Bemerkung zu den Zitationsregeln:

Table 6.1 Basic Citation Styles

Type of citation	First citation in text	Subsequent citations in text	Parenthetical format, first citation in text	Parenthetical format, subsequent citations in text
One work by one author	Walker (2007)	Walker (2007)	(Walker, 2007)	(Walker, 2007)
One work by two authors	Walker and Allen (2004)	Walker and Allen (2004)	(Walker & Allen, 2004)	(Walker & Allen, 2004)
One work by three authors	Bradley, Ramirez, and Soo (1999)	Bradley et al. (1999)	(Bradley, Ramirez, & Soo, 1999)	(Bradley et al., 1999)
One work by four authors	Bradley, Ramirez, Soo, and Walsh (2006)	Bradley et al. (2006)	(Bradley, Ramirez, Soo, & Walsh, 2006)	(Bradley et al., 2006)
One work by five authors	Walker, Allen, Bradley, Ramirez, and Soo (2008)	Walker et al. (2008)	(Walker, Allen, Bradley, Ramirez, & Soo, 2008)	(Walker et al., 2008)
One work by six or more authors	Wasserstein et al. (2005)	Wasserstein et al. (2005)	(Wasserstein et al., 2005)	(Wasserstein et al., 2005)
Groups (readily identified through abbreviation) as authors	National Institute of Mental Health (NIMH, 2003)	NIMH (2003)	(National Institute of Mental Health [NIMH], 2003)	(NIMH, 2003)
Groups (no abbreviation) as authors	University of Pittsburgh (2005)	University of Pittsburgh (2005)	(University of Pittsburgh, 2005)	(University of Pittsburgh, 2005)

8.1 Zitieren IM Text

8.1.1 Werke mit zwei AutorInnen:

Wenn im Text auf eine Arbeit Bezug genommen wird, welche von zwei AutorInnen gemeinsam veröffentlicht wurde, so werden die beiden Namen mit „und“ verbunden. Wenn auf diese Autoren in einer Klammer Bezug genommen wird, werden die beiden Namen mit „&“ verbunden:

...wie Schmid und Maier (2003) zeigten...

...wurde früher schon gezeigt (Schmid & Maier, 2003)

8.1.2 Bei mehr als zwei Autoren : bei 1. Nennung ausgeschrieben:

Schmid, Müller, Jensen und Maier (1991) zeigten...

...wurde früher schon gezeigt (Schmid, Müller, Jensen & Maier, 1991)...

bei nachfolgender Nennung:

Schmid et al. (1991) ...

...wurde früher schon gezeigt (Schmid et al., 1991).

8.1.3 Werke mit mehr als fünf Autoren:

Im Text jeweils nur erster Name und „et al.“ (auch beim ersten Auftreten).

Im Literaturverzeichnis werden die ersten *sechs* Autoren mit Namen angegeben, auf die restlichen wird mit „et al.“ referenziert.

8.1.4 Mehrere Angaben in einem Klammerausdruck:

in alphabetischer Reihenfolge der Erstautoren (danach nach Erscheinungsjahr) gereiht.

(Beck, 1979, 1982; Müller et al., 1976; Schmid & Maier, 1968)

8.1.5 Autoren mit gleichem Familiennamen:

Bei Autoren mit gleichem Familiennamen werden im Text die Initialen vor den Familiennamen gesetzt, auch wenn das Erscheinungsjahr unterschiedlich ist:

S. Sternberg (1969) führte die Teilberichtsmethode ein, R. J. Sternberg (1985) postulierte ein neues Intelligenzmodell.

8.1.6 Wörtliche Zitate

werden in Anfangs- und Schlusszeichen und unter Angabe der Seitenzahl gesetzt: Auch Ewert (1983) stellt fest, dass „leider ... die Verwendung des Begriffes ‚Egozentrismus‘ nicht eindeutig“ (S. 117) ist.

8.1.7 Längere Zitate (mehr als 40 Wörter)

werden als eigener Absatz dargestellt und der Text wird eingerückt. Sie enthalten keine Anführungszeichen. Autor(en), Erscheinungsjahr und Seitenzahl werden entweder zu Beginn oder am Ende aufgelistet.



Bsp. 8.1.1 (Literaturangaben von wissenschaftlichen Artikel):

American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th edition). Washington, DC: APA.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Metz, C. E., & Kronman, H. B. (1980). Statistical significance tests for binormal ROC curves. *Journal of Mathematical Psychology*, 22, 218-243.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53, 53-81.

Peterson, L. R., & Peterson, M. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, 58, 193-198.

9. Anhang



Prinzip:

Der Anhang beginnt auf einer neuen Seite und mit einer eigenen Seitennummerierung. Er enthält im Allgemeinen:

- ◆ Die standardisierten Instruktionen
- ◆ Das vollständige Stimulusmaterial
- ◆ Die Rohdaten
- ◆ Optional Tabellen (mit fortlaufender Nummerierung, z.B. Tab.A1, Tab.A2, etc.) mit detaillierten Ergebnissen, die für die Hauptfragestellung nicht so interessant sind, aber trotzdem berichtet werden sollen.

**Bemerkung:**

Falls das Stimulusmaterial oder die Rohdaten zu komplex sind, so werden sie nicht im Anhang aufgeführt (daher die Wendung „im Allgemeinen“).

10. Style

10.1 Zahlen ausgedrückt als Zahl:

Zahlen 10 oder grösser; wenn im Zusammenhang mit einer Messgrösse (z.B. cm), wenn mathematische Funktionen ausgedrückt werden (z.B. multipliziert mit 5); Daten, Skalen etc.

10.2 Zahlen ausgedrückt als Wort:

Wenn am Beginn des Satzes (möglichst vermeiden): z.B. Achtundvierzig Versuchspersonen nahmen am Experiment teil; bei universell akzeptierten Ausdrücken z.B. die Zwölf Apostel.

10.3 Dezimalzahlen:

Eine Null wird nur in dem Fall vor dem Punkt genutzt, wenn die Angabe theoretisch 1 übersteigen kann. D.h. dies ist nicht der Fall bei p Werten, jedoch bei F Werten.

10.4 Schriftgrösse und -art

Extra: Arial, size 12, Zeilenabstand 1.5. Bericht max. 10 Seiten.

Bachelor & Masterarbeit: Times New Roman, size 12, Zeilenabstand 1.5, 30 bis 40 Seiten inkl. Referenzliste.

Anhang A: Boxplots

In diesem Anhang wird erklärt, aus welchen Elementen ein Boxplot besteht und welche Erkenntnisse sich mit seiner Hilfe über die Verteilung einer oder mehrerer Variablen gewinnen lassen.

Der *Boxplot* (auch Box-Whisker-Plot) ermöglicht eine informative und zugleich platzsparende graphische Darstellung der Verteilung (pseudo-) kontinuierlicher Variablen (vgl. Abb.-A. 1). Er vermittelt schnell einen Überblick über den Bereich, in dem die Daten liegen und wie sie sich über diesen Bereich verteilen. Ein wesentlicher Vorzug des Boxplot besteht darin, dass sich leicht die Verteilung einer Variablen in verschiedenen Subgruppen vergleichen lässt. Die Werte der darzustellenden Variable(n) werden üblicherweise auf der y -Achse abgetragen, auf der x -Achse können mehrere Subgruppen dargestellt werden. Ein Boxplot besteht immer aus einem *Kasten* („Box“) und zwei vertikalen Linien die sich unter- und oberhalb der Box befinden. Die vertikalen Linien werden als *Whisker* bezeichnet.

Das zentrale Element des Boxplot ist der Kasten. Er repräsentiert so genannte *Quantile* der abgebildeten Verteilung. Ein p -Quantil ist derjenige Wert x_p ($0 < p < 1$), der die Werte in zwei Gruppen teilt, so dass mindestens $p \cdot 100\%$ der Daten kleiner oder gleich x_p und mindestens $(1-p) \cdot 100\%$ grösser oder gleich x_p sind. Bekannte Spezialfälle für Quantile sind der *Median*, welcher dem .5-Quantil entspricht und die so genannten *Quartile*, die sich aus dem .25-Quantil (Q_1), dem .5-Quantil (Q_2) und dem .75-Quantil (Q_3) zusammensetzen. Folglich teilt der Median die Werte in zwei Hälften auf, während die Quartile die Verteilung in vier Teile aufteilen, in denen jeweils ca. 25% der Werte liegen.

In die Mitte des Kastens wird der *Median* der Verteilung als durchgezogene Linie eingezeichnet. Die durchgezogene Linie zeigt damit die zentrale Tendenz der Verteilung auf.

Das untere Kastenende repräsentiert das erste Quartil (Q_1), das obere Kastenende repräsentiert das dritte Quartil (Q_3). Die Höhe des Kastens entspricht damit dem Bereich, in dem die mittleren 50% der Daten liegen. Er wird als *Interquartilsabstand* bezeichnet entspricht der Differenz zwischen dem ersten und dritten Quartil:

$$IQ = Q_3 - Q_1.$$

Aus der Höhe des Kastens lässt sich ablesen, wie stark die mittleren 50% der Werte streuen.

Die Whisker verlängern den Kasten nach unten und oben. An ihnen zeigt sich, wie weit die Daten über den Interquartilsabstand hinausreichen. Ihre Länge ist auf maximal $1.5 \cdot IQ$ beschränkt. Die Whisker sind jedoch nicht notwendigerweise genau $1.5 \cdot IQ$ lang. Das Ende des oberen Whisker liegt bei maximal $Q_3 + 1.5 \cdot IQ$. Kommt der Wert $Q_3 + 1.5 \cdot IQ$ in den Daten vor, so endet der Whisker genau dort. Kommt der Wert nicht in den Daten vor, so endet der Whisker beim höchsten real auftretenden Wert, der sich noch innerhalb von $Q_3 + 1.5 \cdot IQ$ befindet. Dasselbe Prinzip gilt für den unteren Whisker. Tritt der Wert $Q_1 - 1.5 \cdot IQ$ in den Daten auf, so endet der Whisker genau dort. Tritt der Wert nicht in den Daten auf, so endet er beim nächst größeren beobachteten Wert.

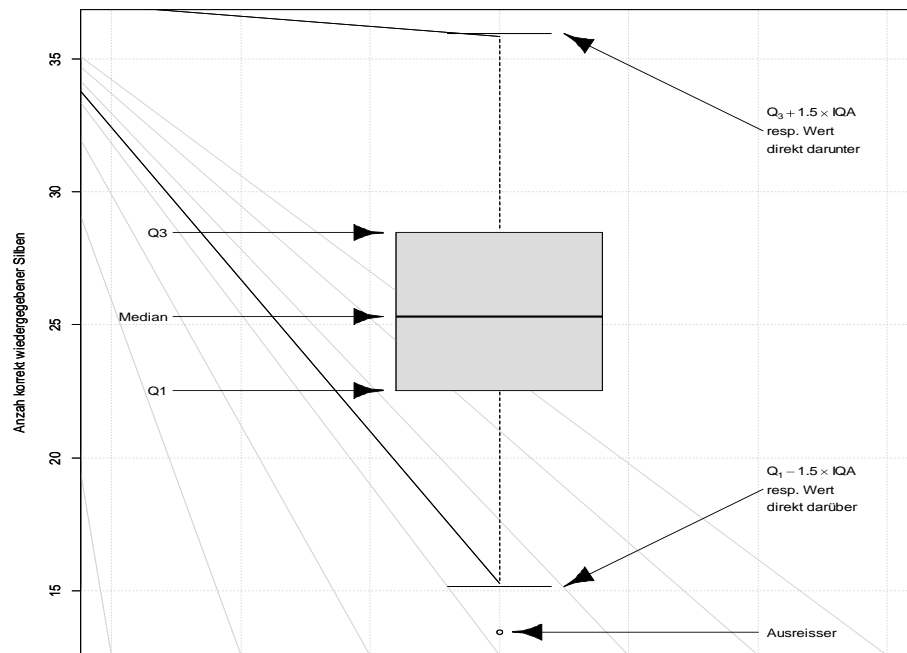


Abb-A. 1: Boxplot der Anzahl korrekt wiedergegebener Silben.

Im Boxplot werden *Ausreisser* als Punkte unter- resp. oberhalb der Whisker eingezeichnet. Ein Ausreisser ist dadurch definiert, dass er ausserhalb des Bereiches liegt, der von den beiden Werten $Q_1 - 1.5 \cdot IQ$ und $Q_3 + 1.5 \cdot IQ$ gebildet wird. Es sind dies also Werte, die stark vom generellen Trend der Daten abweichen. Ausreisser, die kleiner als $Q_1 - 3 \cdot IQ$ oder grösser als $Q_3 + 3 \cdot IQ$ sind, werden Extremwerte genannt. Sie werden von einigen Statistikprogrammen optisch von den anderen Ausreissern abgehoben.

Abb-A 1 zeigt Boxplots von drei verschiedenen Datensätzen und stellt sie den entsprechenden Histogrammen gegenüber. Damit sich die Boxplots besser mit den Histogrammen vergleichen lassen, wurden die Boxplots entgegen den üblichen Konventionen horizontal dargestellt.

An der Form des gesamten Boxplots erkennt man, ob es sich um eine symmetrische oder asymmetrische Verteilung handelt. Liegt dem Boxplot eine symmetrische Verteilung zugrunde, befindet sich der durchgezogene Strich, der den Median repräsentiert, in der Mitte des Kastens und beide Whisker sind ungefähr gleich lang. Liegt der Median näher beim unteren Kastenende, ist die Verteilung linkssteil, liegt er näher beim oberen Kastenende, ist die Verteilung rechtssteil. Die Länge der Linien im Vergleich zur Höhe des Kastens zeigt, wie stark die Enden der Verteilung gestreckt sind. Aus Abb-A 1. wird anhand des Boxplot für Datensatz I sofort ersichtlich, dass die Verteilung linkssteil ist und, dass die Verteilung rechts einen deutlich längeren Schwanz aufweist. Datensatz II und III sind hingegen symmetrisch, was sich daran zeigt, dass der Median sich in der Mitte des Kastens befindet und die Whisker ungefähr gleich lang sind. Der Boxplot vermag jedoch nicht aufzudecken, dass es sich bei Datensatz III um eine bimodale Verteilung handelt, was die Notwendigkeit demonstriert, empirisch gewonnene Daten immer auf mehrere Arten zu visualisieren.

Die ausserhalb der Whisker gezeichneten Punkte zeigen diejenigen Werte, die stark von der zentralen Tendenz der Verteilung abweichen. Insbesondere wenn diese sehr weit vom Ende der Whisker entfernt sind, sollte man sich überlegen, wie diese Werte zustande kommen, weil Ausreisser Mittelwert und Varianz der Stichprobe stark ver-

zerren können. Möglicherweise handelt es sich um einen Fehler bei der Eingabe oder der Aufbereitung der Daten. In *Abb-A 1* weisen alle drei Datensätze einige Ausreisser auf. Insbesondere in Datensatz III weichen die Ausreisser – im Vergleich zur Variabilität der übrigen Daten – relativ stark vom Rest der Daten ab und es sollte überprüft werden, ob diese Werte auf Fehler zurückgehen und überhaupt sinnvoll interpretierbar sind.

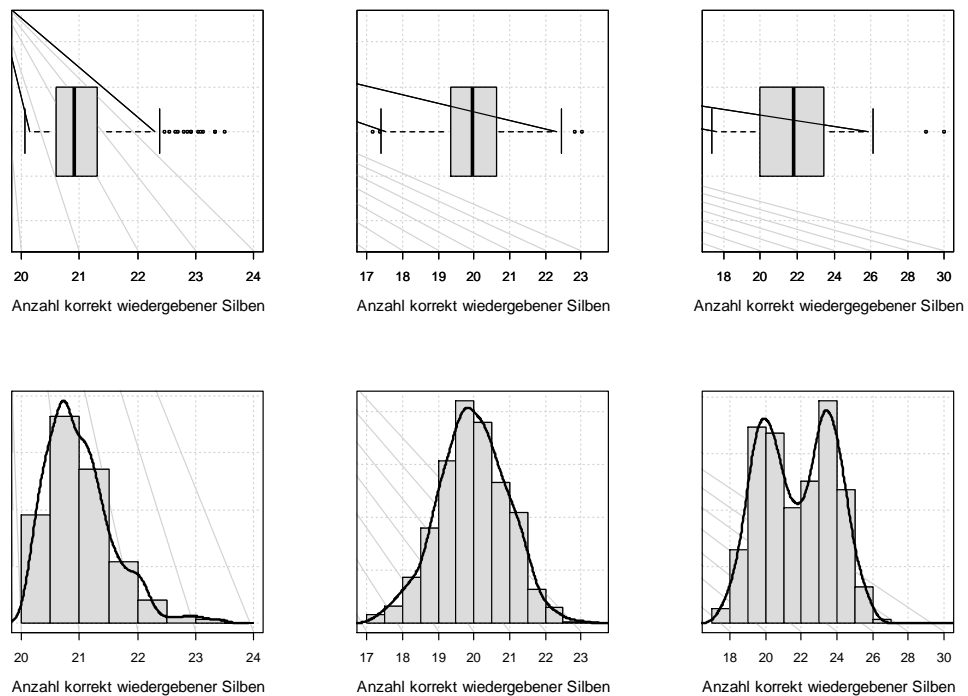


Abb-A 1.: Boxplots der Anzahl korrekt wiedergegebener Silben mit zugehörigen Histogrammen für drei verschiedene Datensätze. Die durchgezogenen Linien in den Histogrammen repräsentieren Kerndichteschätzer. Die Histogramme beruhen auf denselben Daten wie die jeweils direkt darüber liegende Boxplots.

Anhang B: Interpretation und Darstellung von Interaktionen

In Abschnitt 6.2.1 wurde eine korrekte Interpretation einer Interaktion präsentiert. Konkret wurde in Bp 6.2.1.2 eine Interaktion zwischen Latenzzeit und Stimulustyp geschildert. In diesem Anhang soll erklärt werden, wie Interaktionen zwischen zwei Variablen zu interpretieren sind.

Das Vorhandensein einer 2-weg Interaktion weist darauf hin, dass keine konsistenten Haupteffekte der beiden unabhängigen Variablen, welche miteinander interagieren, vorliegen. Dies bedeutet, dass die Grösse des Effekts der einen unabhängigen Variable (UV) davon abhängt, welchen Wert die andere UV besitzt. So hängt in Bsp 6.2.1.2 die Grösse des Effekts der UV *Latenzzeit* davon ab, welchen Wert die UV *Stimulustyp* besitzt: Falls die verwendeten Silben ähnlich sind, so ist der Effekt der Latenzzeit gering. Sind hingegen die verwendeten Silben unähnlich, so hat die Latenzzeit einen grossen Einfluss auf die Wiedergabeleistung (umgekehrt hängt die Grösse des Effekts der UV *Stimulustyp* davon ab, welchen Wert die UV *Latenzzeit* besitzt).

Die inhaltliche Interpretation eines Interaktionseffekts vereinfacht sich, wenn man letzteren als so genannten Moderatoreffekt betrachtet. Ein *Moderatoreffekt* liegt vor, wenn der Effekt einer UV auf die abhängige Variable durch eine andere Variable *moderiert* (lateinisch *moderare* bedeutet „steuern“, „lenken“, „mässigen“) wird. Dies bedeutet, dass der Effekt, den die Hauptvariable auf die abhängige Variable (AV) ausübt, vom Wert bzw. von der Stufe der Moderatorvariable abhängt (vgl. die Beschreibung im unmittelbar vorangegangenen Absatz).

Um eine Interaktion als Moderatoreffekt zu interpretieren, wählt man eine UV als Hauptvariable aus, während die andere als *Moderatorvariable* fungiert. Die Wahl von Haupt- und Moderatorvariable hängt von inhaltlichen Überlegungen ab: Die inhaltlich bedeutsamere UV wird als Hauptvariable und die unwichtigere als Moderator betrachtet. So wurde in Bsp 6.2.1.2 die UV *Latenzzeit* als Hauptvariable konzipiert, deren Einfluss auf die korrekte Wiedergabe durch die zweite UV *Stimulustyp* moderiert wird. Im konkreten Fall bedeutet dies: Der Effekt der *Latenzzeit* auf die Wiedergabeleistung ist abgeschwächt, wenn die Silben ähnlich sind. Der Ausdruck „Effekt der Latenzzeit“ bedeutet, dass bei Vorliegen einer Latenzzeit die Wiedergabeleistung höher ist als ohne Latenzzeit.

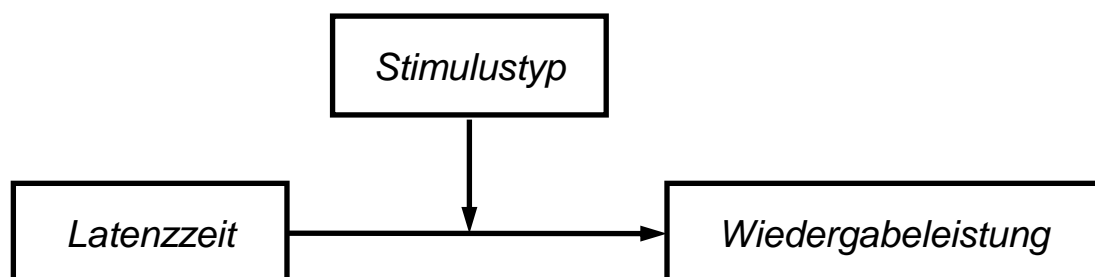


Abb-B. 1: Graphische Darstellung eines Moderatoreffekts: Der Effekt der Latenzzeit auf die Wiedergabeleistung wird durch den Stimulustyp moderiert.

Abb-B. 1 zeigt eine graphische Darstellung eines Moderatoreffekts. Die Rechtecke symbolisieren die abhängigen und unabhängigen Variablen. Der Effekt der Latenzzeit auf die Wiedergabeleistung wird durch den Pfeil zwischen den beiden Variablen symbolisiert. Der Moderatoreffekt wird ebenfalls durch einen Pfeil symbolisiert. Dieser Pfeil ist jedoch auf einen Pfeil und nicht auf eine Variable gerichtet, wodurch dargestellt werden soll, dass es sich um den Einfluss auf einen Effekt handelt, d.h.

die Variable Stimulustyp hat einen Einfluss auf den Effekt (der Latenzzeit auf die Wiedergabeleistung).

Die Interaktion kann auch in umgekehrter Weise interpretiert werden, indem der Effekt des Stimulustyps als durch die Latenzzeit moderiert konzipiert wird. Welche Interpretation gewählt wird hängt von rein inhaltlichen Gesichtspunkten ab. Im aktuellen Beispiel ist diese Interpretation der Interaktion inhaltlich weniger ansprechend als erstere.

Interaktionsgraphiken

Zur graphischen Darstellung von Interaktionen verwendet man so genannte Interaktionsgraphiken. Die Interpretation einer Interaktion kann durch die Auswahl jener Variable, welche die x-Achse der Graphik bildet, erleichtert oder auch erschwert werden. Die Interaktionsgraphik sollte so gestaltet sein, dass die inhaltliche Interpretation der Interaktion optimal illustriert wird.

Die Bedeutung der Wahl der Variable, welche die x-Achse der Interaktionsgraphik bildet, wird durch den Vergleich von Abb-B. 2 und Abb-B. 3 illustriert. Interpretiert man die Interaktion zwischen *Latenzzeit* und *Stimulustyp* auf die oben dargestellte Weise, wonach die Latenzzeit die Haupt- und der Stimulustyp die Moderatorvariable bildet, so ist die Darstellung in Abb-B. 2 besser geeignet als jene in Abb-B. 3. Erstere macht nämlich unmittelbar deutlich, dass bei unähnlichen Silben die Latenzzeit einen grossen Effekt hat (die Fehlerbalken überlappen sich nicht), während bei ähnlichem Stimulusmaterial die Fehlerbalken sich klar überlappen und daher (vermutlich) kein Effekt der Latenzzeit bei ähnlichen Silben gegeben ist.

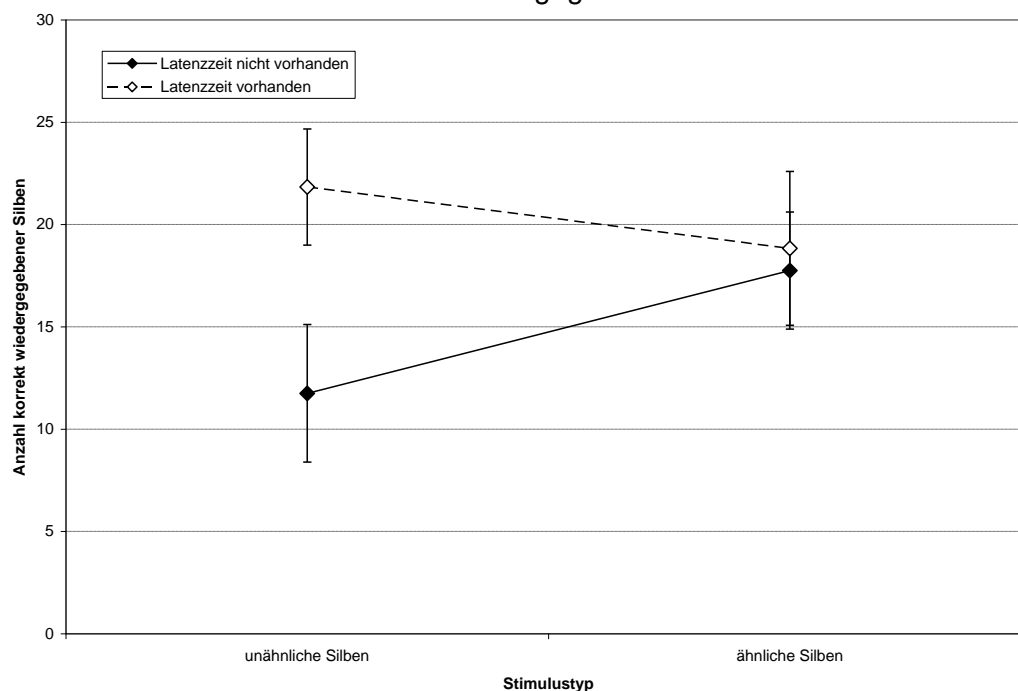


Abb-B. 2: Interaktionsgraphik der Latenzzeit \times Stimulustyp – Interaktion mit der Variable Stimulustyp als x-Achse. Die Fehlerbalken repräsentieren zwei Standardfehler des Mittelwerts.

Aus Abb-B. 3, in welcher die Latenzzeit die x-Achse bildet, ist dieser Moderatoreffekt nicht sofort erkennbar.

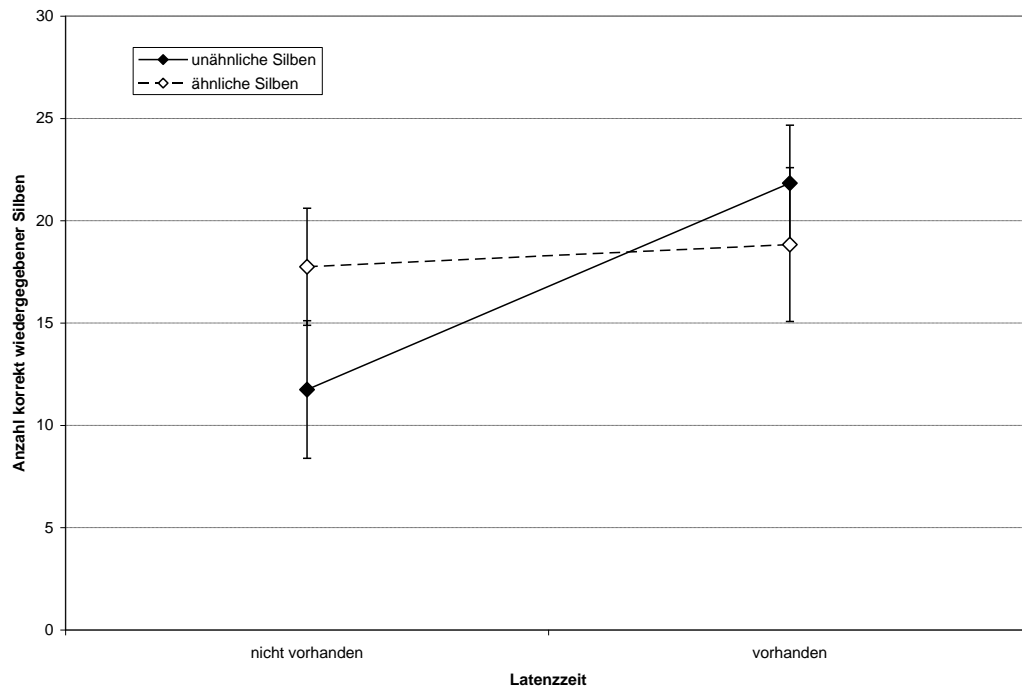


Abb-B. 3: Interaktionsgraphik der Latenzzeit \times Stimulustyp – Interaktion mit der Variable Latenzzeit als x-Achse. Die Fehlerbalken repräsentieren zwei Standardfehler des Mittelwerts.

Im Text wurde ein Beispiel für die korrekte Interpretation einer Interaktion präsentiert (vgl. Bsp 6.2.1.2). Hier seien noch zwei Beispiele angeführt, welche mögliche Fehler bei der Interpretation von Interaktionen illustrieren.



Bsp-B. 1:

Die Interaktion zwischen den Bedingungen Silbenauswahl und Latenzzeit führte zu einem signifikanten Unterschied ($F(1,44)=7.783$, $p < .01$), wobei die UV2 Latenzzeit als Hauptvariable und die UV3 Silbenauswahl als Moderatorvariable fungierte. Die Interaktion beruht auf der Tatsache, dass die unterschiedlichen Bedingungen der Silbenauswahl alleine keine signifikanten Effekte aufwiesen ($F(1,44)=.865$, $p > .35$) in Verbindung mit der Latenzzeit jedoch hochsignifikant war.

Kritikpunkte:

- (i) Gewöhnlich wird nicht explizit angegeben, welche Variable als Haupt- und welche als Moderatorvariable fungiert. Für die Einteilung der UVn in eine Haupt- und eine Moderatorvariable gibt es keine formalen bzw. statistischen Gründe. Die Unterteilung hängt vielmehr davon ab, welche Variable von der Autorin des Berichts (aufgrund inhaltlicher Überlegungen) als zentral betrachtet wird. Diese Variable fungiert dann in der Interpretation als Hauptvariable. Aus diesem Grunde macht es wenig Sinn zu schreiben, dass „[...] die [...] Latenzzeit als Haupt- und die [...] Silbenauswahl als Moderatorvariable fungierte“.
- (ii) Die Beschreibung der Interaktion ist völlig falsch. Insbesondere weist sie folgende Probleme auf:
 - ♦ Aus der Beschreibung geht nicht hervor, dass der Effekt der *Latenzzeit*, welche als Hauptvariable fungiert, durch die Variable *Silbenauswahl* moderiert wird. In der Beschreibung scheint vielmehr die *Silbenauswahl* zur Hauptvariablen mutiert zu sein.

- ◆ In obiger Bemerkung zur Interpretation von 2-wege Interaktionen wurde darauf hingewiesen, dass Moderation bedeute, dass der Effekt, welchen die Hauptvariable auf die AV ausübt, von der *Stufe* der Moderatorvariablen abhängig ist. In der Beschreibung der Interaktion befindet sich jedoch kein Hinweis darauf, wie der Einfluss der Latenzzeit in Abhängigkeit von der Stufe der Silbenauswahl variiert.

**Bsp-B. 2:**

Die Interaktion zwischen den Behaltensintervallen und den Blöcken wurde auch signifikant ($F(9,396)=1.958, p < .05$). Die Interaktion beruht auf der Tatsache, dass die verschiedenen langen Behaltensintervalle über die Blöcke hinweg einen Effekt aufweisen, wobei der Blockeffekt bei kurzen Behaltensintervallen grösser ist als bei langen.

Kritikpunkte:

Die gegebene Darstellung der Interaktion weist die folgenden Mängel auf:

- (i) Die Tatsache, dass der Faktor *Behaltensintervall (BI)* einen Effekt über die Blöcke hinweg aufweist, ist für die *BI*×*Block* Interaktion und deren Interpretation irrelevant (Relevant ist hingegen der *differentielle* Effekt, d.h. wie der Effekt des *BI* sich über die *Blöcke* hinweg verändert).
- (ii) Die Darstellung weist einen schwer nachvollziehbaren „Perspektivenwechsel“ auf: Zuerst wird auf den Effekt des *BI* verwiesen. Man würde daher annehmen, dass die Autorinnen das *BI* als Hauptvariable betrachten, deren Effekt über die Blöcke hinweg variiert. Der *Block* würde somit als Moderator fungieren. Im letzten Teilsatz wird jedoch der *Block* als Hauptvariable betrachtet, dessen Effekt durch das *BI* moderiert wird.
- (iii) Es macht inhaltlich wenig Sinn, den *Block* als Hauptfaktor zu betrachten und das *BI* als Moderator, da das *BI* inhaltlich die zentralere Variable darstellt, deren Einfluss auf die Behaltensleistung in der Studie untersucht wird.

11. Eigenständigkeitserklärung

Für Bachelor Arbeiten siehe <http://www.unifr.ch/psycho/de/teaching/bachelor/ba-thesis>