

# Diplomarbeit

Titel der Diplomarbeit

*163 Wörter als Berufseinstiegshürde*

*Die Entwicklung von schriftlichen Zulassungsprüfungen in Deutsch  
beim Österreichischen Bundesheer*

Verfasser

Robert Rozinsky

angestrebter akademischer Grad

Magister der Philosophie (Mag. phil.)

Wien, 2012

Studienkennzahl lt. Studienblatt: A332

Studienrichtung lt. Studienblatt: Deutsche Philologie

Betreuer: Univ.-Prof. Dr. Hans-Jürgen Krumm

## **Persönlicher Dank**

Ein erster Dank gilt jenen Personen, die mir diese Studie am Sprachinstitut des Bundesheeres ermöglicht haben. Besonders seien hier *Wolfgang Zecha* und *Walter Rys* erwähnt, die mich trotz ihrer beruflichen Verpflichtungen sowohl in fachlicher als auch in institutioneller Hinsicht maßgeblich unterstützt haben. Sie haben die vorliegende Studie nicht nur in organisatorischer und fachlicher Hinsicht begleitet, sondern sich auch auf vorgegebene Termine wie z. B. Experteninterviews eingelassen. Nicht zu vergessen dabei ist die wertvolle Unterstützung bei der Kontaktaufnahme zu anderen an der Studie beteiligten militärischen Dienststellen.

Wie die Untersuchungsschritte zeigen, wurde die Studie an unterschiedlichen Dienststellen des Österreichischen Bundesheeres durchgeführt – die Datenerhebungen fanden am *Sprachinstitut des Bundesheeres* in Wien sowie an den militärischen Dienststellen in *Absam*, *Enns* und *Ebelsberg* statt. Ein Dank gebührt daher allen beteiligten Dienststellen für die Unterstützung und den zahlreichen Datenlieferantinnen für das Ausfüllen der Fragebögen. Nicht zu vergessen ist hier die wertvolle Hilfe von *Peter Kadletz*, der als Tester die notwendigen Auswertungsdaten lieferte und *Gerhard Simperl*, der sich als zweite Beobachtungsperson zur Verfügung gestellt hat.

Ein weiterer Dank gilt den Expertinnen *Drorit Lengyel*, *Marion Döll* und *Martina Glaboniat* für die fachliche Beratung.

Hervorheben möchte hier meinen Betreuer *Hans-Jürgen Krumm* von der Universität Wien. Er hat durch seine Erfahrung in Umgang mit wissenschaftlichen Arbeiten und seine umfassende fachliche Kompetenz meine Untersuchung bereichert und die nötigen Impulse gesetzt. Herzlichen Dank für die in jeder Hinsicht professionelle Hilfe.

Dank gilt auch meiner Familie, insbesondere meiner Frau *Christine* für den nötigen privaten Freiraum und das Verständnis für mein nächtliches Arbeiten an der Diplomarbeit.

# Inhaltsverzeichnis

---

Danksagung	2
<b>Einleitung</b>	<b>5</b>
<b>1 Einführung in das Fachgebiet</b>	<b>9</b>
1.1 Sprachausbildung beim Bundesheer	9
1.2 Militäreigene theoretische Grundlagen	10
1.3 Bildungsstandards, Referenzrahmen und Sprachrealität	12
1.4 Sprachüberprüfungen und sprachliche Förderung beim ÖBH	16
1.5 Das Sprachinstitut als Prüfungsinstitution	19
1.6 Arbeitsgruppe Prüfmittel Deutsch	20
<b>2 Das Diktat innerhalb der Zulassungsprüfung Deutsch</b>	<b>22</b>
2.1 Funktionalität und Nützlichkeit des Verfahrens	22
2.2 Testdurchführung	24
2.3 Die Zielgruppe	25
2.4 Bestehensbedingungen - Auswirkung auf Zulassungsentscheidung	27
<b>3 Die Testaufgabe und ihr Umfeld</b>	<b>28</b>
3.1 Sprachliche Handlungsfähigkeit im Beruf	28
<b>4 Sprachtests zur Erfassung von Sprachkompetenzen als Untersuchungsbasis</b>	<b>30</b>
4.1 Bumerang	32
4.2 Prozessbegleitende Diagnoseinstrumente	35
4.3 Kommunikatives Testen und Qualitätssicherung beim ÖSD	37
4.4, Hamburger Schreibprobe	39
<b>5 Anforderungen an Sprachtests</b>	<b>41</b>
5.1 Testtheoretische und testmethodische Welten	41
5.2 Allgemeine Anforderungen an Sprachtests	47
5.3 Qualitäts- und Nützlichkeitskriterien	49
<b>6 Der Testgegenstand: Die Rechtschreibfähigkeit</b>	<b>58</b>
6.1 Der Fehlerbegriff in der Forschung	58
6.2 Expertise zum Diktat aus der Fachwelt	61
<b>7 Forschungsziele – Hypothese</b>	<b>64</b>
<b>8 Methodischer Ansatz – empirische Vorgangsweise</b>	<b>65</b>
8.1 Bewertung und Erfassung der Testauswertungsdaten	66
8.2 Fragebogenuntersuchung	67
8.3 Teilnehmende Beobachtung	69
8.4 Experteninterview	70
<b>9 Machbarkeitsanalyse und Sicherstellung der Untersuchungsvoraussetzung</b>	<b>72</b>
9.1 Zugangsmöglichkeiten zu den Daten	72
9.2 Personell und finanziell verfügbare Mittel	73
9.3 Datenlieferanten	73
9.4 Voruntersuchung	73

<b>10 Forschungsplan</b>	<b>74</b>
<b>11 Beschreibung der Stichproben</b>	<b>76</b>
11.1 Erhebung 1 in Ebelsberg	77
12.2 Erhebung 2 in Enns	80
<b>12 Darstellung der Ergebnisse</b>	<b>82</b>
12.1 Ergebnisse der Stichproben im Vergleich	82
12.2 Teilnehmende Beobachtung	86
12.3 Ergebnisse aus der Fragebogenerhebung	89
12.4 Ergebnisse aus den Experteninterviews	95
<b>13 Interpretation der Ergebnisse</b>	<b>105</b>
13.1 Interpretation der Testergebnisse	105
13.2 Interpretation der Ergebnisse aus den Fragebögen	108
13.3 Interpretation und Zusammenschau der Experteninterviews	112
<b>14 Nützlichkeit und Qualität des Diktats als Schreibaufgabe</b>	<b>116</b>
14.1 Überprüfung nach den Nützlichkeitskriterien	116
14.2 Auswirkungen auf die Zulassungsentscheidung	119
14.3 Folgenabschätzung in pädagogischer und gesellschaftlicher Hinsicht	120
<b>15 Schlussfolgerungen zur Untersuchung</b>	<b>120</b>
15.1 Hinweise auf Mängel	120
15.2 Bestätigung der Hypothese	120
15.3 Vergleich mit anderen Prüfverfahren	123
<b>16 Erprobungs- und Evaluierungsschritte bei der Einführung des Testformats NEU</b>	<b>124</b>
16.1 Berufskompetenzmodell als Basis zur Entwicklung neuer Testverfahren	124
16.2 Vom Diktat zur Bildschreibung – Erprobung einer Alternative zum Diktat	126
16.3 Gewichtung, Bewertung und Punktevergabe mittels Auswertungsraster	130
<b>17 Vergleichsstudie – Einsatz des neuen Testformats</b>	<b>133</b>
17.1 Untersuchungsmethodik	133
17.2 Beschreibung der Stichprobe	134
17.3 Testergebnisse der Vergleichsstudie	137
17.4 Fragebogenergebnisse aus der Vergleichsstudie	138
<b>18 Interpretation der Vergleichsstudie</b>	<b>141</b>
<b>19 Schlussfolgerung und Auswirkungen auf die DaZ/DaM-Didaktik</b>	<b>142</b>
<b>Anhang</b>	
Literaturverzeichnis	144
Abkürzungsverzeichnis	147
Abbildungsverzeichnis	147
Abstract	148
Lebenslauf	149

## **Einleitung – „Fehler weisen mir den Weg“**

### Ausgangspunkt – Erkenntnisinteresse

Eine Definierung sprachlicher Kompetenzen und Fertigkeiten findet sich nicht nur in internationalen Lernzielvereinbarungen wie im Gemeinsamen europäischen Referenzrahmen für Sprachen und in den nationalen Bildungsstandards, sondern vielfach auch in diversen standardisierten Testverfahren wieder.

Vor dem Hintergrund der gegenwärtigen gesellschaftlichen und wirtschaftlichen Verhältnisse und den mit Globalisierung und Migration einhergehenden Wandlungsprozessen rückt der gesellschaftliche und soziale Stellenwert von Sprache zusehends in den Vordergrund. Spätestens seit Einführung der PISA-Studie sowie der Durchführung von Deutschprüfungen bei Zuwanderern in Österreich haben Testverfahren sowohl politische als auch ethische Dimensionen erreicht. Sie prägen die Vorstellungen von Sprache und deren Beherrschung. Trotz „PISA-Schock“ wurden seitens der Politik allerdings oftmals naive Konsequenzen gezogen, die sich vor allem bei den Zugangsbestimmungen von Menschen mit nichtdeutscher Muttersprache deutlich zeigen. Mangels nachhaltiger Integrationsbemühungen fungiert Sprache offenbar zunehmend als Macht- und Selektionsinstrument, insbesondere, wenn es um die Legitimierung einer Sprache in Verbindung mit nationaler Zugehörigkeit geht.<sup>1</sup>

Trotz reger empirischer Forschungsaktivitäten im Bereich des Messens und Prüfens von sprachlicher Handlungs- und Kommunikationsfähigkeit orientieren sich Steuerungs- und Selektionsmechanismen bei Sprachtests gerade bei Berufseignungs- und Zulassungsprüfungen noch vielerorts an der normgerechten, korrekten Umsetzung von Sprache. Wie die Praxis zeigt, wird bei schriftlichen Leistungsüberprüfungen nach wie vor die korrekte Rechtschreibung als alleiniges sprachqualifizierendes Messkriterium, als vordefinierter Standard herangezogen. Denn erfahrungsgemäß tritt gerade beim Schreiben die Einhaltung und Nichteinhaltung der Norm viel stärker hervor als beim Sprechen.<sup>2</sup>

---

<sup>1</sup> vgl. dazu Ehlich, Konrad (2005): eine Expertise zu „Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuellen Sprachförderung von Kindern mit und ohne Migrationshintergrund“. In: Gogolin, Ingrid u. a. [Hrsg.]: Sprachdiagnostik von Kindern und Jugendlichen mit Migrationshintergrund. Dokumentation einer Fachtagung am 14. Juli 2004 in Hamburg. Münster: Waxmann S. 33 – 50.

<sup>2</sup> vgl. dazu Glaboniat, Manuela (2009): „Wie viel Standardisierung braucht ein Test?“ Vom Nutzen der Standardisierungstendenzen im Bereich des Sprachenlernens. In: ÖDaF-Mitteilungen, Heft 2/2009, S. 44.

Wie groß die Akzeptanz sprachqualifizierender Messkriterien ist, zeigt sich deutlich in staatlichen Verordnungen, insbesondere in der Festlegung amtlicher Regelwerke wie der seit August 2006 verbindlichen neuen deutschen Rechtschreibung.

Die Nichtbeachtung, ein Verstoß gegen orthographische Regeln, wird nicht selten sanktioniert. Schreiben wird zum Rechtschreiben, zum Schreiben nach dem Recht!

Die Beherrschung der Orthographie gilt, wie Gerhard Augst ironisch feststellt, „als Ausweis für Intelligenz [...], als Gradmesser der schulischen Sozialisation“.<sup>3</sup> Demzufolge scheint der jeweilige Bildungserfolg ausschließlich vom normgerechten Umgang mit schriftlichen Texten abzuhängen. Dass der normgerechte Umgang mit schriftlichem Textmaterial Menschen mit deutscher Muttersprache naturgemäß leichter fällt, scheint nur wenige zu stören.

Unter dieser Perspektive ist anzunehmen, dass Standardisierungsbemühungen in Verbindung mit der Erfüllung von Testgütekriterien wie Zuverlässigkeit, Vergleichbarkeit, aber auch Objektivität sich anscheinend als Hemmschuhe für die Erstellung offenerer handlungsorientierter Aufgabenstellungen innerhalb von Sprachtests erweisen und so seitens der Testinstitute aufgrund pragmatischer Überlegungen oft auf herkömmliche Prüfungsformate zurückgegriffen wird. Unter diesen Voraussetzungen gilt die Beurteilung von freien produktiven schriftlichen Leistungen nach wie vor als problematisch und schwierig. Gemeinsame Bezugssysteme wie allgemeingültige Diagnosemittel fehlen, vor allem für MuttersprachlerInnen und bei Menschen mit nichtdeutscher Muttersprache, was erfahrungsgemäß auch zu Lasten von kommunikativen und pädagogischen Ansprüchen geht.<sup>4</sup>

Angesichts dieser Tatsachen ist festzustellen, dass sowohl individuelle Unterschiede beim Spracherwerb, insbesondere nichtlineare Entwicklungen von Sprachfähigkeiten bei Menschen mit nichtdeutscher Muttersprache, als auch motorische Schwächen in den Testinhalten zu wenig berücksichtigt werden. Orthographische und grammatische Schwächen sind oft ausschlaggebend für einen misslungenen Berufseinstieg.<sup>5</sup>

---

<sup>3</sup> Augst, Gerhard (2004): Orthografie/Orthography. In: Ammon Ulrich et al. [Hrsg.]: Soziolinguistik: Ein internationales Handbuch zur Wissenschaft und Sprache und Gesellschaft. 2., vollständig neu bearbeitete und erweiterte Auflage. Berlin/New York: de Gruyter, S. 648.

<sup>4</sup> vgl. dazu Glaboniat (2009), S. 41, S. 50.

<sup>5</sup> vgl. dazu Krumm, Hans-Jürgen (2005): Sprachtests: Politische und ethische Dimension. In: Bolton, Sybille: Probleme der Bewertung bei den Produktiven Fertigkeiten in den Deutschprüfungen des Goethe-Instituts. In: Kühn, Ingrid / Lehker, Marianne / Timmermann, Waltraud [Hrsg.]: Sprachtests in der Diskussion. Reihe: Wittenberger Beiträge zur deutschen Sprache und Kultur - Band 4. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Lang S. 16 – 20.

Zu den Prüfmitteln, die ausschließlich Verstöße gegen die sprachliche Norm beurteilen und in keiner Weise die Komplexität des Schreibens berücksichtigen, lässt sich auch der schriftliche Teil der Zulassungsprüfung Deutsch beim Bundesheer, ein Diktat, bestehend aus zehn Sätzen, zuordnen.

Ausgelöst wurde mein Interesse an einer Evaluierung der Zulassungsprüfung Deutsch zum einen durch die Auswertung von Prüfungsdaten bei schriftlichen Leistungen von Soldaten mit nichtdeutscher Muttersprache beim Österreichischen Bundesheer und zum anderen durch meine Lehrtätigkeit bei Deutsch-Förderseminaren. Bei der Zulassungsprüfung Deutsch traten Ergebnisse auf, die vermuten ließen, dass insbesondere durch das Prüfungsformat „Diktat“ Menschen mit Migrationshintergrund und Menschen mit motorischen Schwächen benachteiligt werden. Außerdem zeigte sich, dass die Situation und Zeitknappheit beim Diktat so manchen Testteilnehmer bei der Erfüllung der Aufgabe hemmte. Insgesamt entstand für mich der Eindruck, dass der schriftliche Teil der Zulassungsprüfung weder den sprachlich erwartbaren berufsspezifischen Anforderungen entspricht, noch ausreichend Rückschlüsse auf schriftsprachliche Kompetenzlücken zulässt.

#### Einordnung der Arbeit – Forschungsgegenstand – Ziel der Studie

Spätestens seit der Institutionalisierung des Österreichischen Sprachdiploms Deutsch 1994 wurden die Schwächen von ausschließlich formal orientierten Testverfahren im Bereich Deutsch erkannt, wo in erster Linie eine Überbewertung orthographischer Fehler erfolgte, wenn auch zum damaligen Zeitpunkt nur für Deutsch als Fremdsprache. Mit dem ÖSD wurde ein mehrstufiges Prüfverfahren für Deutsch als Fremdsprache entwickelt, das nicht alleine die Norm, sondern die Annäherung an reale Sprachverwendungssituationen zum Ziel hatte. Der „Kommunikative Ansatz“, der auf Unterrichtsebene bereits Fuß gefasst hatte, sollte nun auch auf der Testebene durch *Messen und Überprüfen von kommunikativer Sprachhandlungsfähigkeit* deutliche Spuren hinterlassen.<sup>6</sup>

In den Jahren der internationalen Anerkennung des ÖSD, aber nicht zuletzt auch durch die negativen Ergebnisse der PISA-Studie, fanden zahlreiche Evaluierungen und Validierungen von bereits erprobten Testverfahren, aber auch von neu entwickelten Sprachprüfungen statt. Aus bildungspolitischer Sicht geht es seit diesem Perspektivenwechsel nicht nur darum, die Qualität des Sprachtests nach den herkömmlichen Qualitätskriterien zu beurteilen, sondern die Testformate unter der Berücksichtigung von Zweisprachigkeit und möglichen

---

<sup>6</sup> vgl. dazu Glaboniat, Manuela: Kommunikatives Testen im Bereich Deutsch als Fremdsprache. Eine Untersuchung am Beispiel des Österreichischen Sprachdiploms. Innsbruck-Wien: Studienverlag 1998. S. 7 - 8.

Förderentscheidungen zu untersuchen. Wesentlich dabei sind der Aspekt der Fairness sowie die Auswirkungen des Testens auf das Lehr- und Lernverhalten.<sup>7</sup>

Unter Berücksichtigung dieser Entwicklungen und der daraus formulierten Richtlinien für Testersteller und Testanbieter wurden im Zuge von Sitzungen der Arbeitsgruppe „Prüfmittel Deutsch“ unter Verantwortung des Sprachinstituts des Bundesheeres Überlegungen angestellt, die vorhandene „Zulassungsprüfung Deutsch“, insbesondere den schriftlichen Teil dieser Prüfung, anhand einer empirischen Untersuchung zu evaluieren und modifizieren. Grund dafür waren nicht nur die neuen Qualitätsanforderungen an Sprachtests, sondern auch die immer höher werdende Anzahl von Testteilnehmern mit nichtdeutscher Muttersprache. Nicht zu vergessen sind hier die Veränderungen innerhalb der bundesheerinternen Rahmenbedingungen, wie kleinere Probandengruppen sowie der geringere Zeitdruck bei der Auswertung.

In diesem Sinne soll meine Untersuchung auch als schreibdidaktischer Prozess verstanden werden, der vom regelgeleiteten, normbewussten Schreiben hin zu handlungsorientierten Schreibkonzepten führen soll. Sie soll den Stellenwert des Schreibens nicht nur bei der Zulassungsprüfung, sondern auch in der berufsbedingten alltäglichen Kommunikation aufwerten.

Primäres Ziel dieser Untersuchung ist es, anhand der Evaluierungsergebnisse die Grundlagen für ein neues, sach- und kriterienorientiertes und vor allem objektivierendes Prüfmittel für den schriftlichen Teil der Zulassungsprüfung zu schaffen, um einen Beitrag zur Verbesserung der beruflichen Einstiegschancen von MigrantInnen und Menschen mit motorischen Schwächen (Schreiben, Hören) zu leisten.

Übergeordnet geht es bei der Evaluierung der „Zulassungsprüfung Deutsch“ um die Bestimmung und Eingrenzung des testrelevanten Bezugssystems, in unserem Fall um die fachlich-berufliche Kommunikation im militärischen Kontext unter besonderer Berücksichtigung berufsbezogener Themen und berufsinterner Textsorten.

Nebenziel der Evaluierung ist es, das Prüfungsformat „Diktat“, das gegenwärtig nur mehr als Relikt bzw. als Kompromiss in standardisierten Prüfverfahren vorzufinden ist, in Hinblick auf systematisches Sprachenlernen unter veränderten gesellschaftlichen Rahmenbedingungen zur Diskussion zu stellen und durch ein neues Prüfungsformat zu ersetzen.

---

<sup>7</sup> vgl. dazu Krumm, Hans-Jürgen (2005): Was kann eine Sprachdiagnostik bei Kindern und Jugendlichen mit Migrationshintergrund leisten? In: Gogolin, Ingrid/Neumann, Ursula/Roth, Hans-Joachim [Hrsg.]: Sprachdiagnostik bei Kindern und Jugendlichen mit Migrationshintergrund. Dokumentation einer Fachtagung am 14. Juli 2004 in Hamburg. Münster: Waxmann S. 97 – 107.

# 1 Einführung in das Fachgebiet

Zur allgemeinen Strukturierung und Präzisierung der Untersuchung stütze ich mich bei allen Forschungsschritten auf untersuchungsleitende Fragestellungen, die den Einstieg zu den jeweiligen Untersuchungsschritten liefern sollen.

## 1.1 Sprachausbildung beim Bundesheer

Das Ziel und der Zweck der Sprachausbildung beim ÖBH ist das Herstellen der sprachlichen Interoperabilität des Bundesheeres (Fähigkeit zur Zusammenarbeit in sprachlicher Hinsicht) für nationale und internationale Einsätze und die Sicherstellung der Aufgabenerfüllung am Arbeitsplatz sowohl in muttersprachlicher als auch in fremdsprachlicher Hinsicht.<sup>8</sup>

Neben den Schwerpunkten Deutsch, Englisch und Französisch als Führungs- und Informationssprachen steht das Erlernen von Einsatzraumsprachen (Balkan, Naher Osten) und Nachrichtensprachen (Informationssprachen) wie Arabisch, Chinesisch, Persisch, Russisch, aber auch Spanisch, im Vordergrund der Ausbildung. Zusätzlich dient die Sprachausbildung der Erhaltung und Pflege von Nachbarschaftssprachen. Die bedarfsorientierte Sprachausbildung betrifft nicht nur Berufssoldaten, sondern auch zivile Angehörige des ÖBH. Neben allgemeinsprachlichen Lernzielen, die sich bei Fremdsprachenkursen aus den Vorgaben der *STANAG-6001-Dokumente*<sup>9</sup> in Verbindung mit dem *Anforderungskatalog Kompetenzen für Einsatz- und Friedensfunktionen im ÖBH 2010*<sup>10</sup> ergeben, dient die Sprachausbildung in erster Linie der Schaffung und Sicherung der militär- und streitkräfterelevanten Fachsprache(n), um das Militärpersonal auf die Aufgabenerfüllung im In- und Ausland vorzubereiten. Dabei gilt es, das Verstehen und Anwenden von Fachtermini entsprechend der Fach- und Führungsfunktion der Militärperson zu schulen.<sup>11</sup>

Nicht zu vergessen ist hier Sprachausbildung in „Deutsch als Fremdsprache“, die innerhalb des ÖBH einen besonderen Stellenwert einnimmt. Auf diesem Gebiet kooperiert das Sprachinstitut des ÖBH (SIB) mit diversen internationalen Dienststellen. Es geht hier in erster Linie um ein interkulturell geprägtes militärisches Austauschprogramm, bei dem vor allem ausländische Berufsoffiziere die Möglichkeit erhalten, in Form eines Sprachkurses in Österreich Deutsch zu lernen und umgekehrt. Als Leitlinie gilt neben fachlichen Inhalten vor

---

<sup>8</sup> vgl. dazu Ausbildungsgrundlage Förderseminar Deutsch S. 3.

<sup>9</sup> Standardization Agreement 6001 (Language Proficiency Levels) ist ein von Österreich im Rahmen der Partnerschaft für den Frieden ratifiziertes Abkommen mit den NATO-Vertragsstaaten, welches die Ebenen der Sprachkenntnisse (Sprachprüfungen) innerhalb der NATO und PfP (Partnership for Peace) regelt und diese damit gegenseitig anrechenbar macht.

<sup>10</sup> Anforderungskatalog Kompetenzen für Einsatz- und Friedensfunktionen im ÖBH 2010.

<sup>11</sup> vgl. dazu Durchführungsbestimmungen für die Sprachausbildung 2010. S. 5 – 6.

allem „Interkulturelle Kompetenz“. Das Angebot, Deutsch zu lernen wird auch von zahlreichen in Österreich stationierten Militärattachés bzw. von Personen aus diversen Botschaften in Anspruch genommen.

Während sich die Fremdsprachenausbildung nach dem Leistungsstufensystem gemäß *STANAG 6001* richtet, wird bei den Deutschfördermaßnahmen für österreichische Soldaten neben dem erwähnten Anforderungskatalog für Berufssoldaten das Pflichtschulniveau der 8. Schulstufe (*seit 2011 Bildungsstandards Deutsch der 8. Schulstufe*) als Grundlage herangezogen. Das betrifft sowohl Muttersprachler als auch Menschen mit nichtdeutscher Muttersprache.

## 1.2 Militäreigene theoretische Grundlagen

Welche heeresinternen Vorgaben bzw. Vorschriften wirken sich auf die Entwicklung und Erstellung des Testformats aus?

Gleichermaßen relevant für die Evaluierung der Zulassungsprüfung Deutsch sind jene militäreigenen Grundsatzdokumente, die als Grundlage und Voraussetzung für die Auswahl und Erstellung von Prüfmaterialien gelten. Neben den *Durchführungsbestimmungen für die Sprachausbildung* und der *Prüfungsordnung für Sprachprüfungen im Österreichischen Bundesheer* sind insbesondere der *Anforderungskatalog Sprachwesen*, aber auch der *Anforderungskatalog Kompetenzen für Einsatz- und Friedensfunktionen im ÖBH 2010* von besonderer Bedeutung.

Der Anforderungskatalog Sprachwesen beschreibt die Kernprozesse des Sprachwesens, die Einteilung der Sprachen, die Kompetenzen und Erfordernisse sowie das Treffen von Ableitungen davon. Er schreibt somit als Querschnittsmaterie die Fähigkeiten oder Kompetenzen von Personengruppen fest.<sup>12</sup>

Dieser Anforderungskatalog legt sprachliche Mindestniveaus in Englisch und Deutsch für Offiziere, Unteroffiziere, Chargen und Zivilbedienstete anhand von Qualitätsstufen fest.

In Hinblick auf die Zulassungsprüfung Deutsch für Unteroffiziersanwärter bedeutet das, dass möglichst frühzeitig jene vorgegebenen Mindestanforderungen erreicht werden sollten, um eine zukünftige funktionsbezogene Verwendung sicherzustellen.

Es soll daher auch möglichst früh erkannt werden, ob entsprechende Fördermaßnahmen erforderlich sind oder nicht. Der nachfolgende Ausschnitt aus dem Kompetenzregister soll

---

<sup>12</sup> vgl. dazu Anforderungskatalog Sprachwesen S. 3 - 4.

zeigen, welche Leistungsprofile sich für die jeweilige Qualifizierungsstufe ergeben. Betroffen sind davon Chargen und Unteroffiziere mit Mutter- bzw. Zweitsprache Deutsch.

QualStu (Qualifizierungsstufe)	Beschreibung QualStu	Sprache	Zulassung	Abschluss
EP/HPA (Einstufungsprüfung Heerespersonalamt)		Deutsch	Sichere Anwendung der Grundkenntnisse der deutschen Sprache in Wort und Schrift auf Pflichtschulniveau in Bezug auf Lesbarkeit und akustische Verständlichkeit	
1	TrpKdt, (Truppenkommandant) StvGrpKdt (stellvertretender Gruppenkommandant)	Deutsch		Sichere Anwendung der Grundkenntnisse der deutschen Sprache in Wort und Schrift auf Pflichtschulniveau in Bezug auf Lesbarkeit und akustische Verständlichkeit
2	GrpKdt (Gruppenkommandant)	Deutsch	Sichere Anwendung der Grundkenntnisse der deutschen Sprache in Wort und Schrift auf Pflichtschulniveau in Bezug auf Lesbarkeit und akustische Verständlichkeit	Sichere Anwendung der deutschen Sprache in Wort und Schrift in Bezug auf Lesbarkeit und akustische Verständlichkeit einschließlich militärischer Fachsprache
Übernahme Lebenslaufbahn	GrpKdt, Fach-UO (Fachunteroffizier)	Deutsch	Sichere Anwendung der deutschen Sprache in Wort und Schrift einschließlich des Verstehens komplexer militärischer Sachverhalte	Sichere Anwendung der deutschen Sprache in Wort und Schrift einschließlich des Verstehens komplexer militärischer Sachverhalte
3	ZgKdt, StbUO, Fach-UO (Zugskommandant, Stabsunteroffizier, Fachunteroffizier)	Deutsch	-	Sichere Anwendung der deutschen Sprache in Wort und Schrift einschließlich des Verstehens komplexer militärischer Sachverhalte
4	Fach-UO, LUO, StbUO (Lehrunteroffizier)	Deutsch	Sichere Anwendung der deutschen Sprache in Wort und Schrift einschließlich des Verstehens komplexer militärischer Sachverhalte	Sichere Anwendung der deutschen Sprache in Wort und Schrift einschließlich Militär- und Verwaltungssprache bzw. Berufsreifeprüfung Deutsch

(Abbildung 1: Auszug aus dem Anforderungskatalog für Sprachwesen S. 12 – 13)

Für den Berufsunteroffizier sind grundsätzlich vier Qualifizierungsstufen vorgesehen, wobei sich die dritte und vierte Stufe jeweils nach den Arbeitsplatzbeschreibungen richtet. Ab der Ebene Stabsunteroffizier, also ab der dritten Qualifizierungsstufe, soll mit Beginn 2013 die Möglichkeit bestehen, die Berufsreifeprüfung in Deutsch abzulegen. Wie in der Tabelle ersichtlich, richtet sich das Zulassungsniveau nach dem Pflichtschulniveau bzw. nach den Bildungsstandards der 8. Schulstufe.

Eine genaue Ausdifferenzierung der Leistungsprofile findet sich im Anforderungskatalog „Kompetenzen für Einsatz- und Friedensfunktionen im ÖBH 2010“.

Der Anforderungskatalog beschreibt – abgeleitet aus definierten Fähigkeiten von Organisationseinheiten – die zur Auftrags Erfüllung in Einsätzen bzw. im Frieden erforderlichen ebenen- und/oder funktionsbezogenen Kompetenzen von SoldatInnen und Zivilbediensteten des Ressorts zur Sicherstellung der Handlungskompetenz im Beruf.<sup>13</sup>

Die bereits erwähnten vier Qualifizierungsstufen finden sich dort in präzisierenden Kompetenzbeschreibungen wieder, wo sprachliche Kriterien sowohl explizit als auch implizit beschrieben werden. Diese Kompetenzbeschreibungen lauten in Kürze wie folgt:

<sup>13</sup> vgl. dazu Anforderungskatalog Kompetenzen für Einsatz- und Friedensfunktionen im ÖBH 2010 S. 5.

- Aktivitäts- und Handlungskompetenz umfasst Dispositionen einer Person, aktiv und in der Gesamtheit selbstorganisiert zu handeln und dieses Handeln auf die Umsetzung von Absichten, Vorhaben und Pläne zu richten.
- Sozial-Kommunikative Kompetenz umfasst Dispositionen einer Person, kommunikativ und selbstorganisiert zu handeln, sich daher auch mit anderen kreativ auseinanderzusetzen.
- Personale Kompetenz umfasst Dispositionen einer Person, reflexiv selbstorganisiert zu handeln, daher sich selbst einzuschätzen.

Die Ausdifferenzierung der Teilkompetenzen erfolgt auf der Basis von drei Lernaspekten:

Kognition (Wissen), Psychomotorik (Tätigkeit) und affektiver Aspekt (Einstellung).<sup>14</sup>

Während für die Qualifizierungsstufe 2 (GrpKdt) innerhalb der *Fach- und Methodenkompetenz – Affektiver Aspekt* z. B. das „*Engagement zur Führung von Gesprächen mit den Soldaten über militärfachliche oder allgemeinpolitische Inhalte*“ angeführt wird, finden sich innerhalb der *Sozial-kommunikativen Kompetenz – Kognitiver Aspekt* ähnliche Beschreibungen wie im Sprachanforderungskatalog wieder, wie z. B. die „*Sichere Anwendung der Grundkenntnisse der deutschen Sprache einschließlich des sinnerfassenden Lesens auf Pflichtschulniveau in Bezug auf Lesbarkeit und Verständlichkeit in Wort und Schrift einschließlich militärischer Fachsprache*“. Beachtenswert sind auch die Ausdifferenzierungen im Bereich *Personale Kompetenz-Kognitiver Aspekt* wie z. B. das „*Erkennen der Bedeutung des lebenslangen Lernens*“.<sup>15</sup>

Wenn auch diese berufsorientierten Kompetenzbestimmungen nicht alle innersprachlichen und außersprachlichen Faktoren abdecken, liefern sie doch zusammen mit den Bildungsstandards einen nützlichen Rahmen für die Einordnung berufstypischer sprachlicher Aktivitäten.

### 1.3 Bildungsstandards, Referenzrahmen und Sprachrealität

Entspricht das Diktat dem angestrebten Sprachniveau? Welche Einflüsse haben bildungspolitische Vorgaben wie die Bildungsstandards der 8. Schulstufe auf die Testerstellung und die Testauswertung?

#### Schreibusus und Norm

Die Bildungsstandards Deutsch 8. Schulstufe sind spätestens seit der Neufassung im Oktober 2006 in den allgemeinen Bildungsstandards festgelegt und gelten verbindlich für den gesamten Pflichtschulbereich in Österreich ab 2011.

<sup>14</sup> vgl. dazu Anforderungskatalog Kompetenzen für Einsatz- und Friedensfunktionen im ÖBH 2010 S. 5 – 6.

<sup>15</sup> vgl. dazu Anforderungskatalog Kompetenzen für Einsatz- und Friedensfunktionen im ÖBH 2010 S. 41, S. 43.

Der Einführung der Bildungsstandards liegt der Leitgedanke zugrunde, im österreichischen Schulsystem mehr Verbindlichkeit anzustreben und grundlegende Kompetenzen bei allen Schülerinnen und Schülern sicherzustellen. Ihre Entwicklung folgt einem im gesamten deutschsprachigen Raum eingeleiteten Paradigmenwechsel, der Nachhaltigkeit und Ergebnisorientierung ins Zentrum der Unterrichtsentwicklung stellt.<sup>16</sup>

Die Funktion von Bildungsstandards liegt in erster Linie nicht in der Steuerung von Lehrplänen, sondern in der Festlegung von Zielen, welche die Schulen zu erreichen haben.

Sie sollen daher aufzeigen was die Schüler nach Beendigung einer Lernperiode, in unserem Fall der 8. Schulstufe, tatsächlich wissen und können. Es handelt sich dabei sowohl in Deutschland als auch in Österreich um *Regelstandards*, die ein durchschnittliches Anforderungsniveau für die vier Kompetenzbereiche *Zuhören* und *Sprechen*, *Lesen*, *Schreiben* sowie für das *Sprachbewusstsein* wiedergeben, um eine curriculare Kohärenz in den betroffenen Schulen zu ermöglichen.<sup>17</sup>

**Bildungsstandards Deutsch, 8. Schulstufe**



**Einsicht gewinnen in Struktur, Norm und Funktion der Sprache als Voraussetzung für Textverstehen, wirkungsvollen Sprachgebrauch und gelungene Kommunikation.**

1. Text- und Satzstrukturen kennen und anwenden
2. Wortarten und Wortstrukturen kennen und anwenden
3. Über einen differenzierten Wortschatz verfügen und sprachliche Ausdrucksmittel situationsgerecht anwenden
4. Grundlegende Regeln der Rechtschreibung und Zeichensetzung anwenden
5. Bedeutung und Wirkung von Sprache in geschriebenen und gesprochenen Texten unter Berücksichtigung eines mehrsprachigen Umfeldes reflektieren

(Abbildung 2: Bildungsstandards der 8. Schulstufe aus der BIFIE-Homepage: Internet- Zugriff 04. 12. 2011)

Dabei müsste man sich die Frage stellen, ob und in welchem Ausmaß Norm die Voraussetzung für Textverstehen und einer gelungenen Kommunikation sein muss und ob die beschriebenen Ziele nicht über den Bereich Sprachbewusstsein hinausgehen. Bezogen auf die Orthographiekenntnisse im Rahmen des Kompetenzbereichs „Sprachbewusstsein“ heißt es weiter:

„Grundlegende Rechtschreibregeln kennen und anwenden, Satzzeichen richtig setzen“ bzw. „Gebrauchswortschatz einschließlich gängiger Fremdwörter orthografisch beherrschen“.<sup>18</sup>

<sup>16</sup> Auszug aus der Homepage des Bundesinstituts für Bildungsforschung, Innovation & Entwicklung des Österreichischen Schulwesens (BIFIE): Internet-Zugriff 04. 12. 2011

<sup>17</sup> vgl. dazu BIFIE-Homepage: Internet-Zugriff 04. 12. 2011

<sup>18</sup> Auszug aus der BIFIE-Homepage: Internet-Zugriff 04. 12. 2011

Inwieweit es sinnvoll bzw. realistisch ist, dass ein Schüler den Gebrauchswortschatz einschließlich gängiger Fremdwörter orthographisch beherrschen soll, sei hier angemerkt.

Mit der Einführung der Zentral-Matura und der endgültigen gesetzlichen Verankerung der Bildungsstandards sollte jedenfalls der Deutschunterricht im Pflichtschulbereich auf einem möglichst einheitlichen Niveau erfolgen. Um dieses Ziel auch zu erreichen, wurden Diagnoseinstrumente zur informellen Kompetenzmessung entwickelt, um den Lehrpersonen entsprechende Diagnosewerkzeuge zur Überprüfung des Sprachstands und zur Erhebung eines möglichen Förderbedarfs zur Verfügung zu stellen. Ob diese Prüfwerkzeuge tatsächlich das halten, was sie versprechen, sei an diesem Beispiel kurz demonstriert:

IKM D7 Sprachbewusstsein

Antwortformat: 2 aus 5  
Kompetenzfeld: Über Rechtschreibbewusstsein verfügen  
Kompetenz/Deskriptor: 50. Schüler/innen beherrschen grundlegende Regeln der Dehnung, der Schärfung, des Stamprinzipis, der Groß- und Kleinschreibung und können diese beim Schreiben anwenden

17) Wie findest du heraus, ob man das Wort groß- oder kleinschreibt?

**K/k**abel

Klicke die beiden Kästchen neben den *zwei* richtigen Antworten an.

- Ich bilde den Plural (die Mehrzahl).
- Ich bilde den Infinitiv (die Nennform).
- Ich setze den Artikel davor.
- Ich spreche mir das Wort vor.
- Ich verlängere das Wort.

Lösung:

- Ich bilde den Plural (die Mehrzahl).
- Ich setze den Artikel davor.

(Abbildung 3: Diagnoseinstrumente zur informellen Kompetenzmessung (IKM) – Sprachbewusstsein für die 6./7. Schulstufe aus der BIFIE-Homepage: Internet-Zugriff 04. 12. 2011)

Bei den Diagnoseinstrumenten handelt sich, wie in diesem Beispiel, zumeist um geschlossene Aufgaben wie MC-Tests, die meiner Ansicht nach ausschließlich Wissensbestände im Bereich Sprachwissen und nicht Sprachkönnen diagnostizieren und somit auch keine Förderoptionen im Bereich Sprachkompetenz zulassen. Offensichtlich haben sich hier, wie schon im BESK-Verfahren<sup>19</sup> für Kindergärten, ökonomische Überlegungen gegenüber pädagogischen Konzepten durchgesetzt.

Wenn auch die zentral formulierten Bildungsstandards, so die Verfasser, lediglich der Orientierung dienen und ausschließlich Grobziele vorgeben, so entsteht doch der Eindruck, dass es hier vielmehr um Angleichungsmaßnahmen an international gültige Prüfungsnormen geht. Dieser Annahme zur Folge müsste ein Schüler, ähnlich wie in diversen überregionalen Sprachprüfungen, in allen Kompetenzbereichen und Fertigkeiten den gleichen Standard

<sup>19</sup> BESK-Verfahren: Beobachtungsbogen zur Erfassung der deutschen Sprachkompetenz 4½- bis 5½ -Jähriger

erreichen. Demnach scheint sich das Bildungswesen offenbar nicht nur von der Schulautonomie sondern auch von der Lernerorientierung zu verabschieden. Insofern bleibt nur zu hoffen, dass sich der Unterricht bzw. auch das Testen nicht nur auf messbare Inhalte wie zum Bsp. die Rechtschreibnorm beschränkt.<sup>20</sup>

Wie die Praxis zeigt, werden Niveau- bzw. Kompetenzbeschreibungen wie z. B. beim Gemeinsamen Europäischen Referenzrahmen für Sprachen, gerne als politisches Mittel zur Legitimierung von Prüfungen verwendet. Als repräsentatives Beispiel sei hier nur der Sprachtest für Zuwanderer erwähnt, dessen Funktion nach Expertenmeinung dahin geht, eine Barriere für die Integration von Zuwanderer zu schaffen.<sup>21</sup>

Testinstitute sind daher angehalten, entsprechende Instrumente für die Erfassung von lese- und schriftsprachlichen Kompetenzen zu entwickeln. Denn die Grundlage jeder sinnvollen Förderung von sprachlichen Fähigkeiten kann nur in einer diagnostischen Erfassung der Sprachkompetenz in Verbindung mit geeigneten Aufgabenszenarien liegen, um der Sprachrealität, insbesondere jener der MigrantInnen möglichst nahe zu kommen.

Die Sprachrealität scheint zumindest bei einigen Schulen, die aufgrund der wachsenden Migrantenzahlen zusehends von bilingualen aufwachsenden Lernenden ausgehen, wahrgenommen zu werden. Diese veränderte Perspektive zeigt sich in einem ganzheitlichen Ansatz in Verbindung mit einem sprachenfreundlichen Umfeld, wo in einem Schulterschluss von DaM- und DaZ-Didaktik Sprachförderung für alle in einem mehrsprachigen Umfeld als Bildungskomponente angestrebt wird.<sup>22</sup>

Ob sich dieses Konzept auch auf die sprachliche Ausgangssituation von TestteilnehmerInnen, für die es um Nachweise in Qualifizierungskontexten wie beim ÖBH geht, nachhaltig positiv auswirkt, werden die nächsten Jahre zeigen.

Für die Zusammenführung und Erfassung der lese- und schriftsprachlichen Kompetenzen von Muttersprachlern und Menschen mit nichtdeutscher Muttersprache innerhalb von Testformaten ist daher noch viel Forschungsarbeit zu leisten.

---

<sup>20</sup> vgl. dazu Krumm, Hans-Jürgen (2006): Prüfungen und Bildungsstandards weltweit: die Globalisierung erreicht den Sprachunterricht. In: ÖDaF-Mitteilungen 2/2006. S. 67 – 73.

<sup>21</sup> vgl. dazu z. B. Krumm, Hans-Jürgen (2009): Sprachenpolitik und Mehrsprachigkeit: Von der Mehrsprachigkeitsrhetorik zur (nicht mehr ganz so traurigen?) Realität. In: Visionen. Gegenwart und Zukunft von DaF/DaZ in Österreich. ÖDaF-Mitteilungen Sonderheft zur IDT 2009. S. 11 – 12.

<sup>22</sup> vgl. dazu Krumm (2005), S. 99 – 101.

## 1.4 Sprachüberprüfungen und sprachliche Förderung beim ÖBH

In welchem institutionellen Zusammenhang steht die Zulassungsprüfung Deutsch?  
Welche Auswirkungen haben die institutionellen Bedingungen auf das Testkonstrukt bzw. auf die Testergebnisse? Warum wird Deutsch beim ÖBH überhaupt geprüft und gefördert?

Zum einen gilt es, sich Gewissheit zu verschaffen, ob die schulisch erlernten Deutschkenntnisse für die jeweilige Funktion wirklich ausreichen oder nicht. Erfahrungsgemäß sind gerade bei Menschen mit scheinbar nahezu identischen Schulabschlüssen (Pflichtschulabschluss) in Wirklichkeit große Niveauunterschiede in Deutsch auszumachen.

Neben der Sicherstellung von berufsbezogenen sprachlichen Kompetenzen soll die sprachliche Förderung nicht nur dem Militärsystem als solchem, sondern auch den Menschen als Investition für die Zukunft dienen. Das ÖBH soll daher auch als Bildungseinrichtung unter dem Motto „Lebenslanges Lernen“ begriffen werden.

Die Feststellung eines Förderbedarfs richtet sich grundsätzlich nach den Ergebnissen zweier Sprachprüfungen. Es handelt sich dabei zum einen um die *Einstufungsprüfung Deutsch beim Heerespersonalamt*, wo die generelle Eignung zum Berufssoldaten festgestellt wird.

Die Deutschprüfung wird im Rahmen eines umfangreichen Auswahlverfahren (Medizinische Untersuchung, Überprüfung der körperlichen Leistungsfähigkeit, psychologischer Test, Rechentest) durchgeführt. Sie besteht aus zwei Teilen, einem Multiple-Choice-Test, der Kenntnisse und Fertigkeiten im Leseverstehen sowie in den Bereichen Lexik, Grammatik und Orthographie überprüft und einem Diktat, das überwiegend auf Orthographie ausgerichtet ist.

Die Auswertung der Prüfung erfolgt durch heereigene, dafür ausgebildete Berufssoldaten aus dem Bereich des Heerespersonalamtes. Obwohl das Ergebnis der gesamten Eignungsprüfung über die Annahme oder Nichtannahme in den Ausbildungsdienst beim ÖBH entscheidet, dient die Einstufungsprüfung Deutsch lediglich als *Prognoseprüfung*. Dabei soll in erster Linie festgestellt werden, ob der/die Kandidat(in) die sprachlichen Fertigkeiten, die bei der *Zulassungsprüfung Deutsch zum Lehrgang Militärische Führung 2* gefordert werden, ohne bzw. mit gezielter Förderung erreichen kann oder nicht.

Zum anderen geht es um die vom Niveau etwas höher angesetzte *Zulassungsprüfung Deutsch zum Lehrgang Militärische Führung 2*, die zur Feststellung der sprachlichen Kenntnisse und Fähigkeiten in Hinblick auf die Erreichung des Ausbildungsziels (Unteroffizier) dient.

Überprüft wird die sichere Anwendung der Grundkenntnisse der deutschen Sprache in Wort und Schrift in Bezug auf Lesbarkeit und akustische Verständlichkeit einschließlich militärischer Fachsprache mit Schwerpunkt auf Leseverstehen und Diktat.

Die Prüfung besteht aus drei Teilen, einem Diktat, einem Leseverstehensteil und einem Lückentext.

Nach der Durchführung des Diktats folgen Leseverstehen und ein Lückentext.

Der Teil Leseverstehen besteht aus drei Aufgaben:

- Einer *Zuordnungsaufgabe*, wo es darum geht, entsprechende Schlagzeilen fünf kurzen Zeitungsartikeln zuzuordnen.
- Einer *Detailverstehensaufgabe*, wo es darum geht, drei Fragen zu einem kurzen Text zu beantworten. Dabei stehen jeweils vier Antwortmöglichkeiten zur Wahl.
- Einer *Textverstehensaufgabe*, wo Absätze in die richtige Reihenfolge gebracht werden müssen. Dabei soll die thematische und chronologische Gliederung von Einzelaussagen innerhalb eines Textes (Textkohärenz) erkannt werden.<sup>23</sup>

Der Lückentext (Textbausteine) soll die allgemeine Sprachfähigkeit, insbesondere Leseverstehen, Wortschatz, aber auch Grammatik und Orthographie überprüfen. Dabei sind *15 Wortlücken* sinngemäß und sprachlich korrekt zu ergänzen. Bei den Lücken handelt es sich ausschließlich um ganze Wörter und keine Wortteile wie bei einem C-Test.

Für den Leseverstehensteil und den Lückentext stehen für die KandidatInnen maximal 30 Minuten zur Verfügung.

Die Auswertung der Zulassungsprüfung erfolgt durch pädagogisch ausgebildetes Personal des Sprachinstitutes. Während beim Diktat die Fehler und ausgelassenen Wörter von der Auswertungsperson persönlich gezählt werden, erfolgt die Auswertung der Leseverstehensaufgaben und des Lückentextes mittels Auswertungsprogramm.

Die Bewertung der Prüfungsteile erfolgt getrennt. Jeder Prüfungsteil (Diktat, Leseverstehen, Lückentext) muss positiv absolviert werden, um die gesamte Zulassungsprüfung zu bestehen.

Die Zulassungsprüfung gilt allerdings erst dann als abgeschlossen, wenn alle Prüfungsteile in Zuge eines Prüfungstermins positiv absolviert werden. Die Zulassungsprüfung kann generell dreimal wiederholt werden.

Im Gegensatz zur Einstufungsprüfung, wo auch zahlreiche zivile Personen die Deutschprüfung ablegen, sind bei der Zulassungsprüfung ausschließlich bereits im System befindliche Berufssoldaten zu erwarten. Die Anzahl der Prüfungsteilnehmer richtet sich nach der Anzahl der Unteroffiziersanwärter, die jährlich auszubilden sind.

---

<sup>23</sup> vgl. dazu Ausbildungsgrundlage Förderseminar Deutsch 2011

Die Ergebnisse der Zulassungsprüfung bilden die Basis für weitere Fördermaßnahmen. Der Fördersystematik liegt als Kern eine Vertiefung und Ergänzung der schulisch erworbenen Kenntnisse und der vier Fertigkeiten zu Grunde. Die Konzentration liegt aufgrund der Prüfungserfordernisse auf den Bereichen Leseverstehen, Rechtschreibung und Grammatik. Zur Erreichung der erforderlichen Kenntnisse und Fertigkeiten werden nach Feststellen des Förderbedarfs drei Lehrveranstaltungstypen angeboten:

Das *Förderseminar Deutsch (Ausmaß ca. 60 Unterrichtseinheiten)* wird als zweiwöchiger Intensivkurs angeboten, wo durch Gastlehrer und Sprachtrainer nicht nur die Vorbereitung auf die Zulassungsprüfung Deutsch, sondern auch das Erlernen und Anwenden spezifischer Leseformen, Grammatik- und Orthographieregeln sowie das Kennenlernen von Lernstrategien erfolgt. Am Ende eines Förderseminars Deutsch haben die Teilnehmer verpflichtend die Zulassungsprüfung zum MilFü 2 im Sinne einer Seminarabschlussprüfung abzulegen. Diese Prüfung wird durch das Personal des SIB durchgeführt.<sup>24</sup>

Die *Kaderfortbildung Deutsch (Ausmaß nach Bedarf)* erfolgt als Kleingruppenttraining an der eigenen Dienststelle, wo auf den individuellen Förderbedarf der Zielperson eingegangen werden kann. Im Gegensatz zum Förderseminar Deutsch erfolgt hier grundsätzlich keine Abschlussüberprüfung.

*Im Rahmen des Vorbereitungslehrganges/MilFü 2 findet ein Förderunterricht Deutsch an der Heeresunteroffiziersakademie im Umfang von ca. 21 Unterrichtseinheiten mit Hilfe von Gastlehrern statt, wo ähnlich dem Förderseminar in erster Linie auf prüfungsrelevante Themen eingegangen wird. Am Ende des Vorbereitungslehrganges gilt die Verpflichtung, die Zulassungsprüfung Deutsch abzulegen.*<sup>25</sup>

Zu Beginn der jeweiligen Förderausbildung werden grundsätzlich die spezifischen Förderbereiche der Teilnehmer auf Grund der Einstufungsprüfung beim Heerespersonalamt bzw. bereits abgelegten Zulassungsprüfung ermittelt. Falls die Ergebnisse zu weit zurück liegen, werden zu Beginn der Ausbildung Einstufungsprüfungen durchgeführt, um eine möglichst präzise Diagnose zu erhalten.

Eine besondere Rolle bei der Fördersystematik spielen die vom SIB eigens dafür ausgebildeten Sprachtrainer Deutsch<sup>26</sup>, die sowohl für die Förderausbildung (Ablauf, Materialien usw.) als auch für die pädagogische Unterstützung der Gastlehrkräfte zuständig

---

<sup>24</sup> vgl. dazu Ausbildungsgrundlage Förderseminar Deutsch 2011

<sup>25</sup> vgl. dazu Rahmencurriculum MilFü2

<sup>26</sup> Sprachtrainer sind ausgebildete Deutsch Tutoren beim ÖBH.

sind. Zudem fungieren die Sprachtrainer auch als Bindeglied zwischen militärischen Dienststellen und zivilen Gastlehrkräften.

Abschließend sei darauf hingewiesen, dass die Inhalte der Fördermaßnahmen künftig adaptiert werden müssen, wenn sich die Zahl jener Teilnehmer erhöht, die Deutsch nicht als Muttersprache haben. Das bedeutet, dass dann z. B. verstärkt Fertigkeiten wie „Hörverstehen“ und „Mündlicher Gebrauch“ in das Förderprogramm einfließen müssen.

### **1.5 Das Sprachinstitut des Bundesheeres (SIB) als Prüfungsinstitution**

Das Sprachinstitut des Bundesheeres an der Landesverteidigungsakademie in Wien ist eine Institution innerhalb des Verteidigungsressorts, welche im Zusammenwirken mit den zuständigen Fachabteilungen im Bundesministerium für Landesverteidigung und Sport für unterschiedliche Aufgaben verantwortlich ist. Neben *terminologischen Forschungsarbeiten*<sup>27</sup>, die sich überwiegend auf *streitkräfterelevante Fachsprachen*<sup>28</sup> beziehen (*Militärwörterbücher, Sprachfibeln, Grammatiklehrbücher*), Dolmetsch-Einsätzen und Übersetzungen steht die fremdsprachliche und fachsprachliche Weiterbildung aller Berufssoldaten in Englisch sowie in einer zweiten Fremdsprache, meist Französisch im Vordergrund. Es handelt sich dabei zum einen um eine integrierte Sprachausbildung innerhalb militärischer Lehrgänge und zum anderen um Intensiv-, Förder- und Sonderseminare, die vielfach auch dezentral in den Bundesländern stattfinden.

Am Sitz des Sprachinstituts in der Stiftskaserne werden laufend etwa zehn Sprachen unterrichtet und geprüft. Neben Englisch und Französisch stehen Aus- und Weiterbildungen in Spanisch, Italienisch, Russisch, Arabisch, Serbisch, Slowenisch und Ungarisch auf dem Programm. Nicht zu vergessen dabei sind Deutschkurse und Deutschtutorien für ausländisches Militärpersonal. Für die Sprachausbildung ist das Sprachinstitut in mehrfacher Hinsicht zuständig. Neben der Planung, Koordinierung und Durchführung von Sprachausbildungen gehören die Auswahl von Sprachenfachpersonal und Lehrmitteln (Evaluierung und Beschaffung von geeigneten, am Markt befindlichen Lehr- und Lernbehelfen, Wörterbüchern und Fachlexika) ebenfalls zu den institutsinternen Aufgaben

---

<sup>27</sup> Bei der Terminologiearbeit wird unterschieden zwischen normierender und komparativer Terminologiearbeit. Sie ist Teil der sprachwissenschaftlichen Forschung und beinhaltet sowohl das Festlegen von Begriffen als auch das Auffinden von Äquivalenzen zwischen Ausgangs- und Zielsprache. Zur Terminologiearbeit gehört aber auch deren anwenderbezogene Bereitstellung durch beispielsweise Datenbanken, Wörterbücher und Glossare.

<sup>28</sup> Unter streitkräfterelevanter Fachsprache werden jene Terminologien bezeichnet, deren Wortschatz direkt oder indirekt mit dem Militär zusammenhängt, wobei die aus zivilen Bereichen stammenden Termini in der militärischen Sphäre mit einer modifizierten oder gänzlich neuen Bedeutung Verwendung finden.

wie die Herstellung, Verwaltung und Evaluierung von Prüfmaterialien. Insgesamt finden etwa 6000 Sprachprüfungen jährlich statt.

Nicht nur bei den Fremdsprachenprüfungen, sondern auch bei der Zulassungsprüfung Deutsch für Unteroffiziersanwärter lassen sich die Aufgaben des Sprachinstituts keinesfalls auf die Durchführung und Auswertung der Prüfung reduzieren. Neben der Erstellung und Erprobung von Prüfmaterialien steht ebenso die Ausbildung und Schulung der Prüfer im Vordergrund. Darüber hinaus beschäftigt sich seit 2007 eine Arbeitsgruppe ausschließlich mit der Evaluierung und Modifizierung von Prüfmitteln für Deutsch.

## 1.6 Die Arbeitsgruppe Prüfmittel Deutsch

Welche Rolle und welchen Einfluss hat die Arbeitsgruppe (AG) Prüfmittel Deutsch bei Deutschprüfungen, insbesondere bei der Entwicklung und Einführung der Zulassungsprüfung Deutsch?

### Voraussetzungen und Ausgangspunkt der Zulassungsprüfung Deutsch (Rückblick)

Die AG Prüfmittel Deutsch beschäftigt sich grundlegend mit Fragen der Prüfungssystematik innerhalb des Bundesheeres. Das betrifft nicht nur Prüfmittel Deutsch, sondern auch Prüfmittel in Fremdsprachen wie z. B. Englisch. Bei den Teilnehmern der Arbeitsgruppe handelt es sich um Fachpersonal des SIB, Experten aus dem BMLVS, dem Streitkräfteführungskommando, aber auch um Personen aus Prüfungsstellen wie der Heeresunteroffiziersakademie und dem Heerespersonalamt.

In Bezug auf das Bildungs- und Anforderungsprofil der Unteroffiziersausbildung NEU wurden bei einer Besprechung der AG Prüfmittel Deutsch bereits 2007 erste grundlegende Überlegungen angestellt, eine möglichst nachhaltige Prüfungssystematik in Verbindung mit einer bezugsgruppenorientierten Sprachförderung einzuführen.

Aufgrund der Annahme, dass eine lediglich auf die Rechtschreibung konzentrierte Überprüfung in Deutsch zu kurz greift (Überprüfung in Deutsch beschränkte sich auf die Rechtschreibkenntnisse in Form eines Lückendiktats), ergab sich der Auftrag an die AG Prüfmittel Deutsch, durchgängige Prüfmittel zur Überprüfung der Deutschkenntnisse zu erarbeiten, welche alle sprachlichen Fertigkeiten erfassen soll. Das neue Prüfmittel soll sowohl die Fähigkeit zur Sinnerfassung (Hör- und Leseverstehen) als auch Aufgaben zum „sinnschaffenden“ mündlichen und schriftlichen Ausdruck enthalten. Darüber hinaus galt es, entsprechende Fördermaßnahmen wie Kursangebote, Einsätze von Gastlehrern und Sprachtrainern an die Prüfung zu koppeln. Es ging daher nicht mehr darum, KandidatInnen auszuscheiden, sondern vielmehr, das sprachliche Potential und die Willenskraft zur Verbesserung zu erkennen.<sup>29</sup>

---

<sup>29</sup> Auszug aus dem Besprechungsprotokoll der AG Prüfmittel vom 13. 3. 2007

Gemäß Berichterstattungen der AG Prüfmittel Deutsch erfolgte nach einer etwa halbjährigen Erprobungsphase ab Jänner 2008, der Einsatz von neuen Prüfmitteln als Deutsch-Einstufungsprüfung beim Heerespersonalamt. Trotz der Vorgabe, alle vier sprachlichen Fertigkeiten zu überprüfen, wurde auf „Sprechen“ verzichtet, weil die psychologische Testung ohnehin sprachliche Mängel immanent diagnostiziert. Außerdem wurde aus praktikablen Gründen das heute noch verwendete Testformat „Diktat“ anstatt eines eigenen Hörverstehenstest eingesetzt. Beim Diktat wurde zunächst folgender Auswertungsschlüssel eingesetzt (wurde mit 2010 auf zwei Auswertungskategorien reduziert):

bis 7%	Fehler (der Gesamtwörteranzahl):	kein Förderbedarf
7-14%	Fehler:	Förderbedarf
14-21%	Fehler:	großer Förderbedarf
mehr als 21%	Fehler:	sehr großer Förderbedarf

Bei der Fehleranalyse wurde in zwei Auswertungskategorien unterschieden: Fehlende Worte und falsche Worte, die allerdings bei der Eingabe in die Datenbank als Gesamtprozentsatz gewertet werden. Das bedeutet, dass fehlende Worte und fehlerhafte Worte gleich gewichtet wurden. Zusätzlich wurden anfänglich auch Satzzeichen in die Beurteilung miteinbezogen.

Aufgrund der guten Mess- und Fördererfahrungen mit dem neuen Testformat bzw. den ebenfalls nützlichen Testwerten bei nachfolgenden Zwischenprüfungen im Rahmen von UO-Vorbereitungslehrgängen wurde seitens der Arbeitsgruppe beschlossen, ein ähnliches Prüfungsformat, bestehend aus Leseverstehensteil (drei Einzelaufgaben), Lückentext und Diktat für die Zulassungsprüfung Deutsch einzuführen. Die Fördermaßnahmen sollten für die Probanden solange erfolgen, bis alle Einzelteile der Prüfung in einem Zuge positiv abgelegt wurden.

Nach einer Evaluierungs- und Erprobungsphase wurde mit Beginn 2009 die Zulassungsprüfung Deutsch erstmals eingesetzt. Das Schwierigkeitsniveau wurde gegenüber der Deutschprüfung beim Heerespersonalamt etwas angehoben, der Auswertungsschlüssel blieb jedoch vorerst gleich. In weiterer Folge beschränkte man sich jedoch auf leichter handhabbarere Auswertungsrichtlinien (Förderbedarf/kein Förderbedarf). Abgesehen von umfangreichen Fördermaßnahmen hinsichtlich Leseverstehen, Orthographie und Grammatik wurde auch der Umgang mit Legastheniker (nachgewiesene Rechtschreib- und Leseschwäche) festgelegt.

Mit Ende 2010 erfolgte, bedingt durch veränderte Rahmenbedingungen - geringere Personenanzahl, mehr Zeit zur Auswertung bzw. Nicht-Muttersprachler als mögliche Bedarfsträger - ein Sichtwechsel hinsichtlich der Validität der vorhandenen Testformate.

Insbesondere geriet das Diktat aufgrund der scheinbar ständig sinkenden Bestehens-Aussichten, zunehmend unter Kritik. Es galt, wie schon 2006 - angeleitetes freies Schreiben wurde bei der Einstufungsprüfung beim Heerespersonalamt erprobt, aber aufgrund der großen Teilnehmeranzahl und der zu knappen Zeitvorgabe für die Auswertung wieder verworfen - Überlegungen anzustellen, anstelle des Diktats eine berufsbezogene Schreibaufgabe bei der Zulassungsprüfung zu stellen, deren Ergebnis präzisere und sinnvollere Fördermaßnahmen ermöglicht.

In diesem Zusammenhang gab es auch eine Pilotphase, in der eine Bildgeschichte im Rahmen der Zulassungsprüfung vorgesehen war, welche aber aus ökonomischen Gründen verworfen wurde,

## **2 Das Diktat innerhalb der Zulassungsprüfung Deutsch**

Die Bestimmung des Testkonstrukts laut Testersteller lautet: Orthographiekennntnisse auf einen gewissen Niveau abzufragen. Die Frage stellt sich, ob das Diktat diesem Konstrukt entspricht bzw. ob noch andere Fähigkeiten mit dem Diktat erfasst werden und ob diese Erweiterung des Testkonstrukts überhaupt erwünscht ist!  
Welche Interpretationsmöglichkeiten lässt die Auswertung zu?

### **2.1 Funktionalität und Nützlichkeit des Verfahrens**

Die Zulassungsprüfung Deutsch erfolgt im Rahmen der Zulassung zum Unteroffizierslehrgang 2 (Ausbildung zum Gruppenkommandanten) und ist Voraussetzung für zukünftige Verwendungssituationen.

Bei jener Prüfung geht es in erster Linie um die Selektion von Probanden aus einer größeren Gruppe und erst danach um die Zulassung an sich. Das bedeutet, dass dem Schwellenwert, der Grenze zwischen „Bestanden“ und „Nicht Bestanden“ eine große Bedeutung zugemessen werden muss. Die Funktion des Diktats beschränkt sich auf das Abprüfen von verfügbaren normativen Richtigkeitskriterien in Verbindung mit einer angemessenen Schreibfähigkeit. Genaugenommen geht es bei der Zulassungsprüfung um eine definierte sprachliche Mindestleistung, die sich grundsätzlich an den Spezifikationen und Kompetenzbeschreibungen innerhalb des ÖBH orientieren sollte.

Wenn, wie beim Prüfungsformat Diktat, die sprachliche Korrektheit als alleiniges Beurteilungskriterium maßgeblich ist, entscheidet die Anzahl der Rechtsschreibfehler bzw. fehlenden Wörter/Buchstaben über die Annahme/Nichtannahme des Kandidaten/der Kandidatin. Dem zufolge entscheidet nicht wie etwa im europäischen Referenzrahmen für

Sprachen der Grad der Korrektheit der sprachlichen Realisierung über das Bestehen, sondern genauer betrachtet *ein Rechtschreibfehler* über die Zukunft der Testperson.

Hinsichtlich der Testkriterien und der Abgrenzung des Diktats zu anderen Prüfmitteln beziehe ich mich auf die fünf Fragen von Michaela Perlmann-Balme:

- 1) *Worauf bezieht sich der Inhalt der Prüfung?*
- 2) *Für wen sind die Ergebnisse bestimmt (z. B. für die Geprüften, die Lehrkräfte bzw. außenstehende Entscheidungsträger)*
- 3) *Welche Entscheidungen werden auf der Basis der Ergebnisse getroffen? (z. B. Zulassungsprüfung)*
- 4) *Wie aufwendig oder ökonomisch ist das Verfahren im Hinblick auf die Faktoren Zeit, Arbeitsmittel etc.?*
- 5) *Wie aufwendig oder ökonomisch ist das Verfahren im Hinblick auf die vorgesehenen Korrekturverfahren?*<sup>30</sup>

zu 1)

Testgegenstand ist die Überprüfung der „schriftsprachigen Korrektheit“. Der Testinhalt des Diktats ist die direkte Abbildung der Rechtschreibfähigkeit beim Niederschreiben von Gehörtem. Zusätzlich soll das Diktat als Überprüfung der Schreibfähigkeit dienen.

Hinsichtlich der Form spiegelt das Diktat zumindest ansatzweise berufsimmanente Kommunikationsverfahren wieder, wie z. B. das Niederschreiben von Gehörtem bei Einweisungen, Befehlsausgaben bzw. das Notieren von Gehörtem im Fernmelde- und Funkwesen.

zu 2)

Die Testergebnisse sind in erster Linie für jene Entscheidungsträger bestimmt, die über Zulassung, Nicht-Zulassung bzw. über Förderung bestimmen. Diese sind die heeresinterne Prüfungsinstitution (SIB) und die Ausbildungsinstitution (Heeresunteroffiziersakademie). In zweiter Linie sind die Testergebnisse für den Probanden bzw. für die Lehrkräfte (Sprachtrainer, Gastlehrer) bei heeresinternen Fördermaßnahmen vorgesehen.

---

<sup>30</sup> Perlmann-Balme, Michaela (2010): Formen und Funktionen von Leistungsmessung und Kontrolle. In: Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. Herausgegeben von Hans-Jürgen Krumm, Christian Fandrych, Britta Hufeisen und Claudia Riemer. 2 Bände. De Gruyter Verlag: Berlin S. 996.

zu 3)

Bei der Zulassungsprüfung Deutsch geht es um „Förderbedarf“ (über 7% Fehler) und „kein Förderbedarf“ (bestanden). Bei einem Förderbedarf sind entsprechende Maßnahmen wie Förderseminar oder Einzelförderung durch Sprachtrainer bzw. Gastlehrer einzuleiten.

zu 4)

Das Testverfahren „Diktat“ wird laut Testersteller und heeresinternen Experten als weitestgehend ökonomisch, objektiv und vor allem als praktisch bezeichnet.

Naturgemäß treten orthographische Phänomene in authentischen Diktattexten meist in unterschiedlicher Art- und Weise auf wie z. B. die Groß- und Kleinschreibung. Es gilt daher, Texte auszuwählen, die annähernd die gleichen orthographischen Phänomene beinhalten bzw. den gleichen Schwierigkeitsgrad aufweisen. Die Auswahl der Texte beansprucht daher einen nicht zu unterschätzenden Zeitaufwand (Textanalysen, Stichproben, Paralleltests usw.), um nicht nur den Schwierigkeitsgrad abzustimmen, sondern auch die kontextuelle Validität möglichst hoch zu halten. In materieller Hinsicht kommt es zu keinem erheblichen finanziellen Mehraufwand, weil Räumlichkeiten wie Hörsäle bzw. entsprechende Abspielgeräte (CD-Player) ohnehin ausreichend vorhanden sind. Auch die finanzielle Abgeltung der Testdurchführenden stellt aus ökonomischer Hinsicht nur einen geringen finanziellen Mehraufwand dar, weil die Prüfung zumeist in die Normarbeitszeit der Tester fällt.

zu 5)

Die Unabhängigkeit der Ergebnisse kann durch ein abgestimmtes Korrekturverhalten (Wie gehe ich mit Wiederholungsfehlern, Flüchtigkeitsfehlern, Text- und Wortlücken, sowie unleserlichen Textstellen und Auslassungen von Silben und Wörtern um) in Verbindung mit einer vorher festgelegten Gewichtung von Fehlern bzw. der Errechnung eines Fehlerquotienten gewährleistet werden.

## **2.2 Testdurchführung**

Das Diktat erfolgt im Rahmen der Zulassungsprüfung Deutsch und beansprucht etwa 20 Minuten.

Das Diktat wird mittels CD abgespielt, auf welcher zuerst der Text zusammenhängend im normalen Sprechtempo, anschließend langsam zum Mitschreiben Satzteil für Satzteil und abschließend zur Kontrolle nochmals zusammenhängend im normalen Sprechtempo

vorgetragen wird. Die Länge des Diktats besteht aus zehn chronologisch aneinandergereihten Sätzen mit durchschnittlich 163 Wörtern. Es handelt sich zumeist um authentische Textausschnitte bzw. Textzusammenfassungen aus Zeitungsartikeln. Die Schwierigkeit der Texte soll etwa den Anforderungen der 6. bzw. 7. Schulstufe im Unterrichtsfach Deutsch entsprechen. Die Auswahl der Texte erfolgt aufgrund von Stichproben (Probetests) bzw. Textanalysen durch ausgebildetes Fachpersonal.

Das Diktat wird ausschließlich mittels einer akustisch erprobten CD vorgetragen. Die Tonqualität kann daher als sehr gut eingestuft werden. Die Qualität der Abspielung richtet sich naturgemäß nach den örtlichen Begebenheiten bzw. nach möglichen Umgebungsgeräuschen. Als Ansager fungiert eine geschulte Person (ein Mann) mit Stimm- und Sprechausbildung.

Das Diktat wird insgesamt dreimal abgespielt, einmal im normalen Sprechtempo zum Anhören, anschließend betont langsam zum Mitschreiben mit entsprechenden Pausen nach Satzteilen und schließlich ein drittes Mal im normalen Sprechtempo zur Kontrolle bzw. für etwaige Korrekturen/Ergänzungen. Danach stehen noch etwa drei Minuten zur nochmaligen Kontrolle zur Verfügung.

### Prüfungsorte/Prüfungspersonal

Zulassungsprüfungen finden generell im Zuge der Unteroffiziersausbildung an der Heeresunteroffiziersakademie statt. Bei Nichtbestehen z. B. finden die Prüfungen direkt bei der kursführenden Stelle in Verbindung mit Förderseminaren statt. In Einzelfällen besteht die Möglichkeit, die Wiederholungsprüfung direkt am SIB abzulegen.

Die Durchführung und Auswertung der Zulassungsprüfung Deutsch erfolgt ausschließlich durch fachlich ausgebildete Lehrpersonen des Deutschreferats (SIB) mit Unterstützung durch ressortfremdes Fachpersonal in Form von GastlehrerInnen.

## **2.3 Die Zielgruppe bei der Zulassungsprüfung Deutsch**

Die Zulassungsprüfung Deutsch zum MilFü 2 richtet sich an jene Bundesheerangehörigen, die am Anfang ihrer Unteroffizierslaufbahn stehen. Beim Personenkreis handelt es sich zum einen um Männer, die den Grundwehrdienst bereits vollständig absolviert haben und zum Zeitpunkt der Prüfung einen Ausbildungsdienst absolvieren. Zum anderen geht es um Frauen, die sich mit dem Ziel, Unteroffizier zu werden, zum Ausbildungsdienst gemeldet haben.

Insgesamt sind ausschließlich österreichische StaatsbürgerInnen mit Pflichtschulabschluss und positiv absolvierter Eignungsprüfung (Heerespersonalamt) zu erwarten. Erfahrungsgemäß lässt der Pflichtschulabschluss allerdings nur vage Beurteilungen hinsichtlich des schriftsprachlichen Niveaus zu. Wie die Testergebnisse der letzten Jahre belegen, zeigen sich vor allem bei grammatik- und rechtschreiberorientierten Aufgaben große Leistungsunterschiede.

Nicht zu vergessen sind hier jene Personen mit nichtdeutscher Muttersprache, die trotz Pflichtschulabschluss große Schwächen beim Reproduzieren von Texten zeigen. Gründe dafür scheinen die noch nicht ausgereifte Zweitsprachenausbildung in den Schulen und die Normorientierung der Zulassungsprüfung zu sein.

Das durchschnittliche Antrittsalter liegt etwa bei 22 Jahren. In seltenen Fällen kann es auch vorkommen, dass Personen mit über 30 Jahren an der Prüfung teilnehmen. Angesichts dieser Tatsache sollte dabei mitberücksichtigt werden, dass es sich hier auch um Personen handeln kann, die im Zuge ihres schulischen Deutschunterrichts noch nicht mit den neuen Rechtschreibregeln (Rechtschreibreform 2006)<sup>31</sup> vertraut gemacht wurden. Für die schriftliche Zulassungsprüfung bedeutet das, dass die Rechtschreibreform nicht für alle Prüfungsteilnehmer als bekannt vorausgesetzt werden kann.

Bei der Zielgruppe kann erfahrungsgemäß von Personen mit Pflichtschulabschluss ausgegangen werden, die bereits die Eignungsprüfung beim Heerespersonalamt, somit den Einstieg in die soldatische Laufbahn, erfolgreich absolviert haben. Somit fällt unter diese Personengruppe eine beträchtliche Anzahl an Maturanten und Studienabbrechern bzw. gibt es hier aber in Ausnahmefällen auch Menschen ohne Pflichtschulabschluss.

Die Zulassungsprüfung Deutsch muss spätestens vor Beginn des UOLG/MilFü 2 positiv abgeschlossen werden. Sie kann aber, wie in den meisten Fällen, bereits davor im Zuge vorangehender Lehrgänge abgelegt werden.

Gemäß Anforderungskatalog Sprachwesen wurden für die Prüfungsteilnehmer (Unteroffiziersanwärter) folgende Sprachqualifikationen (Kriterien) festgelegt:

Qualifizierungsstufe 2 (UOLG/MilFü 2):

Deutsch – Sichere Anwendung der Grundkenntnisse der deutschen Sprache in Wort und Schrift in Bezug auf Lesbarkeit und akustische Verständlichkeit einschließlich militärischer Fachsprache – Schwerpunkt Leseverstehen, Diktat – Rechtschreibung

Sie entspricht in etwa B2/Europäischer Referenzrahmen bzw. 2+/STANAG 6001<sup>32</sup>

---

<sup>31</sup> Seit August 2006 ist die neue Rechtschreibreform in Kraft.

<sup>32</sup> vgl. dazu Anforderungskatalog Sprachwesen 2008, S. 12 – 13.

Die Vorbereitung auf das Prüfungsformat erfolgt in der Weise, dass die Probanden mit dem jeweiligen Testformat im Zuge von Vorbereitungsphasen, Förderunterricht und spätestens bei Förderseminaren vertraut gemacht werden.

Zusätzlich werden seitens der Dienststellen mit Hilfe von Sprachtrainern ausreichend Übungssequenzen angeboten und Lernmaterial zur Verfügung gestellt.

## **2.4 Bestehensbedingungen – Auswirkungen auf die Zulassungsentscheidung**

Welche Auswirkungen und Folgen hat das Bestehen/Nichtbestehen der schriftlichen Prüfung auf die Zulassungsentscheidung bei der Unteroffiziersausbildung? Handelt es sich dabei um eine ernstzunehmende Prüfung oder lediglich um einen formellen Akt? Liefert das Diktat zusätzliche sprachrelevante Informationen über die Testperson oder geht es dabei ausschließlich um Selektion?

Wenn, wie beim Prüfungsformat Diktat, die sprachliche Korrektheit als alleiniges Beurteilungskriterium maßgeblich ist, entscheidet die Anzahl der Rechtsschreibfehler bzw. fehlenden Wörter/Buchstaben über die Annahme/Nichtannahme des Kandidaten/der Kandidatin.

Generell erfolgt die Bewertung der einzelnen Prüfungsteile innerhalb der Zulassungsprüfung getrennt, wobei jeder Teil als solcher positiv bestanden werden muss, um das Prüfungsfach „Deutsch“ positiv abzuschließen. Ansonsten ist gemäß Bestehensbedingungen eine entsprechende Förderung in Form von Seminaren und Einzeltrainings vorgesehen. Bei einem negativ abgeschlossenen Prüfungsteil, wie z. B. dem Diktat, ist allerdings die gesamte Zulassungsprüfung zu wiederholen.

Die Prüfung kann grundsätzlich dreimal wiederholt werden. In Ausnahmefällen besteht die Möglichkeit, mittels Entscheidung des jeweilig vorgesetzten Kommandos, auch über die festgelegten Bestehensbedingungen hinaus nochmals anzutreten. Angesichts des Übereinkommens, dass bereits ausgebildete leistungswillige Soldaten, wenn möglich, im System zu behalten sind, besteht in Ausnahmefällen die Möglichkeit, mittels Entscheidung des jeweilig vorgesetzten Kommandos, auch über dreimaliges Wiederholen hinaus nochmals zur Prüfung anzutreten.

*Die Gültigkeit einer positiven Zulassungsprüfung ist generell unbegrenzt. Sie endet jedoch spätestens mit Ablauf des Verpflichtungszeitraumes als Berufssoldat!*

Prognosen hinsichtlich des Sprachlernerfolges erfolgen nur insofern, als den Testpersonen bei Nichtbestehen der Prüfung ein Feedback über Art und Anzahl der gemachten Fehler gegeben

und ein Förderprogramm angeboten wird. Eine detaillierte Information über den Lernfortschritt findet nur dann statt, wenn die Prüfung mindestens zweimal negativ absolviert wurde und beim Kandidaten deutliche Verbesserungen auftreten.

### **3 Die Testaufgabe und ihr Umfeld**

#### **3.1 Sprachliche Handlungsfähigkeit im Beruf**

Soll der Test ausschließlich allgemeine schriftsprachliche Fähigkeiten messen oder darüber hinaus fachliche Termini in Form von Fachwissen abprüfen?

Armeen sind zumeist geschlossene Gesellschaftsformen mit eigenen Regeln und spezifischen Merkmalen. Wie in anderen Lebens- und Berufsbereichen hat sich beim Militär ein spezifischer Sprachgebrauch entwickelt, der sowohl auf konkreten Begriffen aufbaut als auch durch spezielle grammatische Eigenheiten gekennzeichnet ist. Einen großen Einfluss dabei nehmen die Charakteristika der militärischen Gesellschaft ein, die vor allem in Abhängigkeit von Hierarchien und normierten Strukturen zu sehen sind.

Das Ziel des vorliegenden Testverfahrens ist nicht nur, die allgemeine schriftsprachliche Fähigkeit zu messen, sondern auch Aussagen über den Grad der Sprachbeherrschung im institutionellen Zusammenhang treffen zu können.

Neben der qualitativen Überprüfung des vorliegenden Testverfahrens, wo die Nützlichkeit der Ergebnisse im Vordergrund steht, gilt es für mich daher auch, auf das Testkriterium, die schriftsprachliche Handlungsfähigkeit im Beruf, näher einzugehen. Bei der sprachlichen Handlungsfähigkeit geht es mir aber weniger um spezifische Begriffe im Bereich des Militärwortschatzes, sondern vielmehr um die Erfassung der tatsächlichen Arbeits- und Sprachwirklichkeit. Dabei geht es um die verschiedenen funktionalen Eigenschaften von Berufs- und Fachsprachen, denen in der Fachsprachenlinguistik grundlegende kommunikative Charaktereigenschaften wie z. B. Generalisierung, Differenziertheit oder Verkürzung zugeschrieben werden. Nicht zu vergessen sind dabei Stilmittel wie Präzision und Deutlichkeit.<sup>33</sup>

Bei den Sprachverwendungssituationen sollte man aber auch die Konsequenzen einer sprachlichen Äußerung (erwartbare kommunikative Handlungsketten) immer mitberücksichtigen. Das heißt, dass die Text- und Mitteilungsfunktion einer sprachlichen

---

<sup>33</sup> vgl. dazu Dannerer, Monika (2008): „Beschreibungsmöglichkeiten der Fach- und Berufskommunikation im Deutschen“. In: ÖDaF-Mitteilungen, Heft 1/2008. S. 24 - 28.

Handlung immer in Verbindung mit der realen Situation (schriftliche Meldung) und dem Adressaten steht (Vorgesetzter, Untergebener) und zumeist als Entscheidungsgrundlage dient. Es könnte daher bei unpräzisen oder unverständlichen Meldungen durchaus zu Fehlentscheidungen kommen. Um diese Gefahr möglichst gering zu halten, ist die Text- und Mitteilungsfunktion bei institutionsinternen Texten wie dem ÖBH klar geregelt, sie untersteht zumeist Rechts- und Dienstvorschriften, die sowohl die Form als auch die Handhabung vorgeben. Das bedeutet, dass aus Sicherheitsgründen der kommunikative Spielraum eines Sprechers/Schreibers durch genormte Druck- bzw. Textsorten deutlich eingeschränkt wird.<sup>34</sup>

Vor diesem Hintergrund stellt sich allerdings die Frage, warum man sich bei einer heeresinternen schriftlichen Prüfung nicht ausschließlich nach dem militärischen Schriftverkehr richtet. Beispielsweise könnte man dabei auch mediale Varianten in Betracht ziehen, wie z. B. den E-Mail-Verkehr. Demnach könnte man auch annehmen, dass bei Institutionen wie dem Österreichischen Bundesheer, die in einer spezifischen kulturellen Umgebung bestehen, der heeresinterne Kontext für konkrete Sprachverwendungssituationen ausreicht.

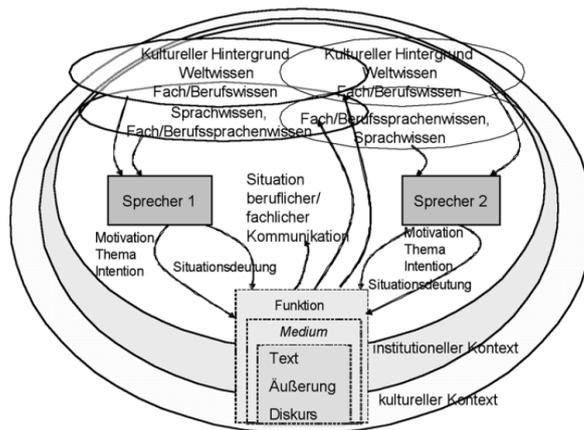
Tatsächlich zeigt aber die Berufspraxis, dass bei der Erfassung von fachkommunikativen Begebenheiten auch im schriftlichen Bereich immer unterschiedliche Wissensbestände und Zielsetzungen zusammenwirken bzw. in Abhängigkeit zueinander stehen.

Wenn man daher davon ausgeht, dass nicht nur das Schreiben von Texten in Verbindung mit Berufswissen, sondern auch allgemeine Schreibkompetenzen in inhaltlicher (Sprach-, Kultur- und Weltwissen) und sprachlicher Hinsicht unter Einbeziehung der operativ-funktionellen Dimension von Sprache relevant sind, sollte das Testkonstrukt der Zulassungsprüfung auch danach ausgerichtet sein. Darüber hinaus gilt, dass der kommunikative Umgang mit Menschen außerhalb des beruflichen Umfelds mitberücksichtigt werden muss (öffentlicher Auftrag des ÖBH) und sich daraus erweiterte kommunikationsspezifische Anforderungen ergeben. Das heißt, dass die berufliche Kommunikation auch über den institutionellen Kontext hinausreichen kann.

Diese mehrfache Einbettung berufssprachlicher Handlungen soll das Modell von Monika Dannerer verdeutlichen:

---

<sup>34</sup> vgl. dazu z. B. Dannerer (2008), S. 30 – 34.



(Abbildung 4: Monika Dannerer – Modell fachlicher und beruflicher Kommunikation. In: ÖDaF-Mitteilungen, Heft 1/2008, S. 23)

Von außen nach lässt sich dieses Modell folgendermaßen beschreiben: Institutionen, die in einer spezifischen kulturellen Umgebung stehen, bilden den Kontext für konkrete Kommunikationssituationen, in denen KommunikationsteilnehmerInnen sprachlich und nonverbal handeln, um ihre Zielsetzungen zu erfüllen. Dabei deuten sie die Situation und greifen auf ihre Wissensbestände zurück: ihr Weltwissen, ihr kulturelles Wissen, ihr berufliches und fachliches Wissen sowie ihr Sprachwissen bzw. ihr fach- und berufssprachliches Wissen. Aufgrund ihrer Situationsdeutung generieren sie entsprechend ihrer Motivationen und Intentionen und ihrer thematischen Mitteilungsabsichten Äußerungen bzw. Diskurse und Texte. [...] Sprachliches Handeln im Beruf ist damit dreifach eingebettet – situational, institutionell und kulturell.<sup>35</sup>

Um eine Basis für ein neues Testverfahren zu schaffen, scheint es notwendig, ein Berufskompetenzmodell auszuarbeiten, wobei sowohl berufsinterne kommunikativ funktionale Schreibfähigkeiten als auch allgemeine schriftsprachliche Fähigkeiten berücksichtigt werden.

#### 4 Sprachtests zur Erfassung von Sprachkompetenzen als Untersuchungsbasis

In Folgenden gebe ich eine Übersicht zu wissenschaftlichen Arbeiten, auf denen die Arbeit aufbaut. Es geht dabei insbesondere um relevante bzw. problembezogene Fachliteratur, die als Materialbasis dienen soll.

Gibt es Literatur zur Fragestellung dieser Untersuchung bzw. welche wissenschaftlichen Untersuchungen führten maßgeblich zu den Hypothesen dieser Arbeit (Hypothesengerüst)? Welchen Einfluss hatte der jeweilige wissenschaftliche Ansatz auf die gewählte Herangehensweise?

<sup>35</sup> Dannerer (2008), S. 23 - 24.

Der PISA-Schock zog nicht nur zahlreiche Fördermaßnahmen in Deutsch mit sich, sondern bewirkte u. a. eine verstärkte Wahrnehmung der mehrsprachigen Wirklichkeit vor allem in den Bildungseinrichtungen deutschsprachiger Länder, im Besonderen eine rege Forschungstätigkeit im Bereich Sprachstandsdiagnostik. Man griff daher erneut die Frage auf: *Wie kann ich kommunikative Sprachhandlungsfähigkeiten und -Fertigkeiten diagnostisch messen und daraus Fördermaßnahmen ableiten?*<sup>36</sup>

Schreib- bzw. Testaufgaben zu gestalten, die den Kriterien der Objektivität, Reliabilität und Validität ebenso entsprechen wie einer realitätsnahen Gestaltung bzw. eine Unterscheidung und eindeutige Abfrage von Teilkompetenzen zulassen, stellt bislang eine nicht vollständig eingelöste Forderung dar. Erfolgreiche Standardisierungen von Testverfahren gelingen offenbar nur um den Preis der Reduktion von Komplexität und Kreativität innerhalb der Aufgaben. Darunter fällt auch der pädagogische Mehrwert einer Prüfung, der meiner Ansicht nach bei strikter Einhaltung der Testgütekriterien zusehends an Bedeutung verliert.<sup>37</sup>

Ein generelles Problem für einen Sprachtestersteller scheint vor allem auch im Bereich der Glaubwürdigkeit des Testkonstrukts zu liegen. Die Frage, in welchem Ausmaß sich die zu messenden Fähigkeiten in den Testergebnissen widerspiegeln bzw. ein direkter, eindeutiger Zusammenhang zwischen Test und Sprachrealität besteht, kann zumeist nur vage beantwortet werden.<sup>38</sup>

Angesichts dieser Tatsachen gleicht die Entwicklung von Sprachtests, zumindest im deutschsprachigen Raum, einem Kurs der Kompromisse zwischen Normorientierung, Standardisierung, Kommunikativer Kompetenz und Glaubwürdigkeit. Nicht zu vergessen sind hier auch bildungs- und migrationspolitische Überlegungen. Trotz dieser eher ernüchternden Fakten gibt es vereinzelt immer wieder erfolgreiche Forschungsprojekte, die sich mit Testen und Prüfen, insbesondere mit Feststellungsverfahren von Sprachkenntnissen beschäftigen. In Zeiten von Migrationsbewegungen und sprachlicher Vielfalt gehen diese Studien zumeist aber nicht von Muttersprachlern, sondern von Menschen mit nichtdeutscher Muttersprache im In- und Ausland aus (DaZ, DaF) aus. Das bedeutet, dass die Anforderungen an zukünftige Test- und Prüfverfahren in erster Linie der Gleichstellung von Mehrsprachigen bzw. Förderung von Mehrsprachigkeit dienen und nicht speziell auf Menschen mit Deutsch als Muttersprache abzielen, was in pädagogischer und sozialer Hinsicht zweifellos als äußerst positiv gesehen

---

<sup>36</sup> vgl. dazu Krumm (2005), S. 98 – 99.

<sup>37</sup> vgl. dazu Glaboniat (2009), S. 41 – 44.

<sup>38</sup> vgl. dazu Glaboniat (1998), S. 26 – 28.

werden kann. Wenn man allerdings davon ausgeht, dass es sich wie bei der Zulassungsprüfung Deutsch überwiegend um MuttersprachlerInnen handelt, deren Deutschunterricht vielerorts noch monolingual ausgerichtet war, scheinen die neuentwickelten Testverfahren in didaktisch- und methodischer Hinsicht nicht immer angemessen zu sein. Warum in Österreich in Zeiten von PISA-Schock und Zentralmatura bislang keine geeigneten, universell einsetzbaren Prüfmaterialien, die sich gleichermaßen für DaM und DaZ eignen, entwickelt wurden, scheint meiner Ansicht nach ausschließlich bildungspolitische Ursachen zu haben.

Um diesen Umstand wusste auch ein Team deutscher SprachforscherInnen und DiagnostikerInnen der *FÖRMIG-Gruppe um Hans. H. Reich*<sup>39</sup>, das sich zwar ausschließlich mit Förderung von Menschen mit Migrationshintergrund beschäftigte, aber gleichermaßen auch Muttersprachler bei den Sprachstandsmessungen mitberücksichtigte. Unter dieser Prämisse wurde meines Erachtens ein sehr geeignetes diagnostisches Verfahren auch für LernerInnen mit Deutsch als Muttersprache entwickelt.

#### **4.1 Bumerang**

Bei „*Bumerang*“ handelt sich um ein sprachdiagnostisches Instrument, das die *Erfassung der Sprachkompetenzen im Übergang von der Schule in den Beruf*<sup>40</sup> zum Ziel hat.

In diesem Zusammenhang ist das HAVAS 5-Verfahren zu erwähnen, wo es um die Ermittlung des Sprachstands von fünfjährigen Kindern geht. Dort stehen wie bei Bumerang förderdiagnostische Komponenten in Verbindung mit Mehrsprachigkeit im Vordergrund.<sup>41</sup> Der große Unterschied zu HAVAS 5 ist, dass beim *Bumerang-Verfahren* die berufsspezifischen fach- und bildungssprachlichen Kompetenzen von Jugendlichen erfasst werden, um die Schnittstelle Schule – Beruf in den Vordergrund zu rücken. Bemerkenswert daran ist, dass dieses Förderinstrument zwar grundsätzlich für Mehrsprachige angelegt ist - es gibt Auswertungsinstrumente in Deutsch, Türkisch und Russisch -, sich aber laut FÖRMIG durchaus auch für monolingual aufgewachsene Deutschsprachige eignet.<sup>42</sup> Dieser positive Aspekt ist der regen Forschungstätigkeit der FÖRMIG-Gruppe zu verdanken, die in ihren Studien auch Menschen mit deutscher Muttersprache mitberücksichtigte.

---

<sup>39</sup> FÖRMIG – Kompetenzzentrum für Förderung von Kindern und Jugendlichen mit Migrationshintergrund – Hamburger Forschungsprojekt unter Leitung von Ingrid Gogolin

<sup>40</sup> vgl. dazu Lengyel, Drorit, Reich, Hans H., Roth, Hans-Joachim, Döll, Marion [Hrsg.] Von der Sprachdiagnose zur Sprachförderung. Waxmann: Münster/New York 2009. S. 139.

<sup>41</sup> vgl. dazu Lengyel u. a. (2009), S. 73 – 74.

<sup>42</sup> vgl. dazu Lengyel u. a. (2009), S. 139.

Grundsätzlich handelt es sich bei *Bumerang* nicht um ein herkömmliches Testverfahren, sondern um eine profilanalytische Erhebung und Auswertung von berufsbezogenen sprachlichen Bereichen, die der weiteren schulischen bzw. beruflichen Sprachförderung dienen soll. Die Auswertungswerkzeuge bei *Bumerang* beziehen sich demnach nicht nur auf die Bildungssprache, sondern auch auf die Fachsprache.<sup>43</sup>

Kennzeichen von Fachsprache		
<b>Morphologie</b> Substantivierungen Adjektive mit bestimmten Suffixen Partizipien Konjunktive	<b>Syntax</b> Funktionsverbgefüge Nominalisierungsgruppen Formen der Unpersönlichkeit diverse Verweisformen Partizipialkonstruktionen	<b>Lexik</b> Fremdwörter (mehrgliedrige) Komposita fachspezifischer Wortschatz fachspezifische Abkürzungen hoher Grad an Exaktheit

vgl. Benholz & Iordanidou 2004, sowie Krischer 2005

(Abbildung 5: Auszug aus der FÖRMIG - Homepage: Internet-Zugriff 04. 12. 2011)

Bemerkenswerterweise gelten für das Bumerang-Verfahren nicht nur der *Referenzrahmen für altersspezifische Sprachaneignung*, sondern auch die *Anforderungen für Verfahren der regelmäßigen Sprachstandsfeststellung für Kinder mit und ohne Migrationshintergrund* von *Konrad Ehlich*. Dieser Anforderungskatalog umfasst sechs Kategorien:

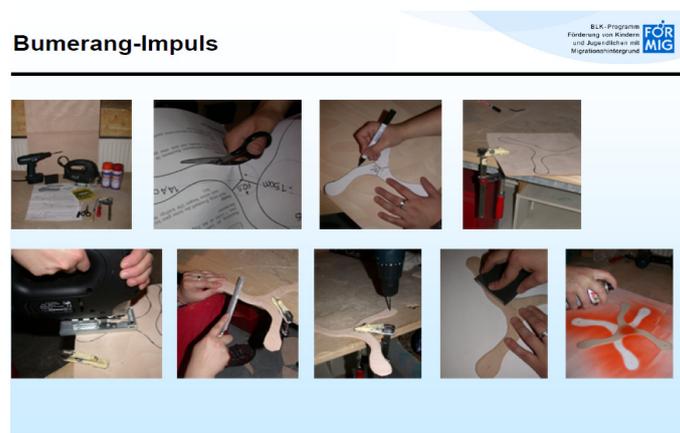
- Allgemeine linguistische Anforderungen wie z. B. Differenzierung nach Qualifikationsbereichen
- Mehrsprachigkeit wie z. B. Einbeziehung beider Sprachen
- Diagnostische Anforderungen wie z. B. die Reichweite, der Stellenwert und die Verankerung im Erhebungsspektrum
- Alterstypische und individuelle Dokumentationen wie z. B. alterstypische Aneignungsbereiche
- Institutionelle Anforderungen wie z. B. eine realistische Förderkultur
- Sprachliches Handeln, diskursive und textuelle Fähigkeiten wie z. B. Sprachdaten aus realen kommunikativen Konstellationen<sup>44</sup>

<sup>43</sup> vgl. dazu Lengyel u. a. (2009), S. 139 – 140.

<sup>44</sup> vgl. dazu Ehlich, Konrad: *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Sprachförderung von Kindern mit und ohne Migrationshintergrund*. München: 2004. S. 46 – 50.

Konrad Ehlich führt anhand seines Anforderungskatalogs vor, dass gegenwärtige Qualitätsstandards für derartige sprachdiagnostische Verfahren nicht ausreichen und dass es zusätzlicher Testkriterien bedarf. Zudem ist Bumerang ein Verfahren, das - im Gegensatz zu vielen herkömmlichen Sprachüberprüfungen - auf reale berufsspezifische Situationen direkt Bezug nimmt.

Mittels Schreibimpuls werden zwei unterschiedliche Textsorten initiiert, ein Bewerbungsschreiben um einen Praktikumsplatz und eine Bauanleitung (Bildbeschreibung) zur Herstellung eines Bumerangs. Bei der Bauanleitung werden die Testpersonen aufgefordert, anhand einer neunteiligen Bildfolge die Herstellung eines Bumerangs schriftlich so zu formulieren, dass der Vorgang auch ohne Bilder nachvollzogen werden kann.



(Abbildung 6: Auszug aus der FÖRMIG – Homepage: Interne-Zugriff 04. 12. 2011)

Während beim Bewerbungsschreiben allgemeine textspezifische Fähigkeiten im Vordergrund stehen, sind bei der Bauanleitung in erster Linie fachkommunikative Fähigkeiten (kausale Bezüge herstellen) in Verbindung mit einem entsprechenden Fachwortschatz gefragt. Der umfangreich gestaltete Auswertungsbogen richtet sich nach den textspezifischen Gesichtspunkten der beiden Aufgaben:

<b>Auswertungskriterien Deutsch</b> (Stand 05/2006)	
<p><b><u>Bewerbung</u></b></p> <ul style="list-style-type: none"> <li>• formale Gestaltung</li> <li>• inhaltliche Gestaltung</li> <li>• Wortschatz (Verben, Nomen, Adjektive)</li> <li>• Satzverbindungen</li> <li>• Lexikalische Dichte (allgemein und fachsprachlich)</li> <li>• Verbspezifität</li> <li>• MLS (mittlere Satzlänge)</li> </ul>	<p><b><u>Bumerang-Artikel</u></b></p> <ul style="list-style-type: none"> <li>• Aufgabenbewältigung</li> <li>• (Fach-)Wortschatz (Verben, Nomen, Adjektive)</li> <li>• Schriftsprachliche Elemente (Nominalisierungen, Attribute, Komposita, Partizipien etc.)</li> <li>• Strukturierung des Textes</li> <li>• Herstellung von Textzusammenhang</li> <li>• Satzverbindungen</li> <li>• Lexikalische Dichte (allgemein und fachsprachlich)</li> <li>• Verbspezifität</li> <li>• MLS (mittlere Satzlänge)</li> </ul>

(Abbildung 7: Auszug aus der FÖRMIG-Homepage: Internet-Zugriff 04. 12. 2011)

Im Gegensatz zu derzeit verwendeten Sprachtests geht es bei Bumerang jedoch nicht um Bestehen oder Nichtbestehen, sondern vielmehr um das Herausfinden von Stärken und Schwächen der Testteilnehmer innerhalb der Bildungs- und Fachsprache hinsichtlich möglicher Fördermaßnahmen.

Bezogen auf den schriftlichen Teil der Zulassungsprüfung Deutsch steht außer Zweifel, dass das Bumerang-Verfahren mehr Vorteile als Nachteile gegenüber dem Diktat hat. Die Vorteile liegen meiner Meinung nach vor allem im Testkonstrukt, wo sich reale berufsspezifische Sprachverwendungssituationen in den beiden Aufgabenstellungen unmittelbar wiederfinden. Die Nachteile von diagnostisch ausgerichteten Verfahren wie bei Bumerang liegen erfahrungsgemäß in der Praktikabilität, insbesondere im Bereich der Auswertung, wo nicht nur der Zeitaufwand, sondern auch die Schulung des Auswertungspersonals eine wesentliche Rolle spielt.

Es darf auch nicht übersehen werden, dass es sich hier um kein Prüfformat, schon gar nicht um eine Zulassungsprüfung handelt, sondern um ein diagnostisches Instrument zur Bestimmung von schriftsprachlichen Fähigkeiten. Geht man allerdings davon aus, dass wie bei der Zulassungsprüfung Deutsch die Förderung im Vordergrund stehen soll, könnte das Bumerang-Verfahren durchaus als berufsspezifische Zulassungsprüfung adaptiert werden (Erstellen einer Meldung, Beschreiben der Spindordnung gemäß Bildimpuls usw.).

Außerdem wurde das Bumerang-Diagnoseverfahren, laut Information der Testersteller, gleichermaßen bei MuttersprachlerInnen und Menschen mit Migrationshintergrund geprüft.

## **4.2 Prozessbegleitende Diagnoseinstrumente**

Im gleichen Maße aufschlussreich wie Bumerang sind die, ebenfalls von der FÖRMIG-Gruppe entwickelten, bisher noch nicht veröffentlichten

prozessbegleitenden Diagnoseinstrumente zur Einschätzung der Schreibfähigkeiten im Deutschen von SchülerInnen der Sekundärstufe I, die in erster Linie zur fächerübergreifenden kollegialen Kommunikation von Deutschlehrkräften und Fachlehrkräften für natur- und sozialwissenschaftliche Fächer geschaffen wurden. Die Einstufung erfolgt anhand von Auswertungsrastern in lexikalischer, syntaktischer und textueller Hinsicht. Gegenwärtig liegen Auswertungsraster für die Sprachhandlungen Berichten, Beschreiben, Erklären und Argumentieren sowie allgemeine Hinweise zur Auswertung vor.<sup>45</sup>

Geradezu auffällig ist bei den Auswertungsrastern, dass sie sich, anders als bei Bumerang, auf drei bzw. vier Hauptkategorien (beim Raster Erklären kommt die Kognition dazu!)

---

<sup>45</sup> vgl. dazu Lengyel u. a. (2009). S. 129 – 130.

beschränken. Ein Grund dafür ist, dass es sich hier um profilartige Beobachtungen von schriftlichen Sprachhandlungen innerhalb der Sekundärstufe I handelt und nicht um die Ermittlung von fachkommunikativen Kompetenzen. Das zeigen auch die sechs gewählten Kompetenzstufen, die meiner Ansicht nach zu linear dargestellt, bzw. deren Übergänge nicht berücksichtigt werden. Dieser Umstand ist dem Forscherteam im Zuge von Evaluierungen durchaus bewusst geworden. Eine Modifizierung der Raster ist derzeit noch geplant.

**Auswertungsraster für die Sprachhandlung Beschreiben**  
Vorläufige Version



Ebene		Stufe 1	Stufe 2	Stufe 3	Stufe 4	Stufe 5	Stufe 6
Lexik, Semantik	Präzision der sprachlichen Abbildung	verbalen Zeigen, Joker	Näherungsbegriffe, umgangssprachlich gehaltener Wortschatz	einfacher, geläufiger Wortschatz, treffende Umschreibungen	differenzierter Wortschatz	sachangemessene Wortwahl unter Einbeziehung des unterrichtlichen und gegebenenfalls fachsprachlichen Wortschatzes	durchgehend präziser und fachsprachlicher Wortschatz
Syntax	Klarheit der Gliederung des Vorstellungsraums	Benennung willkürlicher herausgegriffener Elemente des Beschreibungsgegenstandes	einfache Sätze mit Unsicherheiten in Wortstellung und Wortform (Flexionsformen)	reihende Aussagen, erweiterte Sätze mit einfachen Attributen oder Relativsätzen	Gliederung durch adverbiale Bestimmungen, Genauigkeit durch koordinierte Attribute, Vergleiche und erweiterte Relativsätze	Verdichtung durch komplexe Attribute, komplexe adverbiale Bestimmungen oder Erweiterung durch spezifizierende Nebensätze	Erweiterte Partizipien, Gebrauch von unpersönlichen und/oder konjunktivischen Formulierungen bei Uneindeutigen
Text	Vollständigkeit und Ordnung („Ganz durch den Vorstellungsraum“)	Benennung einzelner, herausgegriffener Elemente des Beschreibungsgegenstandes	Linear angeordnete Aufzählung der Elemente des Beschreibungsgegenstandes	Punktuell vertiefende Darstellung von Teilmassammenhängen	Im Wesentlichen vollständige Darstellung des Beschreibungsgegenstandes, Berücksichtigung über- und unterordnender Gesichtspunkte	Systematisch gegliederte ausführliche Darstellung, z.B. ausgehend von einem Gesamtkontext oder einem hervorgehobenen Element des Beschreibungsgegenstandes und/oder vom Standpunkt des Beobachters	Explizite Kennzeichnung des Beobachterstandortes  Relativierung der Äußerung bei Uneindeutigkeit, Markierung von Vermutungen

© Programmträger FörMig, FÖRMIG AG SEK I, 2009- Universität Hamburg Fakultät EPB, Von-Melle-Park 8, 20146 Hamburg

(Abbildung 8: Auszug aus der FÖRMIG–Homepage: Internet-Zugriff 04. 12. 2011)

Sieht man von der Tatsache ab, dass es sich hier weder um einen Test noch um ein diagnostisches Instrument handelt, eignet sich dieser Kompetenzraster, bei entsprechender Punkteuteilung, durchaus als Bewertungsbogen für berufsspezifische Sprachprüfungen wie der Zulassungsprüfung Deutsch. Diese Umformatierung des Kompetenzrasters würde allerdings voraussetzen, dass es sich um die gleichen Sprachhandlungen handelt bzw. um die gleiche Personengruppe geht. Hinsichtlich der Kompetenzraster „Beschreiben“ und „Erklären“ gäbe es innerhalb des Militärs sogar inhaltlich und sachlich adäquatere Sprachhandlungen (Beobachten und Melden, Erklären von Waffen und Gerät). Sehr nützlich erscheinen mir vor allem die einzelnen Sprachentwicklungsstufen, aus denen sich präzise Förderschnitte ableiten ließen. Auf die Frage, ob die Diagnoseraster auch bei SchülerInnen mit Erstsprache Deutsch überprüft wurden, antwortete Drorit Lengyel:

„Gerade in den oberen Stufen, sind die „Fehler“ von DaZ-LernerInnen und Muttersprachlern häufig sehr ähnlich. In den unteren Stufen gibt es natürlich viel mehr DaZ-Spezifisches. Aus diesem Grund sind die unteren Stufen auch ausdifferenziert, was die Fehler anbelangt (z. B. Unsicherheiten beim Flektieren). Da wir nicht von der Linguistik, sondern von den Anforderungen der Schule ausgehen, in der ja Mutter- und Zweitsprachler gemeinsam unterrichtet werden, war es für uns wichtig, ein Instrument zu entwickeln, das für beide Gruppen durchaus angewendet werden kann. Nach meinem Wissen nach gibt es kein Verfahren, das wirklich für beide Gruppen in allen Facetten geeignet ist.“

Gerade bei Testverfahren, die den psychometrischen Gütekriterien und testtheoretischen Ansprüchen entsprechen, steht dann häufig im Manual: Die Ergebnisse müssen bei Kindern mit Migrationshintergrund sensibel interpretiert werden, da sie in der Stichprobe nur unzureichend berücksichtigt wurden oder ähnliches. Da funktionieren die Profilanalysen oder Instrumente wie das von mir vorgestellte schon besser, weil sie in die Tiefe gehen und „sprachliche Meilensteine“ in der Entwicklung als Ausgangspunkt nehmen.<sup>46</sup>

Insgesamt gesehen liefern sowohl das Bumerang-Diagnoseverfahren als auch die prozessbegleitende Diagnose zur Schreibleistung nicht nur mögliche Vorlagen zur Erstellung eines neuen Testformats, sondern auch eine gute Basis zur Einschätzung schriftlicher Leistungen.

### **4.3 Kommunikatives Testen und Qualitätssicherung beim ÖSD**

Eine weitere gute wissenschaftliche Grundlage, vor allem aus testtheoretischer Sicht, liefert der *Band 2 der Reihe „Theorie und Praxis – Österreichische Beiträge zu Deutsch als Fremdsprache“* von Manuela Glaboniat. Wie schon bei den Messinstrumenten der FÖRMIG-Gruppe setzt sich auch diese schon etwas zurückliegende Untersuchung (1998) von Manuela Glaboniat nicht explizit mit Deutsch als Muttersprache, sondern mit der Frage des Messens und Überprüfens fremdsprachlicher Handlungs- und Kommunikationsfähigkeit auseinander. In erster Linie geht es dabei um die *Umsetzbarkeit und Handhabbarkeit von testtheoretischen Prinzipien am Beispiel des Österreichischen Sprachdiploms Deutsch*<sup>47</sup>, also vor allem um die immer noch aktuellen Bemühungen um Qualitätssicherung von Sprachtests. Dass die Umsetzbarkeit und Handhabbarkeit von testtheoretischen Prinzipien bei DaZ und DaM eine noch größere Rolle spielt, beweist alleine der Umstand, dass es sich in spracherwerbstheoretischer Hinsicht um zwei, zum Teil völlig unterschiedliche Lernergruppen handelt. Während Menschen mit deutscher Muttersprache über genügend Weltwissen, Textsortenwissen und schreibstrategisches Wissen verfügen, um ihr Sprachkönnen, z. B. in der sprachlichen Korrektheit (Grammatik, Orthographie), zu kompensieren, sind Menschen mit nichtdeutscher Muttersprache zumeist ausschließlich auf ihr Sprachkönnen angewiesen.<sup>48</sup>

Im Zuge dieser, am Beispiel des Österreichischen Sprachdiploms Deutsch (ÖSD) unternommenen Untersuchung wurde insbesondere das Verhältnis von Validität, Reliabilität und Objektivität bei diversen Testformaten, wie Multiple-Choice-Verfahren oder C-Tests

---

<sup>46</sup> Ausschnitt aus dem E-Mail von Drorit Lengyel - 16. 03. 2011

<sup>47</sup> vgl. dazu Glaboniat (1998), S. 7 - 8.

<sup>48</sup> vgl. dazu z. B. Krumm (2005), S. 97 – 101.

einer Überprüfung unterzogen. Die Schlussfolgerungen von Glaboniat waren, dass zwar zahlreiche Testformate den Ansprüchen der Reliabilität und Objektivität weitestgehend entsprachen, aber deutliche Mängel im Bereich der Validität hatten. Demgegenüber erwiesen sich offene, sach- und kontextabhängige Testformate wie beim ÖSD, die in erster Linie auf kommunikative Sprachfähigkeit abzielen, als aussagekräftiger und pädagogisch nachhaltiger.<sup>49</sup>

Für die Bewertung bedeutet die Tendenz zu offenen Formaten und Handlungsorientiertheit, dass die Kriterien zur Bewertung der sprachlichen Richtigkeit ergänzt werden durch Kriterien, die es ermöglichen, die kommunikativen Stärken und Schwächen des Testteilnehmers zu beurteilen. Dabei stellen sich folgende Fragen:

- Wie viele und welche Kriterien sollte man festlegen?
- Wie erfolgt die Gewichtung der Kriterien untereinander? Sollen kommunikative Kriterien wie Aufgabenbewältigung, Textkohärenz und Lexik stärker gewichtet werden als die sprachliche Richtigkeit?<sup>50</sup>

Statt die orthographische Korrektheit in den Mittelpunkt zu stellen wie beim Diktat innerhalb der Zulassungsprüfung Deutsch oder mit Richtig/Falsch-Entscheidungen schriftsprachliche Leistungen zu messen, wurden beim ÖSD Kriterien wie Aufgabenbewältigung, kommunikative Angemessenheit, Lexik und auch formale Richtigkeit entsprechend gewichtet und im Sinne einer sprachlichen Gesamtleistung zusammengefasst. Was von früheren formalorientierten Testverfahren blieb, ist die Tatsache, dass das Nichterfüllen eines einzelnen Kriteriums wie zum Beispiel bei der „Formalen Richtigkeit“ automatisch ein Nichtbestehen der Prüfung zur Folge hat. Offensichtlich scheint trotz handlungsorientiertem Bezug kein Testformat ohne Formalkriterien auszukommen.

Wie aus der Untersuchung am Beispiel des ÖSD hervorgeht, verlangt die Entwicklung von handlungsorientierten kommunikativen Testverfahren neue Schwerpunkte in der Testkonstruktion und in der Testbewertung. Ob dabei ein Nützlichkeitskriterienkatalog wie von Bachmann und Palmer bzw. allgemeine Sprachtestanforderungen ausreichen, ist fraglich, vor allem dann, wenn der Begriff der kommunikativen Kompetenz ständig ausgeweitet wird.

---

<sup>49</sup> vgl. dazu Glaboniat (1998), S. 7 - 13.

<sup>50</sup> vgl. dazu Bolton, Sybille (2005): Probleme der Bewertung bei den Produktiven Fertigkeiten in den Deutschprüfungen des Goethe-Instituts. In: Kühn, Ingrid u.a. [Hrsg.]: Sprachtests in der Diskussion. Wittenberger Beiträge zur deutschen Sprache und Kultur - Band 4. Frankfurt u. a.: Lang S. 57 – 58.

Es stellt sich an dieser Stelle die Frage, ob die Grenzen des Überprüfbareren nicht ohnehin schon überschritten worden sind, wenn man bedenkt, dass Schreiben zumeist in Abhängigkeit von Welt- Kulturwissen, Lerner- und Überlebensstrategien sowie von spezifischem Textsortenwissen steht.

In jedem Fall wird in Zukunft eine weitere Eingrenzung des Testgegenstands erforderlich sein, um die komplexen Spracherwerbsprozesse, die hinter jedem Textprodukt stehen, zumindest annähernd einbeziehen zu können.

#### **4.4 Hamburger Schreibprobe**

Dass Orthographie nicht nur im Diktat, sondern auch in diagnostischen Verfahren zum Testgegenstand erhoben wird, will ich am Beispiel der Hamburger Schreibprobe (2002) zeigen.

Im Gegensatz zum Diktat handelt es sich bei diesem Testverfahren um eine qualitative Rechtschreibanalyse, die neben orthographischem Strukturwissen vor allem spezifische Rechtschreibstrategien erfassen soll. Die Hamburger Schreibprobe fußt auf einem Forschungsprojekt von Peter May, wo es um die Definition von orthographischer Kompetenz geht.

Orthographische Kompetenz, so May, umfasst zum einen die Einprägung wortspezifischer Schreibungen und zum anderen eine übergreifende Gesamtstrategie, die sich in fünf Einzelstrategien aufgliedert:

- *Logographemische Strategie (Merken der Form und Anordnung von Zeichen)*
- *Alphabetische Strategie (lautsprachliche Informationen der Schrift nützen)*
- *Orthographische Strategie (Modifikation von Laut-Buchstaben-Beziehungen)*
- *Morphematische Strategie (Ableitung der Schreibung nach verwandten Wortstämmen)*
- *Wortübergreifende Strategie (Satz- und Textebene für Orthographie nützen)<sup>51</sup>*

Aus Mays Studie geht hervor, dass orthographisch richtiges Schreiben in erster Linie als kognitives und strategisches Verfahren begriffen werden soll. Es geht daher nicht um Sprachkönnen, sondern um Strukturwissen in Verbindung mit unterschiedlichen Strategien.

*May bezeichnet darin Strategien als verinnerlichte Regeln, die systematisch wiederkehrend das Handeln steuern.*<sup>52</sup>

---

<sup>51</sup> Ausschnitt aus der Homepage von Peter May: Internet-Zugriff 04. 12. 2011

<sup>52</sup> vgl. dazu Homepage von Peter May: Internet-Zugriff 04. 12. 2011

Bei der Hamburger Schreibprobe werden daher, neben einer quantitativen Auswertung von richtig geschriebenen Wörtern und Graphemen, vor allem Rechtschreibstrategien qualitativ ausgewertet, um Binnendifferenzierungen und entsprechende Fördermaßnahmen einzuleiten.

Die Schreibaufgaben sollen jenes Strukturwissen erfassen, das den Kern orthographischen Wissens und Könnens ausmacht. Die Wörter und Sätze enthalten in genügender Anzahl solche Phänomene, welche die wichtigsten alphabetischen, orthographischen und morphematischen Zugriffe abfordern.

Die Auswertung innerhalb der Hamburger Schreibprobe erfolgt anhand von Lupenstellungen, deren Beachtung den Grad der Herausbildung verschiedener Rechtschreibstrategien oder grundlegender Zugriffsweisen auf Schrift deutlich macht.<sup>53</sup>



Abbildung 4: Christians Schreibungen in der HSP 5-9

(Abbildung 9: Homepage von Peter May: Internet-Zugriff 04. 12. 2011)

Viele Orthographiefehler, die häufig gerne auf mangelndes Sprachkönnen zurückgeführt werden, entpuppen sich, wie die Ergebnisse der Hamburger Schreibprobe zeigen, als strategisches Nichtwissen. Laut May lassen sich jene schriftsprachlichen Mängel nachweisen, die auf Störungen im Aneignungsprozess von Strategiewissen zurückzuführen sind. Insgesamt gesehen ist die Hamburger Schreibprobe (HSP) ein Rechtschreibtest, der hilft, das Rechtschreibkönnen und die grundlegenden Rechtschreibstrategien der Schüler einzuschätzen. Die Hamburger Schreibprobe soll daher als ein Gesamtkonzept zur Feststellung grundlegender Fähigkeiten im Rechtschreiben in der Schule verstanden werden, mit deren Hilfe die Planung von gezielten Fördermaßnahmen erfolgen kann.<sup>54</sup>

Man könnte daher annehmen, dass Rechtschreibfehler in unserem Sinne nur selten vorkommen, weil Fehler, die gegen die sprachliche Norm verstoßen, vielfach auf außersprachliche Phänomene zurückzuführen sind. Das könnte im geringeren Ausmaß auch

<sup>53</sup> vgl. dazu Homepage von Peter May: Internet-Zugriff 04. 12. 2011

<sup>54</sup> vgl. dazu Homepage von Peter May: Internet-Zugriff 04. 12. 2011

auf Grammatikfehler zutreffen. So gesehen wären Prüfverfahren, die sich ausschließlich auf das Zählen von Fehlern beschränken, nicht nur ungerecht, sondern auch nutzlos. Ob diese Hypothese auch auf das Diktat zutrifft, soll meine Untersuchung zeigen.

Demnach zeigt sich die Leistung Peter Mays bei der Entwicklung der Hamburger Schreibprobe meiner Meinung nicht nur im Prüfverfahren, sondern vielmehr in der Beschäftigung mit der Entstehung orthographischer Fehler.

Peter May zeigt uns wie komplex schriftsprachliche Prozesse ablaufen und welche außersprachlichen Störfaktoren den Spracherwerbsprozess entscheidend beeinflussen können.

## 5 Anforderungen an Sprachtests

### 5.1 Testtheoretische und testmethodische Welten

Welche testmethodischen Welten bzw. Anforderungen stehen bei schriftlichen Sprachprüfungen wie der Zulassungsprüfung Deutsch im Vordergrund? Welche Auswirkungen haben die von Palmer vorgeschlagenen Nützlichkeitskriterien auf die Testerstellung bzw. auf die Testauswertung? Warum stehen gewisse Kriterien im Vordergrund und warum werden andere vernachlässigt? Was folgt daraus für das Diktat?

*„Unter Test oder Prüfung wird jegliches Verfahren verstanden, das Individuen unter kontrollierten Bedingungen zu bestimmten Handlungs- und Verhaltensweisen veranlasst, die Rückschlüsse auf bestimmte Fähigkeiten und Fertigkeiten erlauben sollen.“<sup>55</sup>*

Wenn man von Anforderungen an schriftliche Sprachtests spricht, sollte man zunächst der Frage nachgehen, was eigentlich Testen bedeutet.

Beim Testen, insbesondere beim Sprachtesten, geht es zum einen darum, sowohl quantitative als auch qualitative Informationen zur Beurteilung einer Person zu sammeln, um den Ausprägungsgrad einer bestimmten Kompetenz zu messen und entsprechende Entscheidungen daraus abzuleiten. Voraussetzung dafür ist, dass die Erwartungen an die zu messenden Fähigkeiten zumindest annähernd den Sprachniveaus der TestteilnehmerInnen entsprechen.<sup>56</sup>

Eine Person zu testen bedeutet aber nicht nur Messen, sondern zum anderen, die Daten anhand eines festgelegten Bezugsrahmens (Norm bzw. Kriterium) einzuschätzen und zu

---

<sup>55</sup> Kecker, Gabriele: Validität und Validierung von Sprachprüfungen. In: Berndt Annette / Kleppin Karin (Hrsg.): Sprachlehrforschung: Theorie und Empirie Festschrift für Ruediger Grotjahn. Frankfurt: Lang 2009. S. 129.

<sup>56</sup> vgl. dazu z. B. Glaboniat, Manuela (2006): Sprachkenntnisse beschreibbar, vergleichbar und messbar gemacht!? Über den Nutzen und die Grenzen der Europaratsskalen im Bereich des Prüfens und Zertifizierens. In: Krumm Hans-Jürgen, Portmann-Tselikas Paus [Hrsg.]: Begegnungssprache Deutsch - Motivation, Herausforderung, Perspektiven, Innsbruck-Wien. Bozen: Studienverlag S. 84 – 85.

bewerten. Hinsichtlich der Beurteilung von Sprachfähigkeiten ist es daher notwendig, die Frage zu stellen: Worauf wird Bezug genommen?

Wird Bezug genommen auf die Leistungen anderer Testteilnehmer (Normorientierung) bzw. auf Menschen mit gleichem Bildungshintergrund oder bezieht man sich auf Leistungen einer Einzelperson in Bezug zu einem Fähigkeitsgrad (Kriterium)?<sup>57</sup>

Normorientierte Tests werden so konstruiert, dass die Ergebnisse einer Normalverteilung folgen. Wenige Probanden erreichen eine hohe Leistung, viele eine mittlere und wieder wenige eine schwache Leistung. Die Normalverteilung erlaubt die Berechnung eines Durchschnittswertes.

Kriteriumsorientierte Tests prüfen allein, ob Lernziele erreicht oder nicht erreicht werden. Es geht nur darum das festzustellen: Erreicht der Proband im Vergleich auf einen Bezugspunkt die Leistung? Der Bezugspunkt ist das Kriterium. Die Normalverteilung greift hier nicht mehr.<sup>58</sup>

In der Praxis überwiegen meist Mischformen. Zum Kriterium, in unserem Fall der sprachlichen Korrektheit, wird zusätzlich die Testteilnehmergruppe als Bezugsnorm herangezogen, um einen entsprechenden Schwellenwert zu ermitteln. Hier herrscht insofern oft Verwirrung, als normorientierte Tests auf eine Durchschnittsnorm bzw. Realnorm bezogen sind und kriteriumsorientierte Tests sich an einer Idealnorm, die zumeist dem System einer Sprache entnommen wird, orientieren und ein Lernziel bzw. Anforderungsziel (Curriculare Norm) definiert wird. Der Grad der Beherrschung und nicht der Bezug auf eine andere Person ist ausschlaggebend.<sup>59</sup>

Für die Zulassungsprüfung Deutsch beim ÖBH gilt der Pflichtschulabschluss als Normbezug. Der Bezugspunkt innerhalb der Sprachnorm, das Bestehenskriterium, soll laut Experten, bezogen auf das Sprachniveau etwa zwei Schulstufen darunter liegen.

Demnach könnte das Prüfungsformat Diktat ausschließlich als kriteriumsorientiertes Testformat verstanden werden, das sich ausschließlich auf ein Anforderungsziel bezieht. Generell hat die Zulassungsprüfung Deutsch aber nicht nur den Zweck, ein Kriterium abzufragen, sondern hat auch eine Steuerungsfunktion zur Personalaufbringung beim Bundesheer. Das bedeutet, dass beispielsweise bei einem generellen Abfall des Schulniveaus in Verbindung mit einem zu niedrigen Personalstand (zu wenige schaffen die Prüfung!) die

---

<sup>57</sup> Glaboniat (2006), S. 83.

<sup>58</sup> Schelten, Andreas: Einführung in die Berufspädagogik. E., überarbeitete und aktualisierte Auflage, Stuttgart: Steiner Verlag 2010. S. 254

<sup>59</sup> vgl. dazu Glaboniat (2009), S. 44 – 47.

Messlatte (das Niveau) herabgesetzt und an der Normalverteilung ausgerichtet wird. In diesem Fall könnte man das Diktat durchaus auch als normorientierten Test bezeichnen.

Noch vor der Festlegung des Bezugsrahmens (Norm, Kriterium) erfolgt auf Basis des Forschungsstandes zumeist die Entscheidung, welche Funktion und welche Form ein Sprachtest haben soll. Dabei gilt es zu berücksichtigen, unter welchen Voraussetzungen und für welchen Zweck Tests erstellt werden und welche Auswirkungen bestimmte Verwendungszusammenhänge auf das Testergebnis haben.

Bei der Funktion eines Sprachtests geht es in erster Linie darum, welche Entscheidungen aufgrund der Ergebnisse getroffen werden sollen. Diese Entscheidungen fallen meist auf der Basis bildungspolitischer Erwägungen bzw. richten sich auch nach den Interessen von Testinstitutionen und Testerstellern.

Insgesamt verlangt das Thema *Funktion von Sprachtests* nach Chapelle und Read die Beachtung von drei verschiedenen Faktoren:

Die beabsichtigten Interpretationen der Ergebnisse, der beabsichtigte Umgang damit und die beabsichtigten Auswirkungen der Tests. Eine Typologie von Sprachtest nach ihrer Funktion existiert nicht, da in unterschiedlichen Einsatzbereichen eine Vielzahl von Funktionen vorkommt, welche sich nicht genau abgrenzen lassen.<sup>60</sup>

Erfahrungsgemäß herrscht demnach nur selten Klarheit über die Funktion, was darauf zurückzuführen ist, dass sich unterschiedliche Testverfahren meist für unterschiedliche Testfunktionen eignen. Außerdem kann man davon ausgehen, dass sowohl bei Sprachtests als auch bei Sprachprüfungen eine starke Abhängigkeit zwischen Funktion, Form und Inhalt besteht.

Der folgende Abschnitt, wo es um die Beschreibung der Testfunktionen, geht bezieht sich im Wesentlichen auf die Ausführungen von Dlaska und Krekeler<sup>61</sup> bzw. Perlmann-Palme.<sup>62</sup>

Je nach Funktion unterscheidet man grundsätzlich sechs Grundtypen:

---

<sup>60</sup> vgl. dazu Chapelle, C. A. 2001: Computer applications in second language acquisition. Foundations for teaching, testing and research (Cambridge Applied Linguistics). Cambridge: Cambridge University Press. S. 101 - 102.

<sup>61</sup> vgl. dazu Dlaska, Andrea/Krekeler, Christian: Sprachtest. Leistungsbeurteilungen im Fremdsprachenunterricht evaluieren und verbessern. Baltmannsweiler: Schneider Verlag Hohengehren 2009. S. 27 – 31.

<sup>59</sup> Perlmann-Balme, Michaela: Formen und Funktionen von Leistungsmessung und Kontrolle. In: Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. Herausgegeben von Hans-Jürgen Krumm, Christian Fandrych, Britta Hufeisen und Claudia Riemer. 2 Bände. De Gruyter Verlag: Berlin 2010. S. 994 – 1006.

<sup>60</sup> vgl. dazu Dlaska (2009), S. 27 – 28.

## 1. Lernfortschritts- und Kursabschlusstests

Lernfortschritts- und Kursabschlusstests beziehen sich in erster Linie auf den Lernweg und können generell als Leistungstests verstanden werden. Sie werden deshalb vor allem im Rahmen von Kursen und meist in Bezug zu einem Curriculum als Steuerelement eingesetzt. Geprüft wird ausschließlich das, was unterrichtet wurde, sei es am Ende von Lernlektionen oder am Ende von Kursen. Die Testformate werden zumeist von internen Lehrpersonen erstellt und unterliegen ausschließlich institutsinternen Regelungen. Das Besondere dabei ist, dass Lernfortschritt- und Kursabschlusstests eine wichtige Kontrollfunktion sowohl für den Lerner als auch für den Lehrer darstellen. Die Aufgabenstellungen, insbesondere bei Lernfortschrittstests, haben meist einen diagnostischen Charakter und bestehen aus wenigen, wenn nicht gar aus einer einzigen Aufgabenstellung; die Ergebnisse haben keine externen Auswirkungen.<sup>63</sup>

Anders als bei Lernfortschrittstests werden die Ergebnisse bei Kursabschlussprüfungen durchaus für institutsexterne Funktionen verwendet, wie z. B. für Zulassungsprüfungen beim ÖBH. Auffällig ist, dass dabei immer häufiger standardisierte Testverfahren zur Anwendung kommen.

## 2. Feststellungsprüfungen (Qualifikationsprüfungen)

Im Gegensatz zu Lernfortschritts- und Kursabschlusstests geht es bei Feststellungsprüfungen unabhängig von Kursen und Curricula um eine umfangreiche Messung der sprachlichen Fähigkeiten, meist in Form eines Zertifikats wie z. B. beim ÖSD. Bei Feststellungsprüfungen handelt es sich zumeist um standardisierte, international anerkannte Testverfahren, die auch für Außenstehende interpretierbar sein sollen. Durch den meist hohen Bekanntheitsgrad des Prüfungsformats ist eine gezielte Vorbereitung möglich. Die Prüfung selbst besteht in der Regel aus zwei Teilen, einem mit unterschiedlichen Aufgaben aus den Fertigkeiten Hörverstehen, Leseverstehen und schriftlicher Gebrauch und einem zweiten, vorwiegend mündlichen Teil in Form einer kommissionellen Einzelprüfung. Bei Feststellungsprüfungen findet man meist eine Mischung zwischen offenen und geschlossenen Aufgabentypen (z. B. Ja/Nein-Aufgaben beim Leseverstehen bzw. freies Schreiben beim schriftlichen Gebrauch).<sup>64</sup>

---

<sup>64</sup> vgl. dazu Dlaska (2009), S. 28 – 29.

### 3. Diagnosetests

Diagnosetests können im weiteren Sinne auch als Lernfortschrittstests begriffen werden. Allerdings gehen die Ergebnisse eines Diagnosetests darüber hinaus, Aussagen über den Lernfortschritt zu liefern. Anhand von Sprachstandsdiagnosen lassen sich vor allem Rückschlüsse auf den Lernprozess ziehen und daraus Empfehlungen für die nächsten Lern- bzw. Förderschritte bzw. Empfehlungen ableiten. Wie bei den Diagnoseinstrumenten der FÖRMIG-Gruppe geht es zumeist um Sprachprofil-Analysen, die möglichst detaillierte Aussagen über den Sprachstand treffen sollen. Es dominieren daher meist offene bzw. freie Aufgabenstellungen mit dem Zweck, möglichst produktive und kreative Sprachleistungen zu erhalten. Entscheidend ist, dass hier nicht nur die Defizite erkannt werden, sondern deren Zustandekommen. Wenig eignen sich hier geschlossene Aufgabentypen wie Multiple-Choice bzw. halboffene Aufgaben wie beim Cloze-Test.<sup>65</sup>

### 4. Einstufungstests

Bei Einstufungstests lässt sich eine gewisse Ähnlichkeit mit Feststellungsprüfungen und Diagnosetests nicht leugnen. Allerdings geht es hier weder um Zertifizierungen noch um gezielte Fördermaßnahmen, sondern lediglich um die Zuordnung zu einem Kurs bzw. zu bestimmten Lernergruppen. Insbesondere sollen hier die Vorkenntnisse von LernerInnen mit unterschiedlichen Niveaus abgefragt werden, um eine entsprechende Einteilung vornehmen zu können. Da die Einstufung möglichst objektiv und innerhalb kurzer Zeit erfolgen soll, finden geschlossene Testformate wie z. B. „*Multiple-Choice-Aufgaben*“ häufiger Verwendung als offene Aufgabenstellungen (z. B. freies Schreiben), die eine längere Korrekturzeit beanspruchen. In diese Testkategorie fällt auch das „*Cloze-Verfahren*“ bzw. der modifizierte Cloze-Test, wo lexikalische bzw. grammatisch-orthographische Phänomene schnell und effektiv abgefragt werden können.<sup>66</sup>

### 5. Eignungstests

*Eignungstests prüfen grundsätzlich die Fähigkeit zum Erlernen einer Sprache* und sind in den testmethodischen Welten, besonders im deutschsprachigen Raum, kaum anzutreffen.

---

<sup>65</sup> vgl. dazu Dlaska (2009), S. 29.

<sup>66</sup> vgl. dazu Perlmann-Balme (2010), S. 997 – 1000.

Obwohl ein Eignungstest, ähnlich wie ein Intelligenztest, geistige Fähigkeiten abfragen soll, wird dieses Testformat oft im gleichen Atemzug mit Zulassungsprüfungen genannt. Aufnahmeprüfungen, wie z. B. jene für ausländische Studienbewerber in Deutschland, wo entsprechende Deutschkenntnisse zu erbringen sind, können sowohl zu den Eignungstests als auch zu den Zulassungsprüfungen gezählt werden.<sup>67</sup>

## 6. Zulassungsprüfungen

Bei Zulassungsprüfungen geht es nicht nur um die Zulassung im engeren Sinn, also beispielsweise um die Zulassung zu einem Studium, sondern allgemein um die Auswahl von Teilnehmern aus einer größeren Gruppe. Bei Zulassungsprüfungen kommt dem Schwellenwert (oder den kritischen Wert), welcher die Grenze zwischen Bestehen und Nicht-Bestehen darstellt, eine große Bedeutung zu.<sup>68</sup>

Zulassungsprüfungen geben in erster Linie innerinstitutionelle Vorstellungen wieder. Sie richten sich nur ansatzweise an Referenzniveaus bzw. Bildungsstandards aus. Es geht generell um spezifische sprachliche Fähigkeiten, die bei der zulassenden Stelle verlangt werden. Die Testformate werden zumeist institutsintern erprobt und erstellt und stehen in unmittelbarer Nachbarschaft zu Einstufungs- und Eignungstests oder auch Feststellungsprüfungen. Insgesamt fungieren Zulassungsprüfungen im Großen und Ganzen als administrative Verfahren mit Selektions- und Steuerungscharakter in Bezug zur Zukunft. Die Ergebnisse haben selten einen Diagnosecharakter mit dem Ziel der Weiterentwicklung von Sprache, sie dienen vielmehr einer Zugangsbeschränkung und sind mit einem hohen Risiko für Menschen mit nichtdeutscher Muttersprache verbunden.<sup>69</sup>

Auf die Zulassungsprüfung Deutsch beim ÖBH lassen sich mehrere Testtypen zuordnen. Zum einen dient die Prüfung der Auswahl aus einer Gruppe von TestteilnehmerInnen (Zulassungsprüfung) und zum anderen kann man die Prüfung durchaus als Kurseinstiegsüberprüfung (Eignungsprüfung) bezeichnen (MilFü2). Wenn man zusätzlich die aus den Ergebnissen abgeleiteten Fördermaßnahmen mitberücksichtigt, was sicher weitestgehend zutrifft, dann handelt es sich hier auch um einen Einstufungs- bzw. um einen Diagnostest.

---

<sup>67</sup> Perlmann-Balme (2010), S. 1000.

<sup>68</sup> Dlaska (2009), S. 30.

<sup>69</sup> vgl. dazu Perlmann-Balme (2010), S. 1000 – 1001.

Nach der Bestimmung des Verwendungszwecks, in unserem Fall die Zulassung, ist es zielführend, die sprachlichen Anforderungen zu definieren. Das heißt, die realen Anforderungen müssen ins Sprachliche umgesetzt werden. Im weiteren Sinn geht es hier um das erforderliche Sprachniveau, das der jeweilige Testteilnehmer erreichen sollte, um positiv abzuschneiden. Im Falle unserer Zulassungsprüfung entscheiden die beruflich zu erwarteten Kommunikationskompetenzen und deren Zuordnung zum allgemeinen Sprachniveau von Pflichtschulabsolventen über den Schwierigkeitsgrad und die Bewertung von Aufgabenstellungen. Besonders zu berücksichtigen ist dabei, dass die Fähigkeiten der Testteilnehmer, insbesondere bei jenen mit nichtdeutscher Muttersprache, oft im Schreiben schlechter entwickelt sind als im mündlichen Gebrauch.

## **5.2 Allgemeine Anforderungen an Sprachtests**

Wie die Sprachtest-Funktionstypen zeigen, bedingt der jeweilige Typ eine adäquate, repräsentative Form. Das bedeutet, dass es hier nicht nur um die terminologische Abgrenzung zwischen Prüfung und Test geht, sondern vielmehr um die Gewichtung und Aufteilung der vier Fertigkeiten, die Aufgabentypen innerhalb eines Testteiles und um die Testlänge. Demnach eignet sich nicht jeder Aufgabentyp für jede Funktion.

Hinsichtlich des Testinhaltes in Verbindung mit der Aufgabenstellung unterscheidet man innerhalb der Testwelten zwischen Performanz- und Kompetenztests bzw. zwischen direkten und indirekten Tests. Ähnlich wie bei Testfunktion und Form ist auch hier ein Zusammenhang auszumachen, obwohl es hier um völlig unterschiedliche Kategorien geht.

Das Begriffspaar „Performanz“ und „Kompetenz“ geht zurück auf den Linguisten *Chomsky*, der zwischen „*Sprachgebrauch (performance)*“ und „*Sprachkompetenz (competence)*“ unterscheidet. Laut Chomsky liegt der jeweiligen konkreten Realisierung von Sprache (Performanz) eine Fähigkeit (Kompetenz) zu Grunde, woraus zwei Umgangsweisen mit Sprachtests resultieren. Die eine (Kompetenztest) misst die allgemeine Beherrschung von Sprache im Hinblick auf verschiedene Sprachverwendungssituationen, während mit dem anderen (Performanztest) bestimmte Fähigkeiten gemessen werden, um spezifische Sprachverwendungssituationen zu meistern. Die Entscheidung, ob und wann ein

Performanztest oder Kompetenztest zur Anwendung kommt, richtet sich meist nach der Funktion.<sup>70</sup>

Obwohl es bei der Unterscheidung zwischen direkten und indirekten Tests lediglich um den Interpretationsbezug bei Testergebnissen geht, gibt es, wie die Praxis zeigt, durchaus Anknüpfungspunkte zur Performanz bzw. Kompetenz. Während bei direkten Tests die gestellten Aufgaben entsprechend den zu messenden Kompetenzbereich(en) so konstruiert sind, dass ohne weitere Interpretation von der Testleistung auf die Wirklichkeit geschlossen werden kann, fällt die Auswertung der Testergebnisse bei indirekten Formaten deutlich schwerer.<sup>71</sup>

Erfahrungsgemäß gelingt die Konstruktion von direkten Testformaten zumindest bei produktiven Schreib- und Sprechaufgaben relativ problemlos, indem der Forderung nach Authentizität weitestgehend nachgegangen werden kann (handlungsorientierter Ansatz). Einerseits finden sich genügend, an die jeweilige Sprachverwendungssituation angepasste, Schreib- und Sprechanlässe, die sich in die Aufgabenstellung gut integrieren lassen. Andererseits kann aus den Ergebnissen der TestteilnehmerInnen zumindest annähernd auf reale Leistungen geschlossen werden. Bei der Überprüfung von Hör- und Leseverstehen scheint dieses Unterfangen völlig aussichtslos, weil sich diese Fertigkeiten nur mittels schriftlichen und mündlichen Ausdrucks (Ankreuzen bzw. Mitschreibens, Mitteilens) messen lassen. Dass eine Testsituation immer künstlich ist und niemals eine reale Situation simulieren kann, sei an dieser Stelle angemerkt. Aus testtheoretischen und pragmatischen Gründen finden sich gegenwärtig wie schon bei den Testtypen meist Mischformen innerhalb der testmethodischen Welten wieder.

Beim Prüfungsformat Diktat innerhalb der Zulassungsprüfung Deutsch handelt es sich im engeren Sinn um einen direkten, formellen, nach der schriftsprachlichen Korrektheit orientierten Sprachtest, wo es um Reproduktion von Gesprochenem geht. Im weiteren Sinn ist er ein informeller Test, der neben der Orthographie auch das Hörverstehen und andere Fertigkeiten erfassen kann. Hinsichtlich der Überlegung, dass Diktate auch im berufsspezifischen Kontext zumindest vereinzelt vorkommen können, kann man das Prüfungsformat auch als Performanztest ansehen. Wenn man aber mitberücksichtigt, dass das Angesagte in der Praxis zumeist nur stichwortartig, also nicht Wort für Wort niedergeschrieben wird, dann liegt wiederum ein Kompetenztest vor.

---

<sup>70</sup> vgl. dazu Dlaska (2009), S. 24 – 25.

<sup>71</sup> vgl. dazu Dlaska (2009), S. 25 – 26.

### 5.3 Qualitäts- und Nützlichkeitskriterien

Eine nicht zu unterschätzende Rolle spielen dabei die nationalen und internationalen Teststandards, die aktuelle Bemühungen um Qualitätssicherung widerspiegeln. Testorganisationen mit Vorbildfunktionen wie ETS in Princeton<sup>72</sup> oder auch das Test-DaF-Institut in Deutschland<sup>73</sup> üben einen wesentlichen Einfluss auf die Testinstitutionen und Testersteller hinsichtlich Funktion und Form aus.

Standardisierte und nicht standardisierte Tests unterscheiden sich laut Lienert und Raatz wie folgt:

Standardisierte Tests müssen wissenschaftlich entwickelt, hinsichtlich der wichtigsten Gütekriterien untersucht und unter Standardbedingungen durchführbar und normiert sein. Im anderen Falle handelt es sich um nicht-standardisierte oder informelle Tests, wie sie Psychologen und Lehrer gewissermaßen für den Hausgebrauch benützen und auswerten.<sup>74</sup>

Bei standardisierten Tests handelt es sich überwiegend um formelle Verfahren, die einer mehrstufigen wissenschaftlich fundierten Evaluierung (Erstellung, Erprobung, Überarbeitung) unterzogen wurden, um verlässliche Aussagen über Leistungen geben zu können.

Je gravierender und weitgehender die mit der Prüfung verbundenen Auswirkungen mit dem Ergebnis verbunden sind, desto höhere Qualitätsanforderungen werden an die Sprachprüfung gestellt. Es gilt sicherzustellen, dass, soweit möglich, ausschließlich die zu messende Eigenschaft – hier die Sprachfähigkeit – und kein anderes Merkmal innerhalb und außerhalb des Kandidaten zum Ergebnis beitragen oder es verfälschen.<sup>75</sup>

Sprachtests sollen demnach die Leistungen der Probanden möglichst genau erfassen, zuverlässige Daten liefern und objektive Bewertungen zulassen. Qualitätsanforderungen an Sprachtests bringen aber durch die strikte Einhaltung eines Standards auch zahlreiche Beschränkungen mit sich, die sich in der Ausformulierung der Gütekriterien widerspiegeln. Ebenso deutlich zeigt sich der noch nicht abgeschlossene Paradigmenwechsel von Sprachwissen zum Sprachkönnen, der insbesondere bei der Reliabilität und Objektivität eine große Rolle spielt. Um einen entsprechenden Qualitätsstandard von Sprachtests sicherzustellen, haben sich in der klassischen Testtheorie drei Hauptgütekriterien, die Sprachtests unbedingt erfüllen sollen, innerhalb der Testwelten etabliert: Validität, Reliabilität und Objektivität.<sup>76</sup>

---

<sup>72</sup> Educational Testing Service (ETS) in Princeton/USA (TOEFL-Sprachtest)

<sup>73</sup> Entwicklung der international anerkannten Deutschprüfung für den Hochschulbereich

<sup>74</sup> Lienert G. A./Raatz, U. (1994): Testaufbau und Testanalyse, Weinheim: Beltz. S. 7.

<sup>75</sup> Kecker, Gabriele (2009): Validität und Validierung von Sprachprüfungen. In: Berndt Annette / Kleppin Karin [Hrsg.]: Sprachlehrforschung: Theorie und Empirie Festschrift für Ruediger Grotjahn. Frankfurt: Lang. S. 129.

<sup>76</sup> vgl. dazu Kecker (2009), S. 129 – 131.

Bachmann und Palmer haben, um den Ansprüchen und Auswirkungen eines kommunikativen Fremdsprachenunterrichts gerecht zu werden, mit einer Weiterentwicklung der klassischen Testtheorie mit Blickrichtung Nützlichkeit von Sprachtests reagiert, indem sie sechs, etwas abgeänderte Kriterien festlegten: *Reliabilität, Konstruktvalidität, Authentizität, Interaktivität, Auswirkungen und Wirtschaftlichkeit*.<sup>77</sup>

Warum Bachmann und Palmer anstatt der Konstruktvalidität die Reliabilität an die vorderste Stelle ihres Kriterienkataloges setzten, ist für mich nur insofern erklärbar, als die beiden Testtheoretiker vom Modell der Testnützlichkeit und Testbrauchbarkeit im Sinne einer Qualitätssicherung ausgehen und deshalb die Forderung nach Zuverlässigkeit und Vergleichbarkeit insgesamt höher einstufen als die Gültigkeit von Sprachtests.

Insgesamt kann man von einer Wechselbeziehung zwischen Validität und Reliabilität sprechen, wo das eine Qualitätsmerkmal Voraussetzung für das andere ist wie auch umgekehrt, je nach Funktion und Form eines Tests.

Da es bei der Evaluierung von Zulassungsprüfungen vorrangig ist, die Gültigkeit des Testinstruments zu überprüfen, gilt für mich, anders als bei Bachmann und Palmer, die Validierung der Sprachprüfung als wichtigster Untersuchungsaspekt.

Die Validität ist jenes Kriterium, das Auskunft darüber gibt, ob ein Test wirklich das überprüft, was überprüft werden soll. Sie gilt als Grundvoraussetzung für jegliche Testerstellung. Ein Test kann aber nicht als Ganzes valide sein, *sondern zu validieren sind nur die Schlussfolgerungen, die aus dem Ergebnis gezogen werden und die Entscheidungen, die darauf folgen*.<sup>78</sup>

Bachmann und Palmer setzen an Stelle des allgemein gehaltenen Begriffs Validität die Bezeichnung „Konstruktvalidität“, um vor allem *die Testersteller dazu anzuleiten, eine klare Definition des Testkonstrukts vorzunehmen*.<sup>79</sup> Die Bezeichnung „Konstruktvalidität“ soll laut Bachmann und Palmer aber auch als Oberbegriff verstanden werden, welcher verschiedenste Merkmale der Validität im Zusammenhang mit der Interpretation von Testergebnissen zusammenfasst.

Sie richten sich nach der Argumentation von *Messick*, der die *Konstruktvalidität* als übergeordnetes, alles Übrige einschließendes Merkmal definiert. Messick misst räumt damit auch den Konsequenzen, die sich aus einer spezifischen Testverwendung bzw.

---

<sup>77</sup> vgl. dazu Bachmann, L. F., Palmer, A. S. (1996) *Language testing in practice; Designing and developing useful language tests* (Oxford Applied Linguistics). Oxford: Oxford University Press. S. 38.

<sup>78</sup> vgl. dazu Dłaska (2009), S. 36.

<sup>79</sup> vgl. dazu Kecker (2009), S. 138.

Testauswertung ergeben, eine nicht nur auf die Validierung und Validität bezogene sondern vor allem gesellschaftliche Bedeutung ein.<sup>80</sup>

Allerdings scheint mir diese Bezeichnung verwirrend, wenn man bedenkt, dass Konstruktvalidität vielfach auch als ein Teilaspekt der Validität bezeichnet wird. Bachmann und Palmer räumen noch dazu ein, dass die anderen Komponenten wie Interaktivität, Authentizität und Auswirkungen ebenso der Validität zugeschrieben werden können.

Dass die Auffassung und Gewichtung von Validität sehr unterschiedlich sein kann, zeigen vor allem die Veränderungen der Begrifflichkeit im Laufe der letzten 20 Jahre, wo Differenzierungen wie z. B. Inhalts- und Augenscheinvalidität und ein einheitlicher, alles umfassender Begriff von Validität einander abwechselten bzw. gleichzeitig hervortraten. Trotz der weiten Verbreitung und Anerkennung von Messicks Argumentationen scheint die Definition von Validität gegenwärtig noch nicht ausreichend geklärt zu sein.<sup>81</sup>

Um die mit dem Begriff Validität verbundenen Schlussfolgerungen über einen Sprachtest ausführlicher zu erläutern, will ich zumindest auf vier Teilaspekte genauer eingehen, die meines Erachtens für meine Studie relevant sind.

Bei der Konstruktvalidität gehe ich vom Konzept Messicks aus, der sowohl die *Inhaltsvalidität* als auch die *Kriterienbezogene Validität* als Unterbegriffe einordnet.

Unter Konstruktvalidität ist demnach die Gültigkeit der auf der Basis der Testergebnisse vorgenommenen Interpretationen zu verstehen. Die Konstruktvalidität überprüft somit die Interpretation des Testergebnisses, ob sie der zugrundeliegenden Hypothese über die Überprüfung von Sprachfähigkeit (Konstrukt) entspricht. Sie kann deshalb als Basis für alle weiteren qualitätsbezogenen Überprüfungen begriffen werden.<sup>82</sup>

Die *Inhaltsvalidität* soll überprüfen ob die gestellten Aufgaben innerhalb der Testformate eine repräsentative Beispiele aus einem bestimmten zu überprüfenden Sprachbereich sind. Die Repräsentativität von Testaufgaben ist erfahrungsgemäß nur schwer empirisch überprüfbar. Voraussetzung ist in jedem Fall, dass zuvor eine wohldefinierte Grundgesamtheit von Aufgaben erstellt wurde und dadurch die Werte direkt in Bezug zur Grundgesamtheit interpretiert werden können. Allenfalls kann ein Testersteller durch Ausweitung der Testaufgaben die Inhaltsvalidität möglichst hoch halten. Ob und wie weit es überhaupt möglich ist, mittels einer künstlichen Testsituation, die repräsentativ für ein Ganzes ist, das ja

---

<sup>80</sup> vgl. dazu Dlaska (2009), S. 36.

<sup>81</sup> vgl. dazu Kecker (2009), S. 131 – 132.

<sup>82</sup> vgl. dazu Dlaska (2009), S. 36.

auch künstlich interpretiert wird, eine tatsächliche Sprachverwendungssituation wiederzugeben, sei hier zusätzlich angemerkt.<sup>83</sup>

Anders als bei der Inhaltsvalidität stehen bei der *Vorhersagevalidität* zukünftige Leistungen im Bereich des Testkonstrukts im Mittelpunkt der Qualitätssicherung. Es gilt festzustellen, ob anhand der erbrachten Testergebnisse im Bezug zum Testkonstrukt zukünftige Leistungen voraussagbar sind oder nicht. Wie schon bei der Inhaltsvalidität lässt sich dieser Faktor nur durch großen Aufwand, wie durch mehrjährige Untersuchungen überprüfen, indem man z. B. das Ergebnis einer Kurs-Einstufung mit jenem der Kursabschlussprüfung vergleicht.<sup>84</sup>

Die *Augenscheinvalidität* steht im engen Zusammenhang mit der individuellen Beurteilung durch den Testteilnehmer selbst. Sie beschränkt sich meist auf eine Durchsicht der Testaufgaben z. B. im Zuge einer Testevaluierung mittels Fragebogen. Entscheidend ist, ob der Test den betroffenen Teilnehmern gefällt und die Aufgabenstellungen plausibel und transparent erscheinen. Eine nicht zu unterschätzende Rolle spielt hier auch die äußere Form des Tests, die oft den Eindruck verfälschen kann. Aus testtheoretischer Sicht nimmt die Augenscheinvalidität aufgrund ihrer empirischen Vagheit - keine Experten, sondern Laien überprüfen ein Testformat - eine eher vernachlässigbare Größe innerhalb der Qualitätssicherung von Sprachtests ein.<sup>85</sup>

Für die schriftliche Zulassungsprüfung Deutsch beim ÖBH sind meiner Meinung nach alle vier Teilaspekte der Validität wichtig und entscheidend. Wenn die Ergebnisse nicht dem zugrundeliegenden Konstrukt entsprechen, das heißt, die gewollten berufsspezifischen sprachlichen Erfordernisse sich im Ergebnis nicht zeigen, weil das Testkonstrukt nicht klar genug definiert bzw. das Ausmaß der Schlussfolgerungen falsch eingeschätzt wurde, so ist der Test als Prüfmittel unbrauchbar. Gleichermäßen unbrauchbar ist das Testformat, wenn die Voraussagen aufgrund eines Testergebnisses kaum bzw. überhaupt nicht zutreffen. Gerade bei einer Zulassungsprüfung hätte dieser Befund fatale Folgen, die Deutschkenntnisse würden trotz erfolgreich bestandener Prüfung für andere Kursgegenstände nicht ausreichen. Nicht zu unterschätzen ist hier auch die Augenscheinvalidität bei Zulassungsprüfungen. Der erste Blick, der erste Eindruck ist oft ausschlaggebend, nicht nur für die Akzeptanz des Tests,

---

<sup>83</sup> vgl. dazu Kecker (2009), S. 133.

<sup>84</sup> vgl. dazu Dlaska (2009), S. 37.

<sup>85</sup> vgl. dazu Glaboniat (1998), S. 23 – 25.

sondern auch für die Motivation bei der Testdurchführung und schließlich auch für die Testleistung.

Die Ergebnisse eines Sprachtests sollten aber nicht nur eine möglichst hohe Validität aufweisen, sondern auch vergleichbare Ergebnisse sowohl innerhalb des Testformats als auch außerhalb hervorbringen. Ziel des Testens soll sein, Messfehler und Störfaktoren möglichst gering zu halten, damit der reale Messwert und der gemessene Wert so gut wie möglich übereinstimmen.

Testtheoretiker sprechen hierbei von der *Reliabilität*, einem Qualitätsmerkmal, das sich in erster Linie auf die Konsistenz und die Stabilität der Testwerte bezieht. Innerhalb der Reliabilität wird grundsätzlich unterschieden zwischen der „inneren Konsistenz“, wo es um die Konsistenz der Testaufgaben untereinander geht sowie der *Testwiederholungs-* und *Paralleltestreliabilität*, wo die Stabilität der Testwerte eine große Rolle spielt. Die Reliabilität sagt allerdings nichts über die Relevanz der Testwerte aus, sie misst lediglich die Verlässlichkeit der Testwerte. Relevant kann allerdings jenes Faktum sein, dass inkonsistente Aufgaben innerhalb eines Testformats den Testteilnehmern unterschiedliche Rangplätze zuordnen, was erfahrungsgemäß zu Problemen bei der Gesamtbewertung führen muss. Auch unterschiedliche Werte bei einer Testwiederholung bzw. bei einer anderen Testversion sollten hinsichtlich der Messgenauigkeit nicht ignoriert werden.<sup>86</sup>

Bei der *Auswerterreliabilität* hingegen, „*die Auskunft darüber geben soll, wie zuverlässig und präzise ein Test misst, indem der Übereinstimmungsgrad bei der Bewertung ein und derselben Leistung durch mehrere Bewertungspersonen untersucht wird*“<sup>87</sup> stehen andere Kriterien wie ein klar strukturierter Auswertungsraster bzw. Auswertungsmodus sowie möglichst eindeutig formulierte Auswertungsrichtlinien im Vordergrund. Hier hilft meist auch ein entsprechendes Testdesign, damit Messfehler reduziert werden, und der wahre und der errechnete Messwert durch den/die Auswertungsperson/en so weitgehend wie möglich übereinstimmen.

Auf die Zulassungsprüfung Deutsch Bezug nehmend, nimmt die Reliabilität zwar keine Hauptrolle ein, aber es lassen sich durchaus einige relevante Anforderungen daraus ableiten. Bei einer Prüfung wie der Zulassungsprüfung Deutsch, wo das „berufliche Überleben“ im Vordergrund steht, kann die Testwiederholungsreliabilität nie zur Gänze erfüllt werden. Die

---

<sup>86</sup> vgl. dazu Dlaska (2009), S. 35 – 36.

<sup>87</sup> Glaboniat (1998), S. 38.

Gründe für die unterschiedlichen Leistungen einer Testperson sind meist nicht auf mangelnde Stabilität zurückzuführen, sondern liegen eher im psychologischen bzw. lernstrategischen Bereich. Dennoch sollten die Parallelversionen bei einer Zulassungsprüfung annähernd den gleichen Schwierigkeitsgrad aufweisen, um möglichst verlässliche Ergebnisse zu liefern.

Das Problem der inneren Konsistenz hingegen trifft auf das Diktat nicht zu, da es sich hier um einen durchgehenden, authentischen Text handelt, der keine Einzelaufgaben enthält. Wenn man hingegen die gesamte Zulassungsprüfung untersuchen würde, müsste man zumindest die drei Leseverstehensaufgaben, die unter eine Gesamtbewertung fallen, auf Konsistenz überprüfen.

Ein wesentlicher Aspekt innerhalb der Reliabilität ist das Nichtkennen bzw. Kennen des Testformats. Das betrifft insbesondere das Diktat. Faktoren wie „fehlende Vertrautheit mit dem Testformat“ bzw. „*teaching to the test*“ verfälschen das Testergebnis, indem Personengruppen, die das Testformat nicht kennen, meist schlechter abschneiden gegenüber jenen, die ein gezieltes Testvorbereitungstraining absolvieren bzw. zum wiederholten Mal die Prüfung ablegen. Testwiederholungen und Ergebnisse mit anderen, unbekanntem Testformaten haben gezeigt, dass die gleichen beiden Personengruppen (mit und ohne Testerfahrung) unterschiedliche Ergebnisse erzielen.<sup>88</sup>

Entscheidend für die verlässliche Auswertung eines Diktats ist ein erprobter Korrekturmodus, wo sowohl der Umgang und die Bewertung von Fehlern (Wiederholungsfehler, unleserliche Stellen, Flüchtigkeitsfehler usw.) als auch die Fehlermarkierung klar und handhabbar geregelt sein müssen.

Innerhalb der klassischen Testtheorie wird unter *Objektivität* jenes Kriterium angesehen, das die Unabhängigkeit des Testens, Auswertens und Interpretierens misst. Meiner Meinung nach zu Unrecht wird dieses Gütekriterium von Testtheoretikern als rangniedrigstes Qualitätsmerkmal angesehen bzw. scheint bei Bachmanns und Palmers Kriterienkatalog nur als Teilbereich der Reliabilität auf. Dabei können gerade störende und situationsabhängige Umstände allgemeiner Art, wie z. B. akustische Probleme, die oft nur einzelne Kursteilnehmer betreffen sowie das Übersehen von Fehlern bei Auswertern zu Benachteiligungen führen. Nicht zu vergessen sind hier Menschen mit nichtdeutscher Muttersprache, deren Sprachkenntnisse vielfach nicht objektiv beurteilt werden wie z. B.

---

<sup>88</sup> vgl. dazu Schifko, Manfred (2001): „Prüfungen, Zertifikate, Abschlüsse als Planungskategorien für den Unterricht“. In: Helbig, Gerhard u. a. [Hrsg.]: Deutsch als Fremdsprache. Ein internationales Handbuch. 2. Halbband, Berlin: de Gruyter S. 830 - 832.

beim Diktat, wo Hörverstehensprobleme sich als Grammatik- oder Orthographiefehler zeigen. Es geht daher, bei der Objektivität nicht nur um die Objektivität der Beurteilung, sondern auch um die Bedingungen, unter denen der Test stattfindet.<sup>89</sup>

Demnach spielt gerade bei der Zulassungsprüfung Deutsch Objektivität eine nicht zu unterschätzende Rolle. Alleine wenn man den Umstand betrachtet, dass die Prüfungen an unterschiedlichen Orten und zu unterschiedlichen Zeiten (vor Beginn des Unteroffizierslehrgangs, am Ende eines Förderseminars usw.) stattfinden bzw. unterschiedliche Testaufsichtspersonen die Prüfungen durchführen.

Nach der Reliabilität und Konstruktvalidität erwähnen Bachmann und Palmer in ihren Qualitätskriterien den Begriff *Authentizität* als drittes Nützlichkeitsmerkmal.<sup>90</sup> Meiner Meinung betrifft dieses Qualitätskriterium in erster Linie die Konstruktvalidität, wo es um den unmittelbaren Zusammenhang zwischen der realen Sprachverwendungssituation und der Testaufgabe, also auch um den Einsatz möglichst authentischer Materialien geht.

Dass es sich bei einem Test immer um ein *künstliches Produkt* in Verbindung mit einer *künstlichen Situation* handelt, die nur annähernd die Sprachrealität wiedergeben kann bzw. dass authentische Testformate nicht unbedingt nützlich sein müssen, sei hier zusätzlich angemerkt. Außerdem ist Authentizität nur mit großem Aufwand empirisch untersuchbar. Wie schon bei der Konstruktvalidität erwähnt, würden hier nur langjährige Studien nützen, wo allerdings der Zeitfaktor (z. B. hat der/die Proband/in zwischen dem Testergebnis und dem Studienergebnis Erfahrungen gesammelt, die jene Ergebnisse beeinflussen würden?) wiederum das Ergebnis verfälschen würde.<sup>91</sup>

Beim Prüfungsformat Diktat tritt ein zusätzliches Problem in den Vordergrund, das die Einhaltung der Authentizität beeinträchtigen kann. Bei der Wiederholung von Prüfungen ist es nur annähernd machbar, eine gleich schwierige Textaufgabe bei strikter Einhaltung des Authentizitätsprinzips zu finden. So gesehen relativiert sich die Nützlichkeit von authentischen Materialien, insbesondere wenn es um Niederschreiben von Gehörtem geht.

Ein nicht zu unterschätzendes Gütekriterium stellt die *Interaktivität* dar. Bei Bachmann und Palmer wird unter Interaktivität das „*Ausmaß, mit dem die Sprachfähigkeit, das Wissen über das Thema und das Interesse der Teilnehmer an der Erfüllung der Aufgabe*“<sup>92</sup> aktiviert wird,

---

<sup>89</sup> vgl. dazu Glaboniat (1998), S. 37 – 38.

<sup>90</sup> vgl. dazu Dlaska (2009), S. 37 – 38.

<sup>91</sup> vgl. dazu Glaboniat (1998), S. 48 – 50.

<sup>92</sup> Bachmann, Palmer (1996), S. 25.

verstanden. Es soll daher, analog zum Sprachunterricht, die Testperson in einen Lernprozess eingebunden werden, um eine möglichst hohe Motivation zu erreichen. Allerdings ist dieses Kriterium, ähnlich wie die Authentizität meiner Ansicht nach nur schwer empirisch messbar.

Wie die Interaktivität beim Diktat innerhalb der Zulassungsprüfung Deutsch zu beurteilen ist, können nur Fragebögen und Interviews klären. Entscheidend ist hier auch der Zwang, die Prüfung unbedingt bestehen zu müssen. Das Interesse der Teilnehmer an der Erfüllung der Aufgabe sollte gerade bei Zulassungsprüfungen, wo es um Berufschancen geht, in jedem Fall differenziert betrachtet werden. Wenngleich es tatsächlich Testpersonen gibt, die trotz Prüfungsdruck den Test als sinnvolle Lernaktivität sehen.

Im unmittelbaren Zusammenhang mit der Interaktivität stehen die *Auswirkungen eines Test*. Zum einen stehen hier Auswirkungen auf den Lern- und Lehrprozess auf dem Prüfstand, zum anderen sollen die Auswirkungen auf die Gesellschaft, insbesondere auf das Bildungssystem untersucht werden. Im Wesentlichen geht es um spezifische pädagogische Maßnahmen aufgrund einer Testung. Ein großangelegter Test, wie die PISA-Studie z. B., hat bis heute weitreichende bildungspolitische Folgen, sei es auf pädagogische Konzepte oder auf zukünftige Prüfungen bezogen.<sup>93</sup>

Die Wirkung von Tests kann daher durchaus positive Folgen nach sich ziehen, wenn die TestteilnehmerInnen den Sinn und die Bewertungskriterien des Tests verstehen und daraus Schlussfolgerungen für ihren eigenen Lernprozess ziehen können. Hilfreich sind hier Rückmeldungen, also genaue Feedbacks bezüglich des Testergebnisses, die einen direkten Kontakt zwischen Auswerter und Testperson herstellen. Insgesamt lassen sich zumeist nur vage Aussagen über die Auswirkungen von Tests treffen. Oft sind es nur die negativen Ergebnisse, die den Bekanntheitsgrad eines Tests ausmachen. Auswirkungen des Diktats, wie bei der Zulassungsprüfung Deutsch, gibt es genügend. Das Prüfungsformat Diktat hat aufgrund der nachfolgenden Konsequenzen einen großen Einfluss auf die Art und Weise der Prüfungsvorbereitung (*Backwash-Effekt*).

TestkandidatInnen lernen zumeist nicht für die Zukunft, sondern ausschließlich für den Test. Dieser *Rückkopplungseffekt* beeinflusst auch den jeweiligen Deutsch-Förderunterricht, insbesondere das Curriculum, wo das Bestehen der Prüfung im Vordergrund steht (*teaching to the test*). Es geht daher nicht mehr um Sprachkönnen, auch nicht um Sprachwissen, sondern vielmehr um das Training von Lernerstrategien.<sup>94</sup>

---

<sup>93</sup> vgl. dazu Dlaska (2009), S. 38 – 39.

<sup>94</sup> vgl. dazu Schifko (2001), S. 830 - 832.

Außerdem sind die gesellschaftlichen Auswirkungen hinsichtlich des Bestehens und Nichtbestehens der Prüfung von großer Bedeutung. Sei es nur der vielfach negativ propagierte Umstand, dass Nichtmuttersprachler Sprachtests aufgrund unzureichender Sprachkorrektheit (Grammatik, Orthographie) oft nicht bestehen. Dass es bei Menschen mit nichtdeutscher Muttersprache auch bzw. hauptsächlich um Hörverstehen geht, scheint offenbar nur wenige zu interessieren.

Der Aspekt der *Praktikabilität* ist, wie die Praxis zeigt für Testersteller in vielerlei Hinsicht von Interesse. Bei vielen bereits standardisierten Tests stehen erfahrungsgemäß weder die Validität noch die Reliabilität, sondern ökonomische Überlegungen an erster Stelle. Dieser Umstand beruht darauf, dass zumeist nur eingeschränkte Ressourcen wie Personal und technische Geräte zur Verfügung stehen. Relevant ist auch der jeweilige Zeitaufwand, der für das Korrigieren und Interpretieren der Testergebnisse notwendig ist. So werden oft aus praktikablen Gründen äußerst nützliche, kreative Testformate verworfen. Das betrifft in erster Linie offene, kommunikativ orientierte Testformate, deren Erstellung und Auswertung meist einen hohen Einsatz an Personal und technischen Mitteln verlangt. Während ein *Multiple-Choice-Test* mittels einer Schablone sehr rasch und effizient ausgewertet werden kann, ist die Auswertung von offenen Schreibaufgaben wesentlich aufwendiger. Um die Beurteilungszuverlässigkeit von freien Textaufgaben möglichst hoch zu halten, bedarf es nicht nur der Erstellung eines Beurteilungsrasters, sondern auch einer laufenden Schulung der Testauswertungspersonen. Allerdings sind offene freie Testaufgaben wesentlich leichter zu erstellen als geschlossene Antwortformate wie beim Multiple-Choice-Test, der hinsichtlich Aufgabenschwierigkeit und Aufgabentrennschärfe ein mehrstufiges Prüfverfahren mit einer Vielzahl an Stichproben verlangt.<sup>95</sup>

Bei der Einführung des Diktats innerhalb der Zulassungsprüfung Deutsch ist die Entscheidung zugunsten des Testformats offensichtlich aufgrund ökonomischer Überlegungen gefallen. Ob das Diktat aufgrund der Aufgabenerstellung und des Auswertungsverfahrens tatsächlich praktikabler ist als so manche freie Aufgabenstellung, soll die Untersuchung zeigen.

Eine Prüfung des Diktats anhand der Nützlichkeitskriterien von Bachmann und Palmer erscheint mir für meine Untersuchung relevant und passend.

---

<sup>95</sup> vgl. dazu Dlaska (2009), S. 39.

## 6. Der Testgegenstand: Die Rechtschreibfähigkeit

### 6.1 Der Fehlerbegriff in der Forschung

In wieweit sind die Rechtschreibfähigkeit bzw. korrektes Schreiben relevant und aussagekräftig für den erfolgreichen Gebrauch von Sprache beim ÖBH? Welche Auswirkungen haben die Ergebnisse auf die Deutschförderung?

Um die Untersuchungsbasis der Zulassungsprüfung auch in schreibdidaktischer Hinsicht abzusichern, gilt es, auf den Testgegenstand „Orthographische Korrektheit“, insbesondere aber auch auf den Begriff „*Fehler*“ näher einzugehen. Bei den nachfolgenden Ausführungen folge ich weitgehend den Argumentationen von *Karin Kleppin* und *Silke Demme*.<sup>96</sup>

Orthographiefähigkeit bedeutet, dass der Lerner entsprechend dem geforderten Leistungsniveau gelernte Wörter in Textzusammenhängen korrekt schreiben und noch nicht gelernte Wörter durch orthographisches Regelwissen ableiten kann.

Hinzu kommt noch die Beziehung des Wörterbuches bei jenen Wörtern, die nicht bzw. nur unzureichend durch orthographisches Regelwissen erschlossen werden können (Fremdwörter usw.). Auffällig ist, dass bei empirischen Studien über die Rechtschreibfähigkeit stets narrative Texte im Zentrum der Untersuchung standen und nur selten Diktate verwendet wurden. Dass schriftliche Fehler einen besonderen Stellenwert beim Diktat einnehmen, verdeutlichen die Beurteilungskriterien, wo nicht das Gekonnte, sondern ausschließlich die „*Fehlerleistung*“, zum Ergebnis führen soll.

Sieht man von der Tatsache ab, dass zahlreiche Deutschlehrer, aber auch Testauswerter, bei der Beurteilung von schriftlichen Arbeiten nur mehr jene Rechtschreibfehler in Betracht ziehen, die das Verständnis des Textes beeinträchtigen, scheint der Umgang mit Fehlern nach wie vor nicht ausreichend geklärt zu sein. Zentrale Fragen wie: „Was sind überhaupt Fehler? Wie klassifiziert und gewichtet man Fehler? Welche Fehlerquellen gibt es?“ scheinen an Aktualität kaum verloren zu haben.

Fehler sind eine natürliche Gegebenheit, sowohl für Fremd- und Zweitsprachenlerner als auch für Muttersprachler. Sie können generell als Abweichung von einem Regelsystem oder als Verstoß gegen eine sprachliche Norm bzw. gegen den (normalen) Sprachgebrauch verstanden werden. Fehler können auch eine Nichterfüllung von lernstands- und testbezogenen Kriterien

---

<sup>96</sup> Demme, Silke (2007): „... wo Fehler sind, da ist auch Erfahrung“ (Tschechov) – auch beim Lehren und Lernen von Fremdsprachen. In: Eßer/Krumm (Hrsg.): Bausteine für Babylon: Sprache, Kultur, Unterricht ... Festschrift zum 60. Geburtstag von Hans Barkowski. Iudicium, München. S. 206 – 216.  
Kleppin, K. (1998): Fehler und Fehlerkorrektur. Fernstudieneinheit 19, Langenscheidt München.

darstellen. Fehler entstehen aber auch durch Unaufmerksamkeit oder durch Fehlervermeidung. Generell bleibt es der Lehrperson oder dem Testauswerter überlassen, ob und in welcher Form (Klassifizierung, Gewichtung) eine Abweichung von der Norm und ein Verstoß gegen das Regelsystem (Grammatik, Wörterbuch und Institution), als Fehler gewertet werden.<sup>97</sup>

Spätestens seit der Einführung des Faches Deutsch als Zweitsprache haben namhafte Sprachwissenschaftler wie *Corder* und *Edge* erkannt, dass allgemeine Klassifikations- und Kategorisierungsvorschläge für die Analyse von Fehlern zumeist nicht ausreichen. Dieser Annahme zur Folge soll eine Fehleranalyse grundsätzlich nur unter besonderer Berücksichtigung des Spracherwerbs sowohl bei Muttersprachlern als auch bei Menschen mit nichtdeutscher Muttersprache (Einfluss der Muttersprache) erfolgen. In diesem Zusammenhang wurde auch erkannt, dass eine Interpretation bzw. eine klare Abgrenzung von Fehlern nur bedingt möglich ist, da Fehler meist nicht offen bzw. in anderer Form hervortreten. Ein Beispiel dafür sind Fehler in der Groß- und Kleinschreibung beim Diktat, die sich bei genauerer Analyse nicht als orthographische Fehler, sondern als syntaktisches Nichtwissen (Was ist ein Subjekt?) entpuppen.<sup>98</sup>

In der modernen Schreibdidaktik wird bei orthographischen Fehlern grundsätzlich zwischen Fehlern, die eine Aussage unverständlich machen (Kommunikationsabsicht war eine andere als in der schriftlichen Äußerung) und Fehlern, die keine Kommunikationsbehinderung (Sinn wird verstanden) aber einen Regelverstoß bedeuten, unterschieden. Dem entsprechend werden Fehler, die das Verständnis eines Textes nicht beeinträchtigen, nur bei großer Anzahl in die Beurteilung miteinbezogen.<sup>99</sup>

Ein zentrales Anliegen der gegenwärtigen Spracherwerbsforschung ist die Erklärung von *Fehlern und Fehlerursachen* (Fehlerquellen). Bei Menschen mit nichtdeutscher Muttersprache z. B. sollte neben dem Einfluss der Muttersprache bzw. anderen Fremd- und Zweitsprachen auch der Einfluss durch die Lernaltersprache (Deutsch) selbst berücksichtigt werden (Simplifizierungen). Fehler sind zumeist nicht nur das Resultat von sprachimmanenten Faktoren, sondern auf persönliche Störfaktoren (Stress, Angst, Müdigkeit) bzw. auch auf motorische Schwächen des Lernalters zurückzuführen (Schwächen in Hören und Schreiben beim Diktat). Nicht zu unterschätzen ist hier der Einfluss von soziokulturellen Faktoren

---

<sup>97</sup> vgl. dazu Demme (2007), S. 206 - 208.

<sup>98</sup> vgl. dazu Kleppin (1997), S. 19 - 22.

<sup>99</sup> vgl. dazu Kleppin (1997), S. 14 - 16.

bei jenen Lernern, die auf Grund ihrer Herkunft und Sozialisation ein geringeres kulturelles und landeskundliches Wissen mitbringen.<sup>100</sup>

Mitberücksichtigt sollten dabei auch die Lehrpläne der Schulen, wo Deutsch als Zweitsprache explizit unterrichtet wird:

#### **Rechtschreiben**

Von Anfang an sollte danach getrachtet werden, dass der Prozess des Schreibens aktiv vollzogen wird, dh. er sollte vom Sensorischen zum Kognitiven und zum Motorischen führen (Hinschauen/Hören/Merken/Aufschreiben).

Spezifische Schwierigkeiten beim Erwerb der deutschen Rechtschreibung ergeben sich vor allem aus Unterschieden zu den Muttersprachen der Schüler und Schülerinnen, besonders bei:

- Laut-Buchstaben-Entsprechungen,
- Abgrenzung von Wörtern,
- Wortzusammensetzungen,
- Groß- und Kleinschreibung,
- Satzzeichen.

(Abbildung 10: Auszug aus dem Lehrplan der 1. – 4. Klasse Hauptschule: Internet– Zugriff 04. 12. 2011)

Angesichts dieser Tatsachen sollte sich die Fehlerdidaktik keinesfalls nur auf das Zählen, Klassifizieren und Errechnen von Fehlern beschränken, auch nicht bei einer Zulassungsprüfung (Diktat). Das Sanktionieren und Dämonisieren von Fehlern (Fehler als Schande!) führt vielfach zu einer negativen Einstellung gegenüber Korrekturen. Im schlimmsten Fall sind sogar Rückschritte im Lernprozess bemerkbar. Dies wird etwa auch in der Leistungsbeurteilung von Deutschschularbeiten für den AHS-Bereich festgehalten:

Auch laut LBVO sind für die Beurteilung von Schularbeiten aus Deutsch in jedem Fall die Bereiche Inhalt, Ausdruck, Sprachrichtigkeit und Schreibrichtigkeit in gleichem Ausmaß heranzuziehen. Die Rechtschreibung darf also keinesfalls die hauptsächliche oder gar einzige Grundlage der Leistungsbeurteilung sein, und sie darf auch für die Beurteilung von Schularbeiten nicht hauptsächlich herangezogen werden. Verordnungskonform akzeptiert die Arbeitsgemeinschaft für Deutsch auch keine stringenten Fehlerobergrenzen bei der Beurteilung von Schularbeiten.<sup>101</sup>

Diktatkorrekturen wären nur dann didaktisch wertvoll, wenn die Möglichkeit besteht, dass Schüler/Testpersonen ihre Texte selbst korrigieren und reflektieren. Fehler sollten nicht als Mangel, sondern als Chance erkannt werden, sie sind Inhalt jeglichen Lernens. Fehler sind Teil unseres sprachlichen Verhaltens, sie gehören zu uns wie unser Charakter. In pädagogischer Hinsicht sind Fehler vor allem wichtige Hilfsmittel zur Lernfortschrittserfassung.

<sup>100</sup> vgl. dazu Kleppin (1997), S. 29 – 40.

<sup>101</sup> vgl. dazu aus dem § 16 (1) der Leistungsbeurteilungsverordnung von Deutschschularbeiten bei Schülern der allgemein bildenden höheren Schulen.

„Ein Glück, dass Schüler Fehler machen!“<sup>102</sup>

Kein(e) Deutschlehrer(in) beherrscht vollständig die deutsche Orthographie, auch Deutschlehrer (innen) machen Fehler! Auch Normen sind nicht ohne Fehler (Normierungsprozesse), sie müssen immer wieder korrigiert werden.

Für die Zulassungsprüfung Deutsch bedeuten diese Überlegungen, dass das Kriterium „Sprachliche Korrektheit“ immer im Zusammenhang mit anderen Kriterien stehen sollte wie z. B. Wortschatz, Syntax, Aufgabenbewältigung im Sinne einer Gesamtkompetenz.

## 6.2 Expertise zum Diktat aus der Fachwelt

**diktieren** Vb. „vorsprechen (zum Nachschreiben), befehlen, aufzwingen“, entlehnt (15. Jh.) aus gleichbed. lat. *dictāre*, Intensivum zu lat. *dīcere* „sagen“ (s. *zeihen*). Das substantivierte Part. Perf. lat. *dictātum* „Diktirtes“, *dictāta* Plur. „Regeln, Vorschriften“ wird entlehnt als **Diktat** n. „Ansage, Nachschrift“ (Anfang 17. Jh.), „Gebot, aufgezwungene Verpflichtung“ (Anfang 19. Jh.).<sup>103</sup>

Gegenwärtig wird orthographisch einwandfreies Schreiben noch immer als sozialdifferenzierendes Kriterium gesehen, obwohl im Zeitalter von computerbasierenden Korrekturprogrammen orthographische Fehlleistungen weitestgehend ausgeschaltet werden sollten. Gerade vor dem Hintergrund von Migration und PISA-Studien wird orthographische Korrektheit oft zum Kriterium für erfolgreiche Bildung und Integration.<sup>104</sup>

Das Diktat verdankt seine Beliebtheit offenbar der Einfachheit seiner Durchführung in Verbindung mit der scheinbaren eindeutigen Auswertbarkeit und Interpretierbarkeit, die sich meist nur in Form von Fehlerzählen äußert. Außerdem scheint das Diktat tatsächlich vergleichbare und objektive Beurteilungen zuzulassen.

Bis in die späten 80er-Jahre galt das Diktat als aussagekräftiges und bewährtes Prüfungsformat als Übungs- sowie als Testformat insbesondere im schulischen Betrieb. Dieser Umstand darf aber nicht darüber hinwegtäuschen, dass bereits 1967 der Linguist Robert Lado gegenüber dem Diktat als Testaufgabe eine durchaus kritische Position vertrat:

<sup>102</sup> Krumm, Hans-Jürgen (1990): „Ein Glück, dass Schüler Fehler machen!“ Anmerkungen zu Umgang mit Fehlern im lernerorientierten Fremdsprachenunterricht. In: Leupold, Einar/Yvonne Petter (Hrsg.): Interdisziplinäre Sprachforschung und Sprachlehre, Festschrift für Albert Raasch, Tübingen.

<sup>103</sup> DWDS (Wörterbuch der deutschen Gegenwartssprache): Internet-Zugriff 04.12.2011

<sup>104</sup> vgl. dazu Krumm (1990), S. 99 – 102.

„Das Diktat ist ein Verfahren, das im sprachlichen Bereich wenig zu messen vermag. Mit dem Diktat können Orthographie, einige wenige Erscheinungen der Flexion und die Interpunktion gemessen werden“<sup>105</sup>.

Lado verstand das Diktat lediglich als Rechtsschreibtest während Angela Oakeshott-Taylor, aber auch John Oller sich dieses Prüfungsformat durchaus auch als Hörverstehensüberprüfung vorstellen konnten. Oller definierte das Diktat *als integratives Messverfahren, wo ein weites Spektrum an sprachlichen Fertigkeiten erfasst werden – darunter als wichtige Komponente auch das Hörverstehen*. Für dieses Argument sprechen für Oller vor allem die *hohen Korrelationen zwischen Ergebnissen im Diktat und Summenwerten aus Testbatterien diverser fremdsprachlicher Fähigkeiten und Fertigkeiten*.<sup>106</sup>

Gegenüber während des Hörens nicht lösbaren MC-Aufgaben, meint hingegen Oakeshott-Taylor, *hat das Diktat den Vorteil, dass es das Verstehen des sprachlichen Inputs gleichsam online misst und dass damit die Ergebnisse weit weniger als bei entsprechenden MC-Aufgaben von interindividuellen Unterschieden in der Gedächtnisspanne abhängen*. Zudem liefert das Diktat in gewissem Ausmaß auch *diagnostische Informationen – z. B. im Hinblick auf Schwächen bei der Phonemdiskrimination als potentielle Ursache für lexikalische Fehler wie z. B. ‘wise presidency’ für ‘vice presidency’*.<sup>107</sup>

Von großem Interesse für meine Untersuchung ist insbesondere auch Gert Solmeckes Expertise hinsichtlich der Textschwierigkeit in Verbindung mit akustischen Komponenten. So gesehen kann man auch beim Diktat Störungen akustischer Art beim Diktat nicht ausschließen.

---

<sup>105</sup> Lado, Robert (1971): Testen im Sprachunterricht. Handbuch für die Erstellung und den Gebrauch von Leistungstests im Fremdsprachenunterricht. Aus den Amerikanischen von R. Freudenstein. München. (Original: Lado, Robert (1961) Language testing: The construction and use of foreign language tests. London: Longmans.) S. 50.

<sup>106</sup> vgl. dazu Oller, John W., Jr.(1979): Language tests at school: A pragmatic approach. London: Longman. Kap. 10, Macht 1997 sowie auch Kap. 1.7.4.

<sup>107</sup> vgl. dazu Oakeshott-Taylor, Angela (1977): „Dictation as a test of listening comprehension“. In: Dirven, René (Hrsg.): Hörverständnis im Fremdsprachenunterricht/Listening comprehension in foreign language teaching. Kronberg, Ts.: Scriptor, S. 150.

### Schwierigkeitsfaktoren der akustischen Komponente

(aus Solmecke 2000, S. 63)

Textschwierigkeit						
	← eher leicht					eher schwierig →
<i>Akustische Komponente</i>		1	2	3	4	
Störgeräusche	leise					laut
Raumakustik	gut					schlecht
Aufnahmequalität	gut					schlecht
Sprechgeschwindigkeit	langsam					schnell
Artikulation	deutlich					undeutlich
Intonation	gliedemd					monoton
Pausen zwischen Sätzen und Satzteilen	eher länger					eher kürzer

(Abbildung 11: Auszug aus Solmeckes Faktoren der Schwierigkeit von Hörtests)

Die unterschiedlichen wissenschaftlichen Positionen zeigen, dass die Problematik hinsichtlich des Diktats schon frühzeitig erkannt wurde, dass aber weitere empirische Untersuchungen notwendig schienen, um einer Lösung näher zu kommen.

Darüber hinaus, meint Oller, *sind Diktate vergleichsweise leicht zu erstellen und differenzieren in der Regel gut zwischen Lernern unterschiedlicher Kompetenz. Die Auswertung (Scoring) von Diktaten ist für ihn allerdings nicht unproblematisch, da zumeist von der Gesamtzahl der Wörter des Ausgangstextes für jedes falsche oder fehlende Wort ein Punkt abgezogen wird und dabei eindeutige Rechtschreibfehler in der Regel unberücksichtigt bleiben.*<sup>108</sup>

Nach Oller ist das Diktat ein weiteres Testverfahren, das integrierte Fertigkeiten möglichst valide überprüft. Allerdings unterscheidet sich das Diktatverfahren bei Oller von einem herkömmlichen Diktat dahingehend, dass zum einen die Einheiten beim Vorlesen länger sind und zum anderen, dass die Vorlesegeschwindigkeit normalem Sprechtempo entspricht, wo nicht alle Fehler, sondern ausschließlich sinnentstellende Rechtschreibfehler berücksichtigt werden. Nach mehreren Test-Untersuchungen, wo Oller faktorenanalytisch vorging, resümierte er, dass es keinen Subtest gäbe, der eindeutig eine gewisse Fertigkeit misst. So gesehen, ist auch die genaue Trennung der vier Fertigkeiten in Frage zu stellen, also die Hypothese, auf der so manches Testkonstrukt aufbaut, dass Sprachbeherrschung zu trennen sei. Das heißt, dass die vier Sprachfertigkeiten mehr Gemeinsamkeiten aufweisen, als bisher angenommen. Oller argumentiert so mit einem psycholinguistischen Konstrukt einer Erwartungshaltung.<sup>109</sup>

<sup>108</sup> vgl. dazu Oller (1979), Longman. Kap. 10, Macht 1997 sowie auch Kap. 1.7.4.

<sup>109</sup> vgl. dazu Oller (1979), Longman. Kap. 10, Macht 1997 sowie auch Kap. 1.7.4.

Wie diese Ausführungen zeigen, liefern die Ergebnisse des Diktats zwar zahlreiche unterschiedliche diagnostische Informationen, die sich aber aufgrund des Systemwechsels und der Transkodierung nur schwer interpretieren lassen. Angesichts dieser Hypothese scheint es fraglich, ob das Diktat gesicherte Auskünfte über die tatsächlichen Orthographiefähigkeiten geben kann. Insgesamt handelt es sich beim Diktat um einen nach der schriftsprachlichen Korrektheit orientierten Sprachtest, wo es um Reproduktion von Gesprochenem geht. Im engeren Sinn geht es um einen informellen Test, der neben der Orthographie auch das Hörverstehen und andere Fertigkeiten erfassen kann.

Bedingt durch zahlreiche Probleme bei der Bewertung von produktiven kreativen Leistungen, sei es in der Festlegung von Beurteilungskriterien und Deskriptoren oder in den Rahmbedingungen zahlreicher Testinstitute, fungiert das Prüfmittel Diktat vielerorts, wie auch beim ÖBH, als Platzhalter für so manche Neuentwicklungen von Testverfahren.

Im gegenwärtigen schulischen Betrieb bzw. in heeresinternen Sprachkursen dienen Diktate dagegen in erster Linie nicht der Leistungskontrolle, sondern vielmehr der Fehlerdiagnose beim Erwerb orthographischer Fähigkeiten. Sie finden aus schreibdidaktischen Gründen ausschließlich zu Übungszwecken Verwendung (Partner-, Lückendiktate usw.).

## **7 Forschungsziele – Hypothese**

### Gegenstandsbezeichnung, Testgegenstand

Gegenstand meiner eigenen Untersuchung, die im Folgenden dargestellt wird, sind die zu erwartenden sprachlichen Kompetenzen eines Berufsunteroffiziers bei der Produktion von Texten.

Die Definition des Testgegenstandes scheint in diesem Fall schwierig. Sieht man den Testgegenstand als Allgemeinsprache mit Fachausdrücken oder geht es um die kommunikationsspezifischen Voraussetzungen innerhalb des Fachgebietes?

Die Militärsprache gleicht jedenfalls einer Mischung aus Führungs- und Fachsprache. Neben spezifischen Fachausdrücken ist sie geprägt von typischen syntaktischen und morphologischen Phänomenen. Dabei stehen die funktionale und die soziale Bedeutung im Vordergrund.

## Zielsetzung dieser Arbeit

Die Zielsetzung meiner Arbeit ist es, zum einen zu untersuchen, ob das Prüfungsformat Diktat den Zweck der Zulassung zu einer Sprachfördermaßnahme erfüllen kann. Zum anderen soll überprüft werden, ob sich anhand der Diktatergebnisse das beim Militär verlangte berufsspezifische Schreiben messen lässt. Zusätzlich sollen auch außersprachliche Faktoren, die das Testergebnis beeinflussen können wie z. B. Migrationshintergrund, Bildungsstand oder psychische und motorische Schwächen untersucht werden.

## Hypothese

Die Ergebnisse der schriftlichen Zulassungsprüfung Deutsch liefern nur unzureichende Informationen über jene schriftsprachlichen Qualifikationen, die den militärischen Anforderungsprofilen entsprechen. Insbesondere lassen die Ergebnisse nur unzureichend Rückschlüsse auf sprachliche Mängel zu und erfüllen den Zweck der Zulassung zu Sprachfördermaßnahmen nicht. Insbesondere verfälschen außersprachliche Faktoren die Prüfungsergebnisse.

## **8 Methodischer Ansatz – empirische Vorgangsweise**

Um das gewählte Untersuchungsmaterial empirisch erfassen und meine Hypothese überprüfen zu können, habe ich mich für ein mehrperspektivisches Vorgehen in Zusammenhang mit einem explorativen (explanativen) Design entschieden. Gemäß aktueller Forschungsmethodik hat sich bei der Evaluierung von Testverfahren die Kombination verschiedener Datenerhebungsmethoden bewährt. Nicht zuletzt auch, um das Validitätsproblem quantitativer Methoden sowie das Generalisierungsproblem qualitativer Methoden zu lösen.

In der empirischen Sozialforschung wird für die Zusammenführung unterschiedlicher Datenerhebungsmethoden zumeist der Begriff „*Triangulation*“<sup>110</sup> verwendet<sup>111</sup>.

Eine genauere Beschreibung der Begrifflichkeit erfolgt bei *Popp*, die vier Anforderungen an einen effektiven Einsatz von Forschungsmethoden formuliert:

---

<sup>110</sup> Bezeichnung stammt aus der Landvermessung, wo ein Punkt mittels eines Netzes von Dreiecken fixiert wird.

<sup>111</sup> vgl. dazu Stigler, Hubert/Reicher, Hannelore [Hrsg.]: Praxisbuch Empirische Sozialforschung in den Erziehungs- und Bildungswissenschaften. Innsbruck-Wien-Bozen: Studienverlag 2005. S. 88 – 92.

1. *Die Methodentriangulation sollte aus mindestens zwei verschiedenen Datensätzen bestehen, die auf einen Untersuchungsgegenstand gerichtet sind und sich in der Interpretation aufeinander beziehen.*
2. *Mit einer Kombination unterschiedlicher Forschungsmethoden sowohl qualitativer als auch quantitativer Natur können Vor- und Nachteile kompensiert werden.*
3. *Eine „gelungene“ Methodentriangulation beruht auf einem nahezu ausgewogenen Verhältnis zwischen qualitativen und quantitativen Daten, das sich auch in der Präsentation niederschlagen sollte.*
4. *Zeitlich versetzte Erhebungssituationen haben den Vorteil, dass man in Folgeuntersuchungen relevante Entdeckungen gezielt aufgreifen kann.*<sup>112</sup>

Angesichts dieser Ausgangsposition habe ich für meine Untersuchung unterschiedliche methodische Zugänge gewählt. Als Datenquellen fungieren drei repräsentative Stichproben, zwei für die Untersuchung des vorhandenen Prüfmittels Diktat und eine für die Vergleichsstudie (Datentriangulation). Zur Überprüfung meiner theoretisch angeleiteten Hypothese habe ich das Datenmaterial mit folgenden vier unterschiedlichen Methoden untersucht:

- Interpretation der Testergebnisse durch Häufigkeitsanalyse und durch Mittelwerte
- Datenerhebung durch Fragebögen
- Datenerhebung durch Experten-Leitfadeninterviews
- Datenerhebung durch teilnehmende Beobachtung.

## **8.1 Bewertung und Erfassung der Testauswertungsdaten**

Eine meiner Ansicht nach wesentliche Komponente meiner empirischen Untersuchung der Zulassungsprüfung Deutsch ist die Erfassung der Prüfungsergebnisse aller Probanden bei der Stichprobe. Da es sich bei den Stichproben um Teilnehmerergebnisse aus Echtprüfungen handelte, sollten daraus untersuchungsrelevante Schlussfolgerungen hinsichtlich der Qualitäts- bzw. Gütekriterien eines Sprachtests gezogen werden. Insbesondere wurden auch vom Durchschnitt abweichende, extreme Fälle genauer geprüft und in Relation zu den Auswertungsergebnissen der Fragebögen gestellt. Generell galten für mich die Testauswertungsdaten als Basis für weitere untersuchungsmethodische Schritte.

Die Erfassung der Testauswertungsdaten erfolgt grundsätzlich in zwei Schritten. Zuerst werden die Daten (Anzahl der Fehler, Anzahl der fehlenden Worte) in eine eigens dafür

---

<sup>112</sup> vgl. dazu Popp, Ulrike (2002): Geschlechtersozialisation und schulische Gewalt. Weinheim: Beltz S. 287.

vorgesehene, computergenerierte Excel-Tabelle eingegeben und danach in einer Übersicht graphisch dargestellt. Aufgrund des festgelegten Fehlerkoeffizienten wird errechnet, ob die Prüfung (Teilprüfung) als bestanden oder nicht bestanden gewertet wird. Die Tabellendaten geben aber nicht nur Auskunft über das Prüfungsergebnis, sondern auch über die Person selbst. Das bedeutet, dass Auswertungen auch aufgrund von Teilungskriterien wie Alter, Geschlecht und Muttersprache durchgeführt werden können, um personalspezifische Auswirkungen auf das Testergebnis feststellen zu können.

## **8.2 Fragebogenuntersuchung**

Damit in der Untersuchung nicht nur der Forschungsgegenstand, sondern auch das Subjekt, die Versuchsperson zu Wort kommt, habe ich mich für eine schriftliche Fragebogenerhebung entschieden. Die Fragebogenuntersuchung soll als Blick in das Bewusstsein der Betroffenen begriffen werden. Es gilt dabei, die Messwerte einer Stichprobe anhand von Themenkomplexen zu strukturieren und in Beziehung zu bringen. Die quantitative Auswertung der Ergebnisse soll vor allem Aufschluss über untersuchungsrelevante Themenbereiche geben und meine theoretischen Überlegungen stützen. Die Fragebogenerhebungen erfolgten in manueller Form unmittelbar nach den Testungen.

Auf Basis meiner Erkenntnisse aus den wissenschaftlichen Recherchen bzw. aufgrund meiner Erfahrungen aus der Voruntersuchung entwickelte ich einen Fragebogen, der nach folgenden Kriterien aufgebaut ist:

Auf der ersten Seite finden sich allgemeine Informationen über die Deutschförderung beim ÖBH, über das Ziel und den Gegenstand der Untersuchung sowie über den Nutzen und die Verwendung des Fragebogens. Bei der Textfindung war es wichtig für mich, dass die Befragten nicht genau erfahren, um welche Hypothese es in der Untersuchung geht, damit das Befragungsergebnis nicht durch persönliches Kalkül („Das Diktat sollte abgeschafft werden!“) beeinflusst wird. Grundsätzlich geht es mir der ersten Seite des Fragebogens darum, die Wertschätzung gegenüber den Befragungsteilnehmern auszudrücken, um die Motivation möglichst hoch zu halten.

Die Reihung der Themenblöcke richtet sich sowohl beim Fragebogen „Diktat“ als auch bei der „Vergleichsstudie“ neben inhaltlichen Kriterien nach dem Interesse und der Konzentrationsfähigkeit der Befragten. Bei der Ausarbeitung des Fragebogens habe ich versucht, eine möglichst natürliche Sprache zu wählen, um die Aufmerksamkeit und das Interesse der Befragten hoch zu halten. Mittels der direkten Einbeziehung der befragten Person in die Formulierungen (ich, bei mir) ist es zumindest bei einigen Themen gelungen,

einen objektivierenden Fragebogenstil zu vermeiden. Bei der Formulierung von Fragen geht es zumeist um Feststellungen, die zumeist mittels der hier gezeigten „verbalen Ratingskala“<sup>113</sup> zu bewerten sind.

trifft zu	trifft überwiegend zu	trifft eher nicht zu	trifft nicht zu
1	2	3	4

Beim persönlichen Eindruck bzw. bei der Selbsteinschätzung ist hingegen mit „ja“ oder „nein“ bzw. mit „sehr gut“, „gut“, „mittel“ und „eher schlecht“ zu antworten.

Bis auf die Anmerkungen am Schluss handelt es sich ausschließlich um geschlossene Fragen, die sich quantitativ auswerten lassen. Während plausible Fragen wie z. B. „Die äußere Form des Tests löst Missfallen aus“ am Anfang stehen, finden sich komplexe Fragen wie z. B. „*Sprachverwendung beim Bundesheer*“ erst in der Mitte des Bogens. Im zweiten Teil des Bogens geht es größtenteils um persönliche Eindrücke und Einschätzungen sowie um einige Fragen zur Person, die erfahrungsgemäß mehr Zeit in Anspruch nehmen. Am Schluss hat der/die Befragte die Möglichkeit, eigene Gedanken hinzuzufügen. Was beim Fragebogen fehlte, war die Frage nach der Reihenfolge der gelernten Sprachen (Tertiärsprachenforschung!), insbesondere, ob Englisch vor Deutsch, also Deutsch erst als zweite fremde Sprache gelernt wurde.

Durch den Fragebogen lässt sich nicht nur ermitteln, wie viel Prozent der Teilnehmer eine Feststellung mit beispielsweise „trifft zu“ bewertet haben, sondern es lassen sich auch Bezüge zwischen der jeweiligen Bewertung und den persönlichen Angaben herstellen.

Neben Themenkomplexen, die die Versuchsperson aufwerten sollen (Einschätzungen, Vergleiche, Fragen zur Person) finden sich im Fragebogen Feststellungen, die im unmittelbaren inhaltlichen Zusammenhang zur Forschungsfrage stehen. Obwohl mein Fragebogendesign durch die Art und Weise der Bewertung kaum Spielraum für die Befragten zulässt, habe ich zumindest thematisch versucht, die Personen als reflektierende „Versuchspartner“ zu behandeln.

---

<sup>113</sup> Stigler, Hubert (2005), S. 145.

### 8.3 Teilnehmende Beobachtung

Im Zuge der Erhebung der Untersuchungsdaten nützte ich die Möglichkeit, gezielte Beobachtungen durchzuführen, die sich in erster Linie auf das Verhalten der Teilnehmer während der Erfüllung der Testaufgabe(n) konzentrierten.

Dabei habe ich mich für die Methode der offenen, teilnehmenden Beobachtung entschieden, bei der es gilt, *Erkenntnisse über das Handeln und Verhalten oder die Auswirkungen des Verhaltens von einzelnen Personen oder Gruppen zu gewinnen.*<sup>114</sup>

Die Beobachtungsergebnisse hielt ich ausschließlich stichwortartig fest, Videoaufnahmen wurden keine durchgeführt. Um als Forscher nicht in Versuchung zu kommen, im Sinne der Hypothese die Objektivität bei der Beobachtung zu vernachlässigen, habe ich für die zweite Stichprobe eine fachlich geschulte Person mit der Beobachtung betraut.

Um eine Auswahl aus den möglichen Beobachtungskategorien zu treffen, wählte ich solche aus, die im unmittelbaren Zusammenhang mit der Fragestellung meiner Untersuchung standen:

- Gibt es manche TestteilnehmerInnen, die Angst zeigen?
- Gibt es raumakustische Unterschiede während des Abspielens?
- In welcher Schriftart (Schriftart, Druckschrift) wird das Diktat verfasst und wie wirkt sich die Schriftart auf die Schreibgeschwindigkeit aus?
- Wie verhalten sich Menschen mit nichtdeutscher Muttersprache beim Niederschreiben des Diktats während der zweiten Abspielphase?
- Wie wird von den Testteilnehmern die dritte Abspielphase des Diktats genutzt? Was machen dabei Muttersprachler, was Personen mit nichtdeutscher Muttersprache?
- Gibt es TestteilnehmerInnen, die aufgrund von Ermüdung aufgeben und nicht mehr weiter schreiben?
- Sind gegen Ende des Diktats Konzentrationsschwächen zu beobachten?
- Gibt es Unterschiede zwischen jenen TestteilnehmernInnen, die das Diktat zum ersten Mal machen und denen, die bereits zum zweiten oder zum dritten Mal antreten?

---

<sup>114</sup> vgl. dazu Albert, Ruth/Marx, Nicole [Hrsg.]: Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht. Tübingen: Narr Verlag 2010. S. 43 - 44.

## 8.4 Experteninterview

Wie im Kapitel „Methodischer Ansatz“ bereits erläutert, genügt es oft nicht nur eine Methode innerhalb einer Untersuchung anzuwenden.

Zumeist bleiben relevante Fragen offen bzw. es ergeben sich aus dem Befund der ersten Stichproben neue Fragen, die für die Thematik wichtig sind. Dem zufolge haben sich Experteninterviews als Ergänzung bewährt, wenn sie auf Basis theoretischer Vorkenntnisse in Verbindung mit ersten empirischen Ergebnissen fungieren. Sie sollen Ergebnisse klären, die sich aus dem Gebrauch mit anderen Untersuchungswerkzeugen wie z. B. bei der Fragebogenuntersuchung ergeben.

Als Experten werden jene Personen angesprochen, die in irgendeiner Weise Verantwortung tragen für den Entwurf, die Implementierung oder die Kontrolle einer Problemlösung oder die über einen privilegierten Zugang zu Informationen über Personengruppen oder Entscheidungsprozesse verfügen.<sup>115</sup>

Um in einem Interview über das Forschungsthema möglichst konkrete Antworten zu bekommen, habe ich mich für ein Leitfadeninterview entschieden. In der Sozialforschung soll ein Leitfadeninterview als ein strukturiertes Konzept begriffen werden, bei dem in einer festgelegten Reihenfolge vorformulierte Fragen zu möglichst konkreten Antworten führen.

*„Der Leitfaden schneidet die interessierenden Themen aus dem Horizont möglicher Gesprächsthemen heraus und dient dazu, das Interview auf diese Themen zu fokussieren“.*<sup>116</sup>

Durch die festgelegten Themenkomplexe kann das Interview sowohl mit anderen Interviews als auch mit den Ergebnissen der Fragebögen verglichen werden.

Der Nachteil von Leitfadeninterviews liegt meiner Meinung nach in der Einschränkung der Antwortmöglichkeit. Demnach sind bei Erstellung des Leitfadens die Frageabfolgen und Fragestellungen weitgehend offen zu gestalten, dass der interviewten Person eine möglichst freie Formulierung zugestanden wird.

Angesichts dieser Tatsachen hielt ich es für angemessen, ein Probeinterview mit einem Experten aus dem gleichen Fachbereich zu führen. Dabei sollten die Reihenfolge und die Relevanz der Fragen überprüft werden, um unter anderen auch Nachfragethemen auszuformulieren. Als Vorbereitung dazu fertigte ich einen umfangreichen Leitfadencatalog an, um auf möglichst alle für die Forschungsarbeit relevanten Themen aufgreifen zu können.

---

<sup>115</sup> Meuser, Michael & Nagel, Ulrike (1991): Experteninterviews – vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In.: Qualitative- empirische Sozialforschung: Konzepte, Methoden, Analysen. Hrsg. von Garz, Detlev. Kraimer, Klaus. Opladen S. 443.

<sup>116</sup> Meuser (1991), S. 195.

Nach der Durchführung des Probeinterviews galt es für mich, folgende Fragen für das Leitfadenterview besonders zu berücksichtigen:

- *Gibt es Fragen für die Anlaufphase bzw. Fragen zur fachlichen Kompetenz der Person? Werden diese als motivierend bzw. wertschätzend empfunden?*
- *Beziehen sich die Fragen auf den Arbeitsbereich, auf das Forschungsgebiet des Interviewpartners?*
- *Gibt es heikle Fragen, die Probleme bei der Beantwortung bereiteten?*
- *Ist ein thematischer Verlauf in meinem Interview erkennbar?*
- *Wird den Aufforderungen zu Stellungnahmen und Begründungen entsprechend nachgekommen oder wird den Fragen ausgewichen?*
- *Gibt es Fragen, die unverständlich oder mehrdeutig sind?*
- *Gibt es Fragen, die nicht beantwortet wurden? Warum?*
- *Stellten die Fragen eine Überforderung bzw. Unterforderung des Interviewpartners dar?*
- *Waren die Fragen für den Interviewpartner interessant?*<sup>117</sup>

Wie aus meinen Fragestellungen hervorgeht, gibt es unterschiedliche Gesichtspunkte, die ich bei meinen Leitfadenterviews berücksichtige. In der Anlaufphase, das heißt in den ersten Fragen des Leitfadens ging es mir darum, das Vertrauen im Interviewpartner zu wecken. In Mittelpunkt standen dabei Fragen zur Person in Bezug zu deren fachwissenschaftlichen Werdegang. Danach ging es überwiegend um Fragen, die unmittelbar den Untersuchungsgegenstand, die Zulassungsprüfung Deutsch betreffen. In dieser Gesprächsphase sollte der Interviewpartner sein Wissen und seine Meinungen bzw. seine Erfahrungen mit dem Untersuchungsgegenstand darlegen.

Aus vertrauensbildenden Gründen finden sich problembezogene, sensible Fragen deshalb eher im zweiten Teil des Interviews wieder. Jeder Teil ist in Detailfragen, die gezielt zum jeweiligen Themenkomplex hinführen sollen, untergliedert. Insgesamt finden sich auch Fragen, die auf die gesamte Zulassungsprüfung Deutsch, also andere Aufgaben, Bezug nehmen. Ein Grund dafür sind mögliche Zusammenhänge der einzelnen Aufgabenstellungen. Auszuschließen ist hier keineswegs der psychologische Faktor, der meines Erachtens eine wesentlichen Rolle bei der Beantwortung der Fragen spielt.

---

<sup>117</sup> vgl. dazu Stigler (2005), S. 130 – 132.

Insgesamt gewährleistet der Leitfaden nicht nur mir, sondern auch dem Interviewten, alles gesagt bzw. nichts vergessen zu haben. Darüber hinaus bietet der Interviewleitfaden die Möglichkeit, zurückblättern zu können, um Themenkomplexe nochmals aufzugreifen.

Demzufolge war es für mich auch wichtig, den beiden Experten die Fragen bereits vor dem Interview auszuhändigen, damit sie die Möglichkeit erhalten, sich sowohl auf die Chronologie, als auch auf die Fragen vorzubereiten.

In zeitlicher Hinsicht waren für die Interviews jeweils etwa eineinhalb Stunden vorgesehen.

Keineswegs unbeachtet möchte ich die Aufnahmeart des Interviews lassen. Die beiden Experteninterviews wurden nach Zustimmung der Befragten mittels Diktiergerät aufgenommen. Zusätzlich notierte ich die Gespräche stichwortartig, um bei etwaig auftretenden technischen Problemen keine Daten zu verlieren.

## **9 Machbarkeitsanalyse und Sicherstellung der Untersuchungsvoraussetzung**

### **9.1 Zugangsmöglichkeiten zu den Daten**

Um die Evaluierung eines Testverfahrens durchführen zu können, bedarf es geeigneter Rahmenbedingungen und Strukturen. Es galt für mich daher im Vorfeld das Verständnis der Umsetzung auf heeresinterner Ebene zu erreichen. Entscheidend dabei war, all jene Personen dabei zu berücksichtigen und einzubeziehen, die mit der Erstellung, Auswertung, Umsetzung und Genehmigung des Testverfahrens beauftragt waren. Neben dem Leiter des Sprachinstitutes wurde dabei insbesondere die AG Prüfmittel Deutsch in Betracht gezogen, die sich aus Personen des Sprachinstitutes, der Heeresunteroffizierschule, des Streitkräfteführungskommandos, aber auch aus Angehörigen des Bundesministeriums für Landesverteidigung zusammensetzt. Im Zuge von Abstimmungs- und Arbeitsgesprächen wurden sowohl die Zulassung von Auswertungen, Befragungen, Interviews und Beobachtungen als auch die Zugriffsmöglichkeiten auf Datenbanken geregelt.

Nicht zu vergessen ist hier das Einverständnis der Test- und Befragungsteilnehmer, das ich sowohl bei den Verantwortlichen als auch jeweils zu Beginn jeder Stichprobe einholen musste. Insbesondere galt es für mich, bei vertiefenden qualitativen Untersuchungsmethoden wie dem Fragebogen und dem Experteninterview, die auf großteils privatem Datenmaterial basieren, entsprechend den Datenschutzrichtlinien zu handeln.

## **9.2 Personell und finanziell verfügbare Mittel**

Sieht man von der Tatsache ab, dass die Befragung der Testteilnehmer und die Experteninterviews einen entsprechenden zeitlichen Mehraufwand für die Beteiligten darstellten, war nur ein geringfügiger personeller Mehraufwand notwendig, da die Beteiligten aufgrund ihres beruflichen Aufgabenfeldes ohnehin mit der Zulassungsprüfung konfrontiert waren. Als gleichermaßen niedrig kann auch der finanzielle Mehraufwand der Untersuchung eingestuft werden. Die Mehrkosten beschränken sich lediglich auf die Anschaffung von Fachliteratur (acht Bücher) und das Kopieren der Befragungsbögen. Zusätzlich zu erwähnen ist der Ankauf eines Aufnahmegerätes für die Experteninterviews, das ich mit eigenen Mitteln anschaffte.

## **9.3 Datenlieferanten**

Bei den Datenlieferanten handelt es sich ausschließlich um Angehörige des Bundesheeres, sei es bei den Testversuchspersonen oder bei den Experten. Um möglichst repräsentative Ergebnisse zu erhalten, war es notwendig, die Auswahl der Versuchspersonen aus einer realen Probandengruppe zu treffen. Aufgrund der vorausgegangenen Absprachen im Zuge der AG Prüfmittel stand eine ausreichende Anzahl an Probanden zur Verfügung. Da ich mit meiner Untersuchung ein heeresinternes Testinstrument überprüfen will, das nicht zuletzt auf Basis heeresinterner Überlegungen entwickelt wurde, ergab sich die Notwendigkeit, auch Experten aus prüfungsrelevanten Fachbereichen zu befragen. Was die generelle wissenschaftliche Beurteilung des Testformat Diktat betrifft, habe ich auch die Meinung heeresexterner Sprachdiagnostiker in meine Untersuchung miteinbezogen. Besonders zu erwähnen sind hier die fachkundigen Auskünfte von Manuela Glaboniat sowie Drorit Lengyel.

Bei der Erfassung der untersuchungsrelevanten Daten (Stichproben) wurden ausschließlich reale Prüfungsorte ausgewählt wie Hörsäle der Heeresunteroffiziersakademie bzw. Räume der kursführenden Dienststellen im Garnisonsbereich.

## **9.4 Voruntersuchung**

Um die Machbarkeit der Untersuchung, insbesondere die Durchführung und Auswertung der Stichproben, sicherzustellen, führte ich eine Voruntersuchung an zwei kleinen möglichst repräsentativen Probandengruppen (etwa 10 Teilnehmer) durch. Es sollten zum einen der Einfluss von Störfaktoren und zum anderen auch mögliche Gefahren für die Zuverlässigkeit

der Ergebnisse untersucht werden. Als sehr aufschlussreich erwies sich vor allem der erstmalige Einsatz des Fragebogens (Entwurf) in Verbindung mit einer kurzen Nachbesprechung. Dabei zeigte sich, dass etwa 20% der Fragen nicht beantwortet oder missverstanden wurden. Insbesondere stellte sich dabei heraus, dass ich bei einigen Fragen die fachliche Kompetenz der Probanden falsch einschätzte. Auch die Fragebogenlänge schien in Anbetracht der vorher zu absolvierenden schriftlichen Sprachprüfung und der Zeit zum Ausfüllen zu lang. Indessen konnte durch eine entsprechende Sitzordnung der Probanden eine Beeinflussung durch andere Untersuchungsteilnehmer weitestgehend verhindert werden. Auch die Reihung der Fragen schien dem Interesse und der Aufmerksamkeit der Probanden zu entsprechen.

Angesichts dieser Ergebnisse und Beobachtungen wurde der Fragebogen überarbeitet, wobei die Anzahl der Fragen von 28 auf 21 reduziert und auch der Beantwortungsschlüssel geringfügig verändert wurden.

Zusätzlich zu den erwähnten Stichproben führte ich auch ein Probeinterview bei einem Experten durch. Gleichermaßen wie beim Fragebogen hatte ich die fachliche Komplexität der Fragen unterschätzt. Ich musste daher auch beim Interviewleitfaden Änderungen und Reduzierungen vornehmen.

## **10 Forschungsplan**

Um den Zeitrahmen, aber auch die Möglichkeiten und Problemfelder meiner Untersuchung besser einschätzen zu können, sah ich es als notwendig an, einen Forschungsplan anzufertigen.

Aufgrund dieser Überlegung verfasste ich einen skizzenhaft chronologisch aufgebauten Leitfaden, um die jeweiligen zu erwartenden Schritte meiner Untersuchung plausibel und anschaulich darzustellen. Insbesondere ging es mir um untersuchungsleitende Fragestellungen, die zur Hypothesenbildung und schließlich zur Verifizierung oder Falsifizierung der Hypothese führen sollten. Naturgemäß ging es auch um die Frage, wie viel Zeit und welche Materialien für meine Untersuchung zur Verfügung stehen. Dass es hier natürlich auch Überschneidungen gibt, sei zusätzlich erwähnt.

Welche relevanten Forschungsschritte und welche Evaluationsphasen für meine Untersuchung zur Anwendung kommen, zeigt der folgende Ausschnitt meines Forschungsplans:

### Phase 1 – Vorbereitung und Sicherung der Untersuchung: Oktober 2010 – Jänner 2011:

1. Bestandsaufnahme der für die Forschung relevanten theoretischen Grundlagen (Auflistung von militäreigenen Schriften sowie fachlichen und problembezogenen wissenschaftlichen Arbeiten)
2. Beschreibung des Untersuchungsgegenstands anhand von militäreigenen Schriftstücken bzw. Fachliteratur
3. Militärinterne Besprechungen bezüglich Genehmigungen, Finanzierungen, Personalhilfen für die Untersuchung und Datenlieferanten (Festlegung der Probandengruppe für Stichproben)
4. Beurteilung der Vor- und Nachteile von Forschungsmethoden wie Interview, Fragebogen, Beobachtung
5. Entwicklung eines Fragebogens bzw. Leitfadeninterviewbogens
6. Erarbeitung von Fragestellungen und Hypothesen

### Phase 2 - Empirische Vorerhebungen: November 2010 – Jänner 2011

7. Einholung von Expertentipps hinsichtlich des Untersuchungsdesigns (Glaboniat, Lengyel, Döll, Zecha)
8. Modifizierung der Fragestellungen, Hypothesen
9. Theoretische Untersuchung: Erstellung eines Berufskompetenzmodells
10. Durchführung einer kleinen Stichprobe und eines Probeinterviews zur Evaluierung der Fragebögen bzw. des Interviewleitfadens

### Phase 3 – Durchführung und Auswertung der Stichproben und Interviews: Dezember 2010 – April 2011

11. Durchführung von zwei repräsentativen Stichproben (gültige Zulassungsprüfungen) mittels vorhandener Prüfmittel im aktuellen Prüfungsmodus
12. Erfassung der Prüfungsergebnisse mittels Stichproben
13. Auswertung der Fragebögen
14. Befragung von Experten
15. Interpretation der Ergebnisse, Hypothesentestung

### Phase 4 - Entwicklung des neuen Prüfmittels und Vergleichsstudie: März 2011 – September 2011

16. Erstellung eines neuen Prüfmittels als Ersatz für das Diktat

17. Vergleichsstudie anhand einer Stichprobe (Diktat, neues Prüfmittel)

18. Auswertung der Fragebögen

19. Interpretation der Ergebnisse, Schlussfolgerungen, Ausblick

#### Phase 5 – Verschriftlichung der Arbeit: August 2011 – Dezember 2011

Obwohl ich den Zeitrahmen für die Untersuchung etwas unterschätzt hatte – die Erhebung und Auswertung der Stichproben dauerte um zwei Monate länger als anberaumt (die gesamte Untersuchung dauerte etwa dreizehn Monate) - glaube ich, sagen zu können, dass sich die Anfertigung eines Forschungsplans aus mehreren Gründen gelohnt hat. Zum einen konnte ich anhand dieses Grobgerüsts meine Forschungstätigkeit plausibel darstellen und begründen, um die Untersuchungsvoraussetzungen beim ÖBH abzusichern. Zum anderen nützte ich den Forschungsplan, um mit erfahrenen Sprachwissenschaftlern über meine Untersuchung zu sprechen.

### **11 Beschreibung der Stichproben**

Im Zentrum meiner empirischen Forschungsarbeit steht jenes Belegmaterial, das zur Modifizierung meiner Hypothese führen soll: *die Stichprobendaten*. Sie dienen dazu, Auskunft zu verschiedenen, theoretisch abgeleiteten Fragenkomplexen in repräsentativer Form zu geben. Wie die wissenschaftliche Praxis zeigt, ist die Repräsentativität einer Stichprobe für die Aussagekraft einer wissenschaftlichen Arbeit von entscheidender Bedeutung. Ich habe daher für meine Untersuchung eine Durchschnittsgruppe hinsichtlich Alter, Geschlecht, Herkunft und sprachlicher Vorbildung ausgewählt, um möglichst präzise Rückschlüsse auf die Gesamtheit der Betroffenen zu erhalten. Demnach sind bei den Stichproben sowohl mindestens eine weibliche Testperson als auch Personen mit nichtdeutscher Muttersprache vertreten. Das Durchschnittsalter reicht von 22 bis 24 Jahren. Hinsichtlich der *Herkunft* handelt es sich bei allen drei Stichproben (inkl. Stichprobe der Vergleichsstudie!) ausschließlich um Personen mit österreichischer Staatsbürgerschaft, die großteils ihren Pflichtschulabschluss erfolgreich absolviert haben. Bei jenen, die im Fragebogen eine andere Muttersprache als Deutsch angeben, lassen sich unterschiedliche Migrationszeiten feststellen. Zum einen gibt es Personen, die in Österreich die gesamte Pflichtschulzeit absolviert haben und zum anderen gibt es auch Personen, die erst im Erwachsenenalter nach Österreich übersiedelt sind. Die Bundesländerverteilung innerhalb der

drei Stichproben ist annähernd ausgeglichen. Es sind zumeist mindestens sechs Bundesländer vertreten.

Die Frage, wie viele Untersuchungspersonen zur Verfügung stehen, stellte sich bei meiner Untersuchung nicht, weil ich ohnehin die gesamte Probandengruppe eines Prüfungsvorganges zur Verfügung hatte. Die Wirklichkeitstreue und Forschungsbezogenheit war durch den vorgegebenen Untersuchungsablauf (zuerst Test, danach Fragebogen) und die reale Situation weitestgehend gegeben. Um sicherzustellen, dass Abweichungen in den realen Bedingungen wie unterschiedliche Prüfungsorte (Ebelsberg und Enns) und Prüfungszeiten (Tag der Prüfung) keine Störvariablen darstellen, habe ich bei der Untersuchung des Diktats zwei signifikante Stichproben durchgeführt. Generell fanden beide Stichprobenerhebungen (Prüfungen) am Ende des Vorbereitungslehrganges zur Unteroffiziersausbildung statt. Angesichts der Tatsache, dass es sich bei beiden Stichproben um eine Wiederholungsprüfung handelte, kannten die Versuchsteilnehmer das Prüfungsformat Diktat bereits.

In Hinblick auf die *Untersuchungsdurchführung* erfolgten alle drei Stichproben auf Basis eines einheitlichen Ablaufkonzeptes:

- Absolvierung des schriftlichen Teils der Zulassungsprüfung unter realen Prüfungsbedingungen
- Ausfüllen des Fragebogens
- Teilnehmende Beobachtung durch den Forscher

Sowohl für die Durchführung des Diktats als auch die Ausarbeitung der Bildbeschreibung (Vergleichsstudie) beträgt die Testbearbeitungszeit etwa 25 Minuten. Für das Ausfüllen des Fragebogens hatten die Versuchsteilnehmer 15 Minuten Zeit. Die Datenerfassung und die Einordnung der Stichproben in den Forschungszusammenhang erfolgten durch:

- computergenerierte Einzel- und Gesamtauswertung der Testergebnisse (heeresinterne Prüfungsdatenbank)
- computergenerierte Auswertung der Fragebögen
- eigene Mitschriften bei Beobachtungen

## **11.1 Erhebung 1 in Ebelsberg**

### Ort/Stärke - Auswahl der gültigen Testpersonen – Gründe

Die erste Erhebung fand am *20.12.2010* um 0900 Uhr in der Kaserne Linz-Ebelsberg in Oberösterreich statt. Beim Untersuchungsort handelte es sich um einen ausgelagerten

Ausbildungsort der Heeresunteroffiziersakademie, der auch für Prüfungszwecke verwendet wird.

Die Gesamtstichprobe umfasste 23 Versuchspersonen, wobei eine Person davon ausgeschlossen wurde. Der Grund dafür war ein mangelhaft ausgefüllter Fragebogen in Verbindung mit einer unterdurchschnittlichen Ausfüllzeit (unter 5 Minuten). Dabei handelt es sich nicht um Einzelfragen, sondern um ganze Seiten des Fragebogens, die nicht ausgefüllt wurden. Es verblieben daher *22 gültige Versuchsteilnehmer*.

### Testvariante

Bei dem Diktattext „*Wenn Schnee zur Gefahr wird*“ handelt es sich um einen authentischen Text aus der Tageszeitung „Kurier“ vom 24.02.2009.

Der Text gliedert sich in zehn Sätze mit unterschiedlicher Länge und umfasst 163 Wörter.

### Alter

Die vorliegende Stichprobe setzt sich aus folgenden Altersgruppen zusammen:

Altersgruppe	Häufigkeit	Prozent (gerundet)
unter 20-Jährige	1	4,5
20 – 25-Jährige	17	77,4
25 – 30-Jährige	2	13,6
über 30-Jährige	2	4,5
Gesamt	22	100,00

Laut Prüfungsdatenbank entspricht das Durchschnittsalter 25 Jahren. Die Altersverteilung ist breit gestreut, sie reicht von 19 bis zu 34 Jahren.

Angesichts der Altersobergrenze bei Zulassungsprüfungen von 35 Jahren kann das Alter von 34 Jahren durchaus als Grenzwert betrachtet werden.

Bezugnehmend auf die Orthographiekenntnisse ist hier besonders zu beachten, dass drei Personen (älter als 30 Jahre) im Zuge ihrer schulischen Ausbildung die neuen Regeln der Rechtsschreibreform von 2006 mit ziemlicher Sicherheit noch nicht gelernt haben. Diese Annahme kann erfahrungsgemäß auch auf die Personengruppe der 25 - 30-Jährigen zutreffen.

### Geschlecht

Bei Zulassungsprüfungen innerhalb des ÖBH – es geht um den Beruf Soldat – muss von einer ungleichen Verteilung der Geschlechter ausgegangen werden.

An der ersten Stichprobe nahm lediglich eine weibliche Testperson teil.

### Muttersprachen – Deutsch als wievielte Sprache

Bei den erworbenen bzw. erlernten Sprachen der Testpersonen zeigen sich folgende Werte:

Sprachen	Häufigkeit	Prozent (gerundet)
Deutsch als Muttersprache	21	95,5
Deutsch als Zweitsprache	1	4,5
BKS	1	
Gesamt	22	100

Unter den Versuchsteilnehmern befindet sich eine Person mit nichtdeutscher Muttersprache. Es handelt sich dabei um einen Mann mit Herkunftsland Bosnien.

### Erlernte Fremdsprachen

Fremdsprachen	Häufigkeit	Prozent (gerundet)
keine Fremdsprache	-	-
eine Fremdsprache Englisch	18	82,0
zwei Fremdsprachen- Englisch	3	13,5
+ Arabisch	1	
+ Italienisch	1	
+ Spanisch	1	
vier Fremdsprachen Englisch	1	4,5
+ Italienisch, Französisch, Russisch	1	
Gesamt	22	100

Hinsichtlich der erlernten Fremdsprachen ist eine breite Streuung auszumachen. Innerhalb der Stichprobengruppe sind neben Deutsch und Bosnisch sechs erlernte Fremdsprachen vertreten. Bemerkenswert ist der Aspekt, dass drei Nichtmaturanten bereits in der Pflichtschulzeit neben Englisch auch eine zweite Fremdsprache erlernen konnten. Wie die Schulabschlussdaten zeigen, handelt es dabei nicht um Personen mit Hauptschulabschluss bzw. um Personen mit sonstigen Schulabschlüssen wie z. B. Handelsschule.

In der Tabelle ist nicht zu übersehen, dass Englisch nach wie vor als erste gelernte Fremdsprache dominiert, auch bei jener Person mit nichtdeutscher Muttersprache.

Was beim Fragebogen fehlte, war die Frage nach der Reihenfolge der gelernten Sprachen, insbesondere, ob Englisch vor Deutsch gelernt wurde oder Deutsch als zweite erlernte Fremdsprache in Frage kommt.

### Schulabschluss

Laut eigener Angaben der Versuchspersonen ergaben sich folgende Werte:

positiver Schulabschluss	Häufigkeit	Prozent (gerundet)
Hauptschulabschluss	11	50,0
Unterstufe Gymnasium	1	4,5
Matura	2	9,0
andere Schulabschlüsse	8	36,5
Gesamt	22	100

Das Kriterium Schulabschluss zeigt, dass die Stichprobe sehr unterschiedliche Schulabschlüsse ausweist und dass zwei Personen mit Matura an der Stichprobe teilnahmen.

## 11.2 Erhebung 2 in Enns

### Ort/Stärke - Auswahl der gültigen Testpersonen – Gründe

Die zweite Erhebung fand am 07.03.2011 um 0900 Uhr an der Heeresunteroffiziersakademie in Enns/Oberösterreich statt. Beim Untersuchungsort handelte es sich um jenen Ort, an dem der allgemeine Teil der Unteroffiziersausbildung und die Stabsunteroffiziersausbildung stattfinden.

Bei der Teilnehmergruppe handelt es sich um einen reduzierten Teilnehmerkreis aus der Stichprobe 1. Dieser Umstand ist darauf zurückzuführen, dass alle Teilnehmer der Stichprobe 2 die Stichprobe 1 negativ, also mit Förderbedarf absolviert haben.

Die Gesamtstichprobe umfasste 28 Versuchspersonen, wobei 4 davon ausgeschlossen wurden. Gründe dafür waren ausschließlich mangelhaft ausgefüllte Fragebögen in Verbindung mit einer unterdurchschnittlichen Ausfüllzeit (unter 5 Minuten). Dabei handelte es sich, wie schon in Ebelsberg, nicht um Einzelfragen, sondern um ganze Seiten der Fragebögen, die nicht ausgefüllt wurden. Nach der Bereinigung verblieben bei der zweiten Stichprobe 24 gültige Versuchsteilnehmer.

### Testvariante

Beim Diktattext handelt es sich um die gleiche Testvariante wie bei der ersten Stichprobe.

### Alter

Die zweite vorliegende Stichprobe setzt sich aus folgenden Altersgruppen zusammen:

Altersgruppe	Häufigkeit	Prozent (gerundet)
unter 20-Jährige	-	-
20 – 25-Jährige	20	83,0
25 – 30-Jährige	2	8,5
über 30-Jährige	2	8,5
Gesamt	24	100

Das Durchschnittsalter entspricht mit 24 Jahren der angestrebten Zielgruppe. 83% aller Personen sind im Alter von 20 bis 25 Jahren, was etwa dem Durchschnitt aller Geprüften des Jahres 2010 entspricht

Laut Prüfungsdatenbank ist die älteste Person 34 Jahre alt, die jüngsten jeweils 20 Jahre. Es sind keine Personen aus der Altersgruppe „unter 20-Jährige“ vertreten.

Gegenüber der ersten Stichprobe haben etwa gleich viel 25 bis 30-Jährige an der Zulassungsprüfung teilgenommen.

### Geschlecht

In der zweiten Stichprobe findet sich eine weibliche Testperson mit bosnischer Herkunft.

Das entspricht etwa dem weiblichen Anteil der ersten Stichprobe.

### Muttersprachen – Deutsch als wievielte Sprache

Bei den erworbenen bzw. erlernten Sprachen der Testpersonen zeigen sich folgende Werte:

Sprachen	Häufigkeit	Prozent (gerundet)
Deutsch als Muttersprache	21	87,5
Deutsch als Zweitsprache	3	12,5
BKS	3	
Gesamt	24	100

Der Anteil der Personen mit nichtdeutscher Muttersprache ist prozentuell etwa dreimal so hoch wie bei der ersten Stichprobe. Bei der Länderverteilung ist wie schon bei der ersten Stichprobe Bosnien vertreten.

### erlernte Fremdsprachen

Fremdsprachen	Anzahl	Prozent
keine Fremdsprache	2	8,0
eine Fremdsprache Englisch	18	75,0
zwei Fremdsprachen- Englisch	4	17,0
+ Französisch	1	
+ Italienisch	3	
Gesamt	24	100

Nichts Unerwartetes zeigt das Ergebnis in der Zeile „keine Fremdsprache“. Es geht um zwei ältere Männer mit bosnischen und serbischen Wurzeln, die nach eigenen Angaben erst nach dem Pflichtschulabschluss nach Österreich ausgewandert sind.

Dieser Umstand lässt die Vermutung zu, dass beide Personen in ihrer Heimat keine schulische Ausbildung in Englisch erhielten. Wie schon wie bei der ersten Stichprobe zeigt sich, dass Englisch nach wie vor als erste gelernte Fremdsprache dominiert.

Vergleichbar, wenngleich wesentlich höher als bei der ersten Stichprobe ist die Anzahl jener Personen mit Italienisch als zweiter Fremdsprache.

### Schulabschluss

Laut Angaben der Versuchspersonen ergaben sich dabei folgende Werte:

positiver Schulabschluss	Häufigkeit	Prozent (gerundet)
Hauptschulabschluss	12	50,0
Unterstufe Gymnasium	-	-
Matura	5	21,0
andere Schulabschlüsse	7	29,0
Gesamt	24	100

Das Kriterium Schulabschluss zeigt, dass die zweite Stichprobe ganz im Gegensatz zur ersten aus einer ziemlich breitgestreuten Personengruppe bestand und dass im Verhältnis zu anderen Zulassungsprüfungen eine hohe Zahl an MaturantInnen an der Stichprobe teilnahm. Was beim Kriterium Schulabschluss weiters auffällt, ist, dass keine Person mit Unterstufe Gymnasium aufscheint.

## **12 Darstellung der Ergebnisse**

### Präsentationsdesign der Ergebnisse

Die grafische Darstellung und Auswertung erfolgt anhand von Tabellen, die auf Basis von sprachinstitutsinternen Excel-Datenbanken erstellt wurden.

Während der Wert „Häufigkeit“ Aussagen über die jeweilige Anzahl der Versuchsteilnehmer liefert, gibt der aus der Gesamtsumme von Stichprobenteilnehmern errechnete „Mittelwert“ Auskunft über das durchschnittliche Leistungsniveau.

### **12.1 Ergebnisse der Stichproben im Vergleich**

Wenn man zwei Stichproben für die Untersuchung eines Sprachtests heranzieht, ist neben den beiden Fragebogenergebnissen auch ein direkter Vergleich der Prüfungsergebnisse hinsichtlich der Bewertungskategorien „kein Förderbedarf“ (bestanden) und „Förderbedarf“ (nicht bestanden) von empirischem Interesse. Abgesehen von den unterschiedlich strukturierten Personengruppen sind Einflüsse sowohl durch den Prüfungsort als auch durch die unterschiedliche Prüfungssituation (Was passierte am Vortag? usw.) zu erwarten. Wie bereits erwähnt, wurde bei beiden Stichproben die gleiche Testvariante verwendet. Bei den jeweils 163 Wörtern durften maximal 7% der Wörter falsch sein.

	<b>Ebelsberg 20122010</b>		<b>Enns 07032011</b>		Abweichung Mittelwerte
	Häufigkeit	Mittelwert	Häufigkeit	Mittelwert	
kein Förderbedarf	9	3,81 %	11	4,91 %	1,1 %
Förderbedarf	13	9,53 %	13	13,60 %	4,07 %
Insgesamt	22	7,19 %	24	9,44 %	2,25 %

Beim Vergleich der Testergebnisse zeigen sich sowohl bei der Kategorie „kein Förderbedarf“ als auch bei der Kategorie „Förderbedarf“ deutliche Abweichungen bei den Mittelwerten.

Aus den Daten ist die Tendenz ablesbar, dass bei der Personengruppe mit Förderbedarf große Leistungsunterschiede vorherrschen.

Ähnliche Differenzen zeigen sich auch bei der Häufigkeit, allerdings im geringeren Ausmaß.

Im Summenergebnis der Mittelwerte liegen beide Stichproben über der Förderschwelle von 7%. Bei der zweiten Stichprobe kann aufgrund der hohen Prozentanzahl bei der Kategorie Förderbedarf auf insgesamt schwächere Testteilnehmer geschlossen werden als bei der ersten Stichprobe.

#### Ergebnisse der Stichproben im Vergleich: Spitzen- und Migrantengruppe

Wenn es um sprachliche Korrektheit wie beim Diktat geht, sind die Teilergebnisse der Spitzen- und Migrantengruppe von besonderem Interesse.

Bei der Spitzengruppe habe ich aufgrund der annähernd gleichen Anzahl der Testteilnehmer jeweils die besten fünf angenommen, bei der Migrantengruppe alle Teilnehmer.

	<b>Ebelsberg: 20122010</b>		<b>Enns: 07032011</b>		Abweichung
	Häufigkeit	Mittelwert	Häufigkeit	Mittelwert	
Spitzengruppe	5	2,82 %	5	2,45 %	0,37 %
Migrantengruppe	1	14,11 %	3	8,86 %	5,25 %

Bei der Spitzengruppe gab es jeweils eine Testperson mit 0 % (fehlerlos). Die Werte der Spitzengruppen gleichen sich annähernd, obwohl die zweite Stichprobe drei Maturanten mehr umfasste. Aus den Daten der Spitzengruppe ist weiters zu schließen, dass das Diktat für leistungsstarke Personen leicht zu bewältigen war.

Bei der zweiten Stichprobe in Enns gab es einen Teilnehmer der Migrantengruppe, der mit 6,13 % die Prüfung knapp bestand und dadurch das Gesamtniveau der Gruppe (etwa 10%) entscheidend verbesserte.

Um zu zeigen, dass die Eingangsvoraussetzungen in einem unmittelbaren Zusammenhang zu den Testergebnissen stehen können, habe ich einige untersuchungsrelevante Aspekte ausgewählt.

### Alter und Testergebnisse

Aus Platzgründen habe ich bei den Tabellen *kein Förderbedarf durch + ersetzt, bzw. Förderbedarf durch – ersetzt.*

	Ebelsberg: 20122010			Enns: 07032011		
	Gesamt	+	–	Gesamt	+	–
unter 20	1	-	1	-	-	-
20 - 25	17	5	12	20	10	10
26 - 30	3	2	1	2	-	2
über 30	2	2	-	2	2	-
Insgesamt	22	9	13	24	12	12

Hinsichtlich der Fragestellung, ob ältere Personen bei Zulassungsprüfungen besser oder schlechter abschneiden, zeigt sich eine eindeutige Datenlage. Sowohl bei der ersten Stichprobe als auch bei der zweiten haben alle Testteilnehmer über 30 Jahre die Prüfung bestanden.

### Muttersprachen und Testergebnisse – Herkunftsabhängigkeit

*Besonderheiten abhängig vom Herkunftsland*

	Ebelsberg: 20122010			Enns: 07032011		
	Gesamt	+	–	Gesamt	+	–
Deutsch als Muttersprache	21	9	12	21	11	10
Zweitsprache	1	-	1	3	1	2
Insgesamt	22	9	13	24	12	12

Wie schon im Testvergleich der Migrantengruppe ersichtlich, bestätigt sich die Annahme, dass Personen mit nichtdeutscher Muttersprache anteilmäßig beim Diktat schlechter abschneiden.

### Fremdsprachen und Testergebnisse

	Ebelsberg: 20122010			Enns: 07032011		
	Gesamt	+	–	Gesamt	+	–
keine FS	-	-	-	2	-	2
eine FS	18	5	13	18	7	11
zwei und mehr FS	4	4	-	4	4	-
Insgesamt	22	9	13	24	11	13

Ein eindeutiges Bild zeigt sich beim Abschneiden von Personen mit mehr als einer erlernten Fremdsprache. Alle Teilnehmer haben den Test erfolgreich absolviert. Dagegen scheinen

Personen ohne Fremdsprachenausbildung zumindest bei der Stichprobe 2 nur geringe Chancen bei Sprachtests dieser Art zu haben.

### Bildungsniveau/Schulabschluss und Testergebnisse

Wie wissenschaftlich fundierte Testevaluierungen belegen, hat insbesondere der Bildungshintergrund zumeist große Auswirkungen auf Erfolg und Misserfolg bei schriftlichen Testverfahren.

	Ebelsberg: 20122010			Enns: 07032011		
	Gesamt	+	-	Gesamt	+	-
Hauptschule	11	1	10	12	1	11
Gymnasium Unterstufe	1	1	-	5	4	1
Matura	2	2	-	7	6	1
anderer Abschluss	8	5	3	-	-	-
Insgesamt	22	9	13	24	11	13

Hauptschüler liefern zumeist negative Ergebnisse, während der Großteil der Maturanten gut abschneidet. Bei der Kategorie „anderer Schulabschluss“ zeigen sich unterschiedliche Ergebnisse. Hier würde eine genauere Ausdifferenzierung der Schultypen zu einem aussagekräftigeren Ergebnis führen.

### Testergebnisse und Selbsteinschätzung (Fragebogen)

Da es nur wenige Ergebnisse bzw. schulische Belege über das tatsächliche Leistungsniveau der Versuchsteilnehmer in den geforderten Fertigkeiten *Grammatik (G)* und *Orthographie (O)* gibt, möchte ich auch die subjektive Einschätzung der Versuchsteilnehmer als Vergleichskategorie in Betracht ziehen. Dabei greife ich in der Chronologie meiner Untersuchung bewusst etwas vor, indem ich einen Teilbereich der Fragebogenergebnisse herausgreife und in Relation zu den Testergebnissen stelle. Berücksichtigt dabei werden muss, dass bei Selbsteinschätzungen von Sprachkenntnissen erfahrungsgemäß die Bescheidenheit des Teilnehmer eine nicht unwesentliche Rolle spielen kann.

	Ebelsberg: 20122010			Enns: 07032011		
	Häufigkeit		Förderung	Häufigkeit		Förderung
Korrektheit	G	O		G	O	
sehr gut	2	1	9+	3	2	11+
gut	5	5		8	8	
mittel	15	10	11-	11	12	13-
eher schlecht	-	6		2	2	
	22	22		24	24	

Unter der Voraussetzung, dass die Beantwortung der Fragen anonym bleibt, haben die Befragten offensichtlich sehr realistisch ihre Grammatik- und Orthographiekenntnisse beurteilt. Das bedeutet, dass die Selbsteinschätzung fast deckungsgleich übereinstimmt mit den erreichten Testergebnissen. Mitgedacht werden sollte, dass alle Teilnehmer zumindest eine schriftliche Sprachprüfung beim ÖBH absolviert und daraus entsprechende Schlüsse gezogen haben.

## **12.2 Teilnehmende Beobachtung**

Neben den Testergebnissen, der Fragebogenerhebung und den Experteninterviews soll die Beobachtung als vierte methodische Komponente in meiner Untersuchung begriffen werden. Die Beobachtungsergebnisse sollen vor allem als Ergänzung zu den anderen Verfahren verstanden werden. Insbesondere soll die Beobachtung ein Gesamtbild der Testsituation und Aufschlüsse über mögliche Einflüsse auf die Ergebnisse beim Diktat liefern. Um dabei möglichst objektiv zu bleiben, ersuchte ich jeweils eine zweite Person, mich bzw. meinen Partner in Stichprobe 2 bei der Beobachtung zu unterstützen. Angesichts dieser nahezu idealen Forschungsbedingungen hatten ich und meine Partner genügend Zeit, Beobachtungen zu notieren. Zusätzlich bestand auch die Möglichkeit, Teilnehmer nach dem Test zu befragen.

### Allgemeines

Angesichts der Fülle an beobachtbaren Kategorien galt es für mich, vorerst eine Auswahl zu treffen. Im Vordergrund standen dabei typische Phänomene, die auf das Ergebnis beim Diktat Einfluss ausüben können.

### Mögliche Einflussgrößen auf die Ergebnisse beim Diktat

- Testangst
- Akustik
- Geschwindigkeit, Pausen, Umblättern
- Schreibroutine - Ermüdung
- Schreibgeschwindigkeit der Personen mit nichtdeutscher Muttersprache
- Schreibschrift versus Schulschrift
- Psychologische Faktoren wie Testwiederholung
- Abschreiben vom Nachbarn

- Strategie
- Wissen über das Testformat (Bekanntheit)

Auf Basis der aufgelisteten Einflussgrößen gilt es, die beobachteten Sachverhalte zu klassifizieren und zu beschreiben. Bei der Darstellung der Beobachtungsergebnisse ziehe ich beide Stichproben (1 und 2) heran. Die Beobachtungsergebnisse beziehen sich sowohl auf meine Beobachtungen als auch die Beobachtungen meines Forschungspartners.

### *Testangst*

Bei etwa einem Drittel aller Teilnehmer zeigte sich, dass beim Niederschreiben des Diktattextes, vor allem am Anfang, die Schreibhand zittert. Dieses Phänomen verschwand nach drei bis fünf Sätzen, je nach Teilnehmer.

### *Akustik*

Bei beiden Stichproben konnte ich keine Auffälligkeiten hinsichtlich der Akustik wahrnehmen. Allerdings klagten einige Teilnehmer nach der Stichprobe, dass sie einige Wörter nicht richtig verstünden. Nach Überprüfung der Sitzordnung ergab sich, dass die betroffenen Personen unterschiedliche Plätze in den Lehrsälen einnahmen. Probleme mit der Raumakustik können daher nicht ausgeschlossen werden.

### *Geschwindigkeit des Ansagens (Diktierens), Pausen, Umblätter*

Obwohl der Diktattext dreimal abgespielt wurde, habe ich festgestellt, dass ca. 30% aller Teilnehmer dem Tempo des Ansagers, vor allem bei längeren Sätzen, nicht folgen konnten. Darüber hinaus konnte ich beobachten, dass sowohl die Pausen nach den Satzteilen als auch die Pause zum Umblättern (nach fünf Sätzen muss umgeblättert werden!) zu kurz waren. Angesichts dieser Beobachtungen bestätigen sich *Solmeckes Forschungsergebnisse*<sup>118</sup> auch beim Diktat, zumindest bei der Raumakustik und der Geschwindigkeit des Ansagens.

### *Schreibroutine – Ermüdung beim Schreiben*

In unmittelbarem Zusammenhang mit der Schreibgeschwindigkeit steht die Schreibroutine.

---

<sup>118</sup> vgl. dazu Solmecke, Gert (2000): „Faktoren der Schwierigkeit von Hörtests“. In: Bolton, Sybylle [Hrsg.]: TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar. Köln: VUB Gilde, S. 63.

Bereits nach fünf Sätzen begann etwa ein Viertel aller Teilnehmer die Hände auszuschütteln. Nach acht Sätzen war es fast die Hälfte aller Personen. Beim Schreiben der zweiten Diktathälfte (Sätze 5 bis 10) war deutlich zu erkennen, dass einige Teilnehmer, trotz Händeschütteln, langsamer wurden und Wörter auslassen mussten. Es kam vermehrt auch vor, dass Wörter, die im Diktattext öfters vorkamen, am Anfang orthographisch richtig und am Ende falsch geschrieben wurden.

#### *Schreibgeschwindigkeit der Personen mit nichtdeutscher Muttersprache*

Bei drei von vier Personen dieser Gruppe zeigte sich bereits am Anfang, dass die Sprechgeschwindigkeit der Ansage zu schnell war. Alle drei Teilnehmer konnten bereits den zweiten bzw. dritten Satz nicht vollständig niederschreiben. Darüber hinaus war zu sehen, dass die Wort- und Satzlücken gegen Ende des Diktats deutlich zunahmen.

#### *Schreibschrift versus Schulschrift*

Bei ca. 80% aller Personen nahm ich wahr, dass das Diktat mittels Schreibrift (Druckschrift) niedergeschrieben wurde. Nur etwa 20 % der Teilnehmer verwendeten Schulschrift. Erschwerend wirkte sich der Umstand beim Schreibtempo aus.

Auf die Frage, warum das Diktat in Druckschrift und nicht in Schulschrift geschrieben wurde, antworteten die betroffenen Teilnehmer höchst unterschiedlich:

- „Wenn ich in Latein schreibe, kann man meine Schrift nicht lesen“.
- „Ich schreibe seit Jahren nur mehr in Schreibrift.“
- „Ich habe geglaubt, dass Schreibrift verlangt wurde!“

#### *Psychologische Faktoren wie Testwiederholung*

Bei Personen, die den Test wiederholen mussten, zeigten sich vermehrt Anzeichen von Angst. Etwa zwei Personen pro Stichprobe resignierten bei etwa der Hälfte des Diktats und schrieben daraufhin nicht mehr alles mit. Dass sich Testangst unmittelbar auf die sprachliche Leistung auswirkt, zeigt sich hier ganz deutlich.

#### *Abschreiben vom Nachbarn*

Bei der dritten Abspielphase versuchten einige die Gelegenheit zu nützen von ihrem Nachbarn abzuschreiben, was zum Teil auch gelang. Interessanterweise handelte es sich dabei größtenteils um jene Personen, die ohnehin die Zulassungsprüfung geschafft hätten.

### *Wissen über das Testformat (Bekanntheit)*

Trotz ausreichendem Wissen über das Testformat und auch über die Texte gab es diesbezüglich wenig Auffälligkeiten. Allerdings beobachtete ich, dass einige Teilnehmer den Diktattext offensichtlich schon kannten. Sie konnten bereits während der ersten Abspielphase (normales Sprechtempo) einige Sätze vollständig mitschreiben.

### **12.3 Ergebnisse aus der Fragebogenerhebung**

So aufschlussreich die Vergleichsdaten der Testergebnisse auch sind, sie lassen nur vage Aussagen über die Nützlichkeit, insbesondere über die Validität des Testverfahrens zu. Da bei der Fragebogenerhebung die Einschätzung und Beurteilung des Testformates im Vordergrund steht, sind hier aussagekräftigere Daten zu erwarten. Dieser Annahme zur Folge stellt für mich die Fragebogenerhebung den zentralen Teil meiner empirischen Untersuchung dar. Zusätzlich sollen die Daten Auskunft über jene untersuchungsrelevanten Phänomene geben, die in keinem Zusammenhang mit den Testergebnissen stehen (Einschätzungen), aber wesentliche Aspekte zur Untersuchung liefern. Darüber hinaus kann überprüft werden, inwieweit ein Konsens über die Nützlichkeit des Testverfahrens besteht.

Mitgedacht bei der Fragebogenerhebung werden sollte, dass bei derartigen Befragungsmethoden nicht alle ermittelten Daten relevante und eindeutige Ergebnisse liefern können.

Im Gegensatz zu anderen Evaluationsinstrumenten steht bei der von mir angewandten Fragemethode die Gewichtung von Aussagen im Mittelpunkt. Der befragte Teilnehmer soll mittels einer Ratingskala feststellen, ob Aussagen zutreffen oder nicht. Um Missverständnisse und Unklarheiten möglichst zu vermeiden, stand bei allen Stichproben eine Auskunftsperson für etwaige Rückfragen zur Verfügung.

Bei den zu bewertenden Aussagen handelt es sich um untersuchungsrelevante Feststellungen, die in folgende Themenkomplexe chronologisch eingeordnet sind:

- A) Testvorbereitung
- B) Testdesign und Testlänge
- C) Niveau, Qualität und Konsistenz der Testaufgaben
- D) Störende Aspekte/Testangst
- E) Lern- und Gebrauchsstrategien
- F) Sprachvermittlung beim Bundesheer

G) Testaktivität als nützliche Ausbildungszeit

H) Persönlicher Eindruck - Fragen zur Person

Die graphische Darstellung der Daten erfolgt mittels Balkendiagrammen auf Basis einer Datenmatrix-Tabelle, die sowohl die befragten Personen als auch die Variablen (z. B. 1 - 4) repräsentiert. Bei der Darstellung habe ich mich auf statistisch auffällige Ergebnisse und Zusammenhänge beschränkt, die untersuchungsrelevante Schlussfolgerungen zulassen.

Zudem möchte ich auch das Antwortverhalten einzelner Personengruppen in Betracht ziehen um in einem Vergleich zu zeigen, dass beispielsweise Migranten eine andere Meinung zu einzelnen Themenkomplexen haben als Personen mit deutscher Muttersprache.

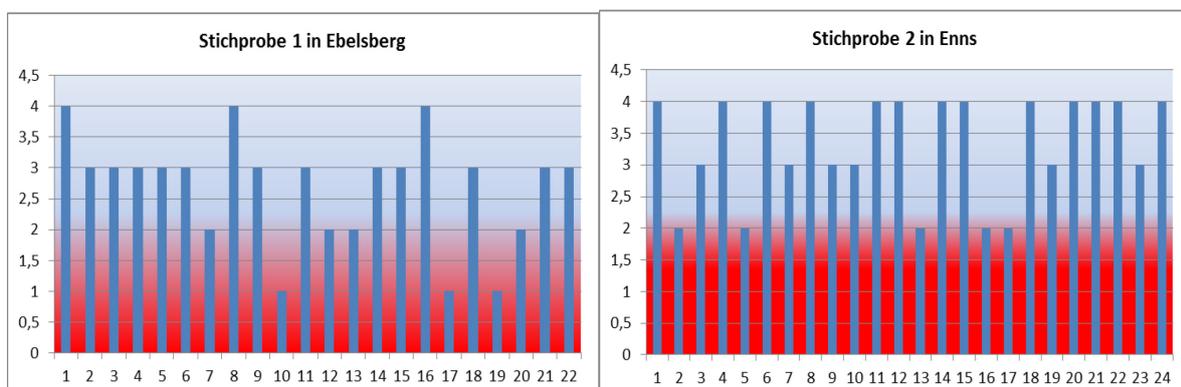
Bei den Auswertungen handelt es um deskriptiv-statistische Analysen, die anhand der Ratingskala präsentiert werden.

trifft zu	trifft überwiegend zu	trifft eher nicht zu	trifft nicht zu
1	2	3	4

Wie schon bei der Gegenüberstellung der Testergebnisse werden beide Stichproben synchron dargestellt. Die rote Hintergrundfarbe gibt jenen Bereich an, wo aufgrund der Feststellungen der Befragten negative Auswirkungen auf die Qualität des Testformats zu erwarten sind

### B) Testdesign und Testlänge:

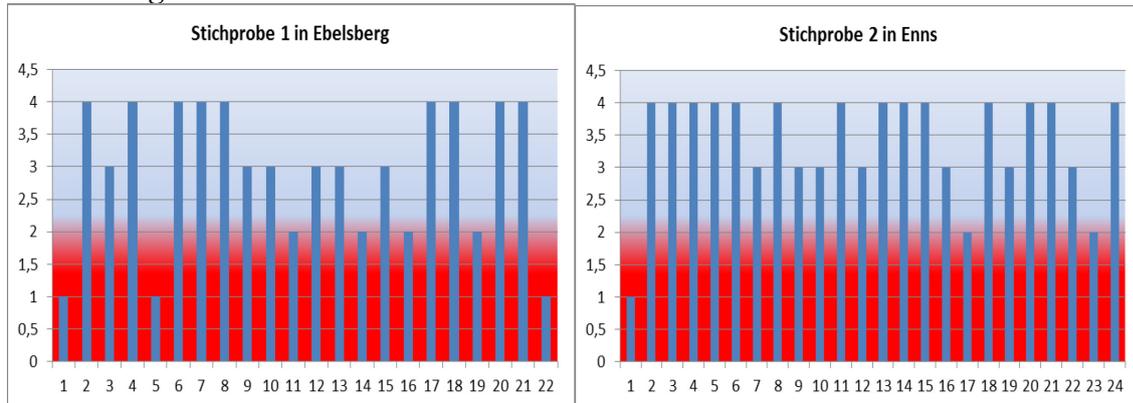
*Die äußere Form des Tests löst bei mir Missfallen aus:*



Wie die Ergebnisse der beiden Stichproben zeigen, löst das Testformat großteils kein Missfallen aus. Besonders auffällig ist, dass 19 von 24 Teilnehmern der Stichprobe 2 der Meinung sind, dass die äußere Form des Diktats kein Missfallen auslöst.

#### D) Störende Aspekte/Testangst:

*Durch die Aufgabenstellung werden Personen mit nichtdeutscher Muttersprache benachteiligt:*

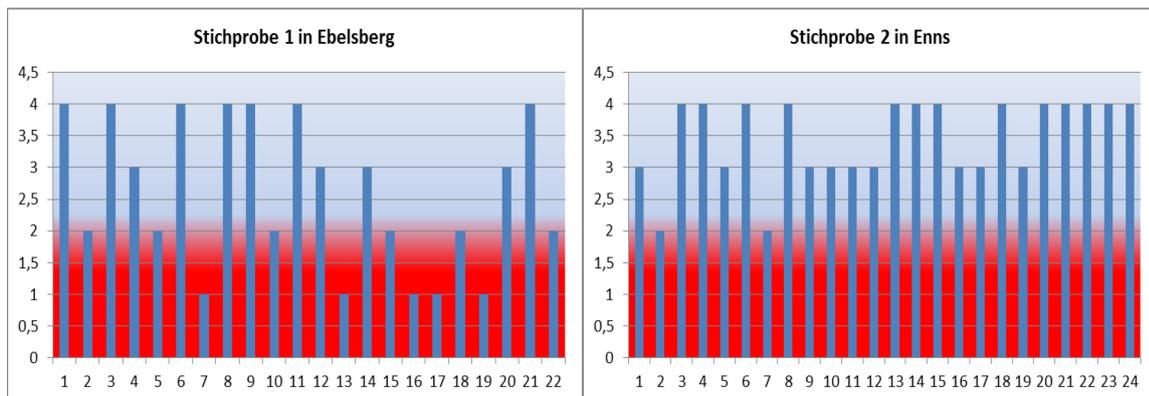


Angesichts der Tatsache, dass Menschen mit nichtdeutscher Muttersprache beim Diktat zumeist schlechter abschneiden als Muttersprachler, gilt diesem Aspekt mein besonderes Interesse. Auffällig ist, dass etwa 70 % der Befragten in der Art und Weise der Aufgabenstellung keine Benachteiligung feststellen können.

#### Differenzierte Auswertung bei den einzelnen Personengruppen:

Auch bei der Gruppe der Migranten zeigen sich ähnliche Ergebnisse. Interessanterweise fühlen sich 75% der Teilnehmer mit nichtdeutscher Muttersprache (4) durch die Aufgabenstellung nicht benachteiligt. Noch eindeutiger fällt die Beurteilung bei jenen Personen aus, die mehr als eine Fremdsprache gelernt haben. 90% aller Personen (10) stimmen der Feststellung nicht zu, dass Nichtmuttersprachler benachteiligt werden.

*Die Zeitknappheit, die Testsituation und die Inhalte des Tests lösen bei mir Angst aus:*



Anders als bei der Themenstellung zuvor, können hier kaum Korrelationen zwischen den beiden Stichproben festgestellt werden. Während bei Stichprobe 1 die Hälfte der Teilnehmer

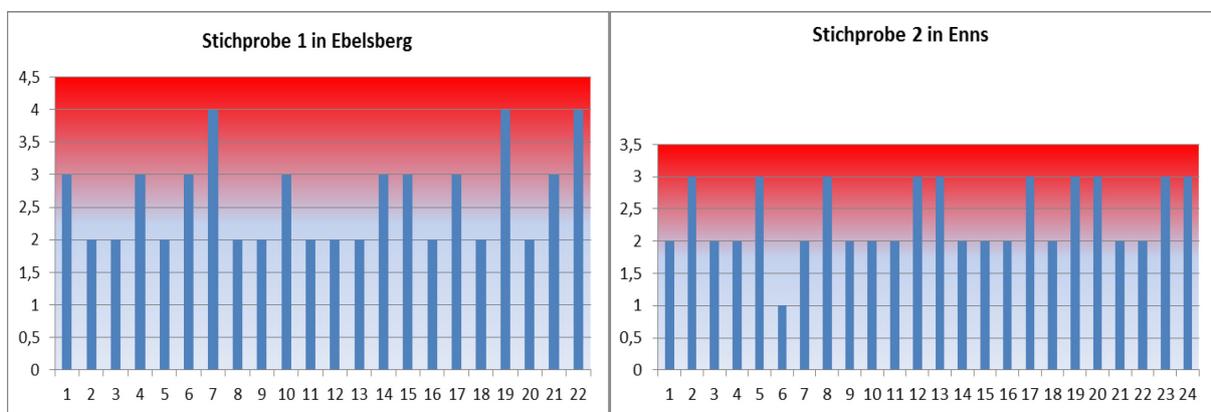
angeben, dass die Zeitknappheit, die Testsituation und auch die Inhalte des Tests Angst auslösen, sind es bei Stichprobe 2 lediglich 2 Teilnehmer.

*Differenzierte Auswertung bei den Personengruppen:*

Bei der Auswertung der Migrantengruppe zeigt sich ein etwas anderes Bild als bei der Themenstellung zuvor. Zwei von vier Migranten stimmen der Aussage zu, dass die Zeitknappheit, die Testsituation und die Inhalte des Tests Angst auslösen.

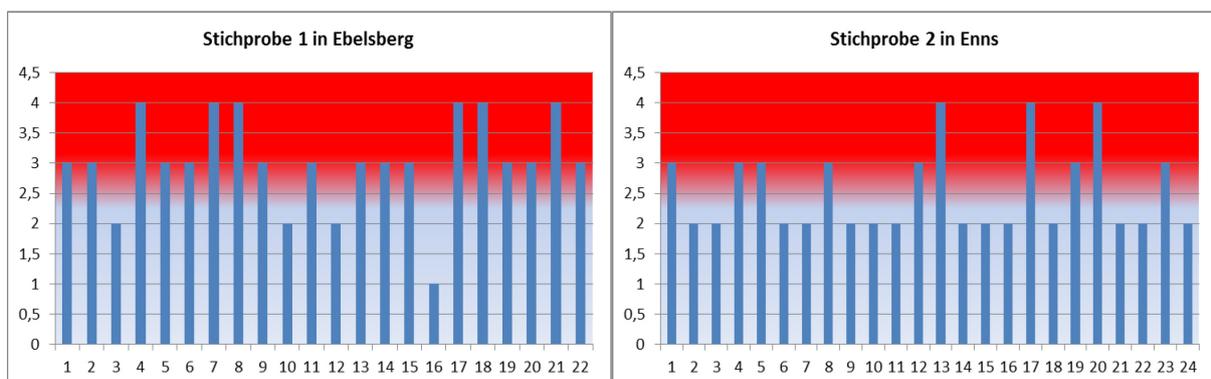
#### F) Sprachverwendung beim Bundesheer:

*Auf spezifische Sprachverwendungssituationen im Bundesheer (streitkräfterelevante Fachsprache) wird im Test eingegangen:*



Eine der wesentlichsten Themenkomplexe stellt die Berufsbezogenheit des Testformats dar. Nach der Einschätzung von 11 bzw. 10 Befragten wird auf spezifische Sprachverwendungssituationen im Bundesheer im Test eingegangen. Einige Befragte, besonders in Stichprobe 2, stimmen dieser Aussage eher nicht zu. Nur drei Teilnehmer insgesamt schließen einen Zusammenhang zwischen dem Diktat und einer berufsspezifischen Schreibtätigkeit völlig aus.

*Das erfolgreiche Lösen der Testaufgabe gibt mir die Gewissheit, den sprachlichen beruflichen Erwartungen zu entsprechen:*

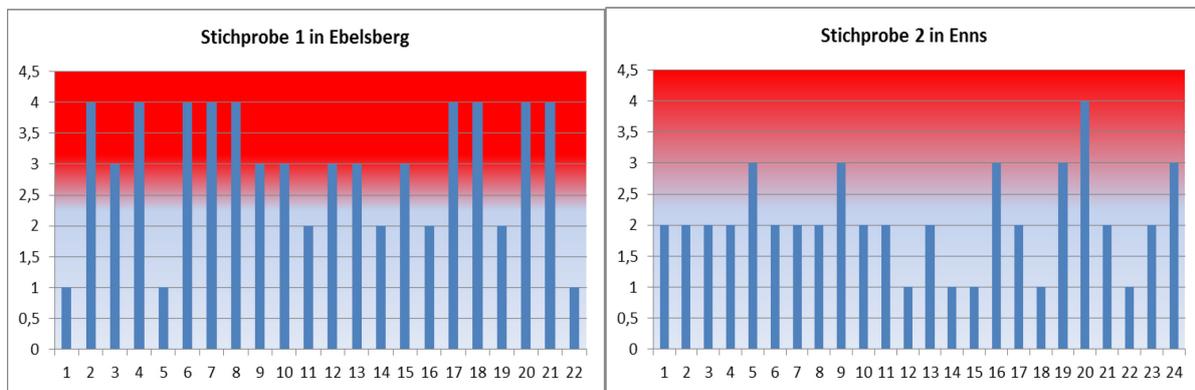


Wie schon beim Thema zuvor geht es hier um den beruflichen Kontext. Dabei zeigt sich insbesondere in Stichprobe 1, ein deutlich anderes Einschätzungsverhalten als bei „Sprachverwendung beim Bundesheer“. Für 18 Befragte gibt es keine Gewissheit, dass die Testaufgabe den sprachlichen Erwartungen entspricht.

Von großem Interesse jedes Testerstellers ist, ob anhand der Testergebnisse zukünftige Leistungen voraussagbar sind. Wenn, wie in Stichprobe 1, die Mehrzahl der Teilnehmer feststellt, dass das Lösen der Testaufgabe keine Gewissheit gibt, den zu erwartenden sprachlichen Leistungen zu entsprechen, dann ist gemäß Nützlichkeitskriterien die Validität des Testformats nicht ausreichend gegeben (Vorhersagevalidität).

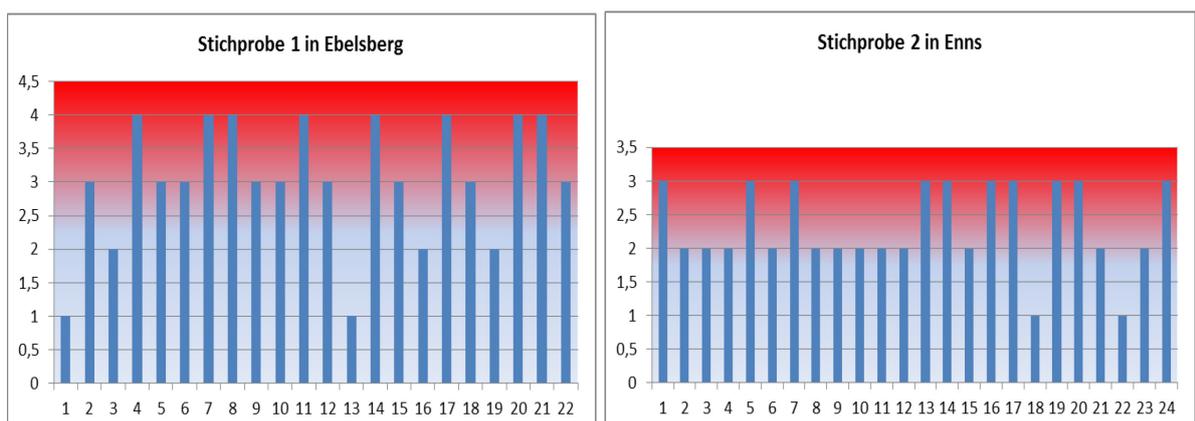
### G) Testaktivität als nützliche Ausbildungszeit:

*Der Test stellt einen sinnvollen Teil der militärischen Ausbildung dar:*



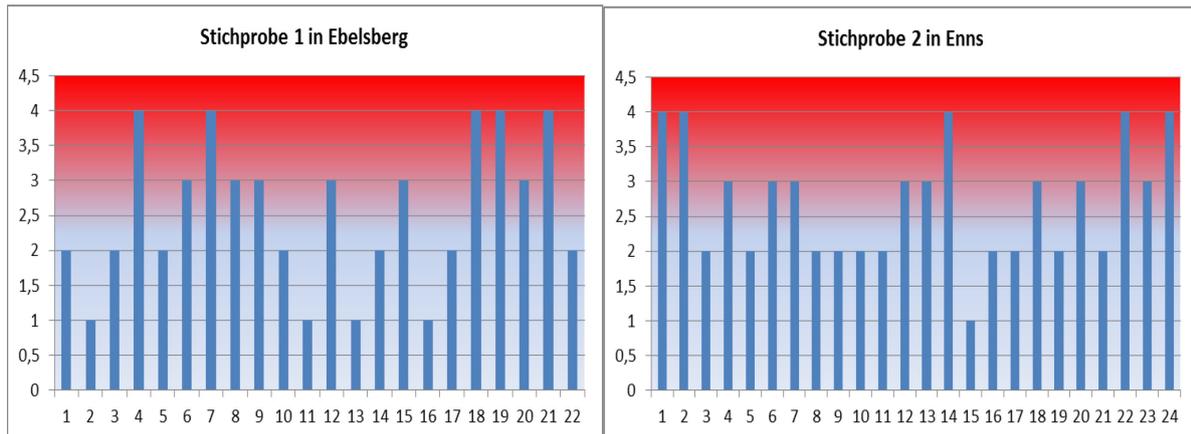
Bei den Ergebnissen der Stichprobe 1 zeigt sich eine deutliche Abweichung zu jenen der Stichprobe 2. Während in Ebelsberg nur sieben Teilnehmer feststellen, dass das Diktat einen sinnvollen Teil der Ausbildung darstellt, gilt für 18 TeilnehmerInnen der Stichprobe 2 das Gegenteil. Wie schon in den vorangegangenen Themenkreisen korreliert das Ergebnis von Ebelsberg zumeist nicht mit dem von Enns.

*Die Testaufgabe ist ansprechend und interessant:*



Bei der Feststellung, ob das Format Diktat ansprechend und interessant ist, zeigt sich eine geringe Korrelation zwischen den beiden Stichproben. Während bei der Stichprobe 1 17 Teilnehmer das Diktat als nicht ansprechend beurteilen, überwiegt bei der Stichprobe 2 ein positives Urteil.

*Beim Lösen der Testaufgabe habe ich Mängel bei meinen sprachlichen Fähigkeiten erkannt:*



Eines der vorrangigen Kriterien einer Testaufgabe sollte es sein, den Sprach- und Lernstand der Teilnehmer abzufragen. Dabei sollten den Teilnehmer ihre Stärken und Schwächen bewusst gemacht werden. Wie in den Diagrammen ersichtlich, korreliert Stichprobe 1 eindeutig mit Stichprobe 2. 50% aller Teilnehmer stellen fest, dass sie durch das Diktat keine sprachlichen Mängel erkennen konnten, 50% meinen das Gegenteil. Das Ergebnis kann daher als ziemlich ausgeglichen gesehen werden.

#### Differenzierte Auswertung bei den Personengruppen

Vom besonderen Interesse ist bei dieser Themenstellung die Gruppe mit Hauptschulabschluss. Das sind zumeist jene Personen, die von Fördermaßnahmen profitieren. Bei der Detailauswertung zeigt sich, dass etwa 90% aller Hauptschüler Mängel bei ihren sprachlichen Fähigkeiten erkennen konnten

#### H) Persönlicher Eindruck:

*15. Welche sprachlichen Teilfertigkeiten werden Ihrer Einschätzung nach beim Diktat geprüft?*

Fertigkeit	Stichprobe 1		Stichprobe 2	
	ja	nein	ja	nein
Hörverstehen	19	3	22	2
Rechtschreibfähigkeit	22	-	20	4
Grammatik	18	4	20	4
Schreibfähigkeit	20	2	17	7
Konzentrationsfähigkeit	21	1	20	4

Etwa 80% der Befragten wissen, welche Teilfertigkeiten beim Diktat geprüft werden.

Was bei dieser geschlossenen Fragestellung fehlt, sind Kategorien wie kognitive Strategien oder Welt- und Sachwissen, die erfahrungsgemäß einen wesentlichen Einfluss auf die Testleistung ausüben können.

#### 12.4 Ergebnisse aus den Experteninterviews

Bei der Auswahl der Interviewpartner handelt es sich um bundesheerinterne Experten, die den Untersuchungsgegenstand genau kennen und die an der Entwicklung der Zulassungsprüfung Deutsch maßgeblich beteiligt waren. Darüber hinaus sind es Personen, die über eine langjährige Praxis in Testen und Prüfen verfügen.

Gleichermaßen wie in der Fragebogenerhebung habe ich auch beim Experteninterview ausschließlich inhaltlich wichtige forschungsrelevante und aussagekräftige Interviewausschnitte aus den Rohdaten ausgewählt, nicht forschungsrelevante Textteile wurden ausgeschieden. Danach wurden die ausgewählten Textausschnitte verglichen, gebündelt und entsprechenden Themenkomplexen zugeordnet, wobei bedeutungsgleiche Informationen zusammengefasst wurden.

Bei der Transkription handelt es sich um eine text- und wortgetreue Wiedergabe der Interviewausschnitte. Paraphrasierungen werden ausschließlich zu Erklärungen und Erläuterungen von Ergebnissen eingesetzt. Bei längeren Aussagen werden Sätze, die in keinem unmittelbaren Zusammenhang mit dem Forschungsgegenstand stehen, nicht angeführt. Bei Antworten, die sich bei beiden Experten inhaltlich decken, habe ich mich auf die Aussage von Experte 1 beschränkt. Die vorgenommenen Kürzungen wurden möglichst behutsam als Beleg genommen, um die Inhalte der Aussagen nicht zu verfälschen.

Um das ausgewählte Textmaterial anschaulicher zu präsentieren, werden die Experteninterviews synchron dargestellt (direkt miteinander verglichen).

Nach jedem Fragenkomplex sollen anhand einer Zusammenschau die Sichtweisen der Experten kurz erläutert werden und in die abschließenden Schlussfolgerungen einfließen.

Meine Interviewpartner waren: <sup>119</sup>

Hauptmann Mag. Walter Rys – *Referatsleiter Deutsch am Sprachinstitut*

Akademische Fachrichtung: Deutsche Philologie

Laut eigenen Angaben: Ausbildung als Deutschprüfer für Sprachprüfungen nach STANAG 6001 und GER

Seit 1999 ist Walter Rys für die Entwicklung und Überarbeitung von Prüfmitteln Deutsch verantwortlich.

Das Interview fand am 20.04.2011 um 13:45 Uhr im Bereich des Sprachinstituts statt.

Es dauerte etwa 1 Stunde und 40 Minuten.

Oberst des höheren militärischen Dienst Mag. Dr. Wolfgang Zecha – *Referatsleiter Prüfungswesen am Sprachinstitut*

Akademische Fachrichtung: Pädagogik/Geschichtstudium

Laut eigenen Angaben: Ausbildung als Deutschprüfer für Sprachprüfungen nach STANAG 6001 und GER

Seit 2001 ist Wolfgang Zecha für die Planung und Durchführung aller Sprachprüfungen des Sprachinstituts letztverantwortlich.

Darüber hinaus ist er bei der Entwicklung, Überarbeitung und Bearbeitung von Prüfmitteln für die Konformität mit STANAG 6001 und Konkordanz mit den Prozessabläufen zuständig.

Das Interview fand am 12.04.2011, um 13:00 Uhr im Bereich des Sprachinstituts statt.

Es dauerte etwa zwei Stunden.

*MilFü2* = Lehrgang militärische Führung 2 an der Heeresunteroffiziersakademie

Bei den Interviewpartnern werden aus Einfachheitsgründen bei den ausgewählten Gesprächssequenzen die Bezeichnungen Experte 1 (Wolfgang Zecha) und Experte 2 (Walter Rys) verwendet.

## **Teil 1:**

### **Sprachspezifische Fundierung des Testverfahrens, Auswahl der geprüften Kompetenzen**

- 1.1 Gab es vor der Einführung des aktuellen Testformats für die schriftliche Zulassungsprüfung Deutsch eine andere/frühere Testversion? Wenn ja, welche? Welche Gründe gab es, die alten Testversionen zu verwerfen?

---

<sup>119</sup> Die Betroffenen waren mit der Namensnennung einverstanden

### Experte 1:

*Seit der Mitte der 1980-iger-Jahre wurde ein Lückendiktattest als Zulassungsprüfung für den Unteroffizierslehrgang (Vorgänger des MilFü 2) an der Heeresunteroffiziersakademie verwendet. Dieser wurde durch den damaligen Heerespsychologischen Dienst ausgewählt und mittels Erlass verfügt. 2006 kamen der Referatsleiter Deutsch als Fremdsprache, Mag. RYS, und ich zur Erkenntnis, dass dieses Prüfmittel nicht mehr zeitgemäß bzw. weder valide noch reliabel ist.*

*Außerdem war der Bekanntheitsgrad der Prüfmittel schon relativ hoch. Es wurde daher mit dem Heerespersonalamt Verbindung aufgenommen, ob es in deren Sinne sei, ein neues Prüfmittel zu entwickeln. Die Antwort war positiv und es wurde auf Antrag des SIB mit Erlass BMLV eine Arbeitsgemeinschaft installiert, welche die Erarbeitung von validen Prüfmitteln Deutsch zur Zielsetzung hatte. Zeitgleich wurde vom BMLV eine Arbeitsgemeinschaft beauftragt, einen Anforderungskatalog für Friedens- und Einsatzfunktionen zu erarbeiten. In dieser Arbeitsgemeinschaft war ich der Vertreter des SIB (Der vorliegende Anforderungskatalog enthält auch die Anforderungen in sprachlicher Hinsicht). Zeitgleich wurde das SIB beauftragt, einen Anforderungskatalog Sprachwesen zu erarbeiten, der unter meiner Federführung erstellt wurde.*

Die Antwort des Experten 2 entspricht in den wesentlichen inhaltlichen Punkten der Antwort von Experten 1.

Demnach gibt es laut Aussage der Experten seit 2006 bereits Überlegungen, das alte Testformat „Lückendiktattest“ aufgrund Nichterfüllens der Qualitätskriterien (Reliabilität und Validität) und militärsprachlichen Anforderungen durch ein neues zu ersetzen. Angesichts dieser Evaluierungstendenzen wurde die Arbeitsgruppe Prüfmittel installiert, um die Entwicklung neuer Prüfmittel voranzutreiben.

Was bleibt ist die Frage, warum das aktuelle Diktat valider bzw. reliabler als ein Lückendiktat sein soll, wenn es doch bei beiden Aufgaben um Hörverstehen, Lexik, Grammatik und Orthographie geht. Ein Grund dafür könnte sein, dass 2007, also noch vor dem Diktat, eine freie Schreibaufgabe angestrebt wurde und dadurch hohe Qualitätsansprüche beansprucht wurden.

- 1.2 Aus welchen sprachtheoretisch fundierten Überlegungen heraus wurde das aktuelle Testformat entwickelt?  
*Wie erfolgte die Entscheidungsfindung zu Gunsten dieses Testformats?*

### Experte 1:

*Durch das SIB wurde zuerst ein MC-Einstufungstest mit einem zusätzlichen kurzen offenen Formatteil im freien Schreiben (10 Sätze zu Hobbys) erarbeitet und an einer Stichprobe im Jänner 2007 pilotiert. Nach weiteren Validierungen und Überarbeitungen in Kooperation mit dem Heerespersonalamt wurde dieser Test nach Schulung der Auswerter mit Mai 2007 beim Heerespersonalamt eingeführt. Hierbei kamen zwei Überlegungen zum Tragen:*

- 1. möglichst standardisierte Auswertungsmöglichkeit*
- 2. Möglichkeit einer Sprachstandserhebung zur Festlegung von Förderbedarf*

*Das Niveau sollte etwa dem Pflichtschulniveau der 8. Schulstufe entsprechen. (HS-Abschluss) Nach einer Zeitspanne von ca. 3 Monaten wurden vor allem die Ergebnisse der Auswertung beim Schreiben einer Evaluierung unterzogen. Hierbei zeigten sich Unterschiede bei der Bewertung des Freien Schreibens. Zeitgleich wurde begonnen an einem Zulassungsprüfmittel zu arbeiten, welches vor allem die Lese- und Schreibkompetenz valide prüfen sollte und mit Jänner 2008 wurden Förderseminare Deutsch im Umfang von 30 Unterrichtseinheiten zuerst auf „freiwilliger“ Basis, ab Dezember 2008 verpflichtend für alle förderwürdigen Unteroffiziersanwärter eingeführt.*

#### Experte 2:

*Pragmatische Überlegung: Staatsbürgerschaftstest entspricht A2 GER<sup>120</sup> Soldat in Führungsfunktion sollte eine Stufe darüber stehen, also B1 bzw. Zertifikat Deutsch. Anfangs hatten wir als Schreibaufgabe eine Bildgeschichte, später das Diktat.*

Die Antworten der Experten können durchaus als chronologische Fortsetzung zur vorangegangenen Fragestellung verstanden werden. Experte 1 gibt zu verstehen, dass bereits 2007 der Versuch unternommen wurde, eine freie Schreibaufgabe als Testformat einzusetzen. Wie aus dem Text weiter hervorgeht, scheiterte dieses Pilotprojekt jedoch aus Gründen der Objektivität und Reliabilität. Auffällig ist hier die unterschiedliche Auffassung der beiden Experten hinsichtlich des Anforderungsniveaus. Während Experte 1 das Pflichtschulniveau der 8. Schulstufe als Ausgangsbasis für die Zulassungsprüfung heranzieht, geht Experte 2 vom Referenzniveau B1 des GER aus. Diese Aussage lässt zumindest vermuten, dass bereits 2007 Menschen mit nichtdeutscher Muttersprache als zukünftigen Bedarfsträger erkannt wurden.

Besonders interessant ist, dass Experte 1 sowohl von Sprachstandserhebungen als auch von gezielten Fördermaßnahmen spricht während Experte 2 auch die zwischenzeitliche Einführung einer Bildgeschichte als Schreibaufgabe erwähnt.

### 1.3 Nach welchen sprachspezifischen Kriterien wurde bei der Auswahl und Formulierung der Testaufgabe vorgegangen?

#### Experte 1:

*Aus den ersten Testergebnissen (Pilotierung des Heerespersonalamt-Tests mit freiem Schreiben) ließ sich ableiten, dass der Förderbedarf bzw. die Defizite in erster Linie in den Bereichen Orthografie und Grammatik lagen. Anfänglich waren zwei Aufgaben mit „Freiem Schreiben“), genaugenommen zwei Bildgeschichten mit ca. 80 Worten als Prüfung vorgesehen.*

*2009 wurde schließlich als Kompromiss das Diktat anstatt des Freien Schreibens zur Überprüfung der orthografischen Fähigkeiten ausgewählt und zur Überprüfung von grammatikalischen Aspekten ein Lückentext entwickelt und eingesetzt.*

---

<sup>120</sup> Gemeinsamer Europäischer Referenzrahmen: Stand 2010

### Experte 2:

*Militärisch relevante Texte in der Hinsicht, dass es sich um Texte handelt, die entweder über das Militär berichten, bzw. von militäreigenen Institutionen stammen. In jedem Fall handelt es sich meist um authentische Texte!*

Beachtet man die Aussagen zu den Fragen 1.1. und 1.2, so lässt sich feststellen, dass seitens der beiden Experten größere Anstrengungen unternommen wurden, ein freies Schreibformat durchzusetzen. Gemäß der Aussage von Experte 1 dient das Diktat lediglich als Kompromiss, nicht als Ideallösung. Gleichzeitig führt Experte 1 an, dass es in erster Linie Defizite in den Bereichen Orthographie und Grammatik gab, womit man die Auswahl des Diktats rechtfertigen könnte.

Nach Sicht von Experten 1 war das Diktat zum damaligen Zeitpunkt vor allem zur Überprüfung von orthographischen Fähigkeiten gedacht, während ein Lückentext die grammatikalischen Aspekte abdecken sollte.

Experte 2 betont die Wichtigkeit der militärischen und authentischen Ausrichtung von Prüfungstexten. Das bedeutet, dass auch die Textvorlagen für das Diktat hinsichtlich militärischer Relevanz überprüft werden.

#### 1.4 Aus welchem Grund wurde das Diktat als Prüfmittel herangezogen?

### Experte 1:

*Das Diktat ergab sich aus der Forderung, gehörte Texte zu verschriftlichen (Befehlsmitschriften, Niederschriften, Funksprüche) bzw. aus der Forderung, möglichst schnell eine voraussichtliches Mengengerüst von bis zu 400 Texten innerhalb von 3 Tagen zu korrigieren. Um das Prüfungsformat Gesprächsnotiz zu akzeptieren, hätte man bei dieser Menge eine große Anzahl an akademischem Personal benötigt!  
Das Diktat wurde daher lediglich als Kompromiss aus Prüfungsmengengerüsten und der Notwendigkeit korrekte Schreibergebnisse zu überprüfen, ausgewählt.*

### Experte 2:

*Die Entscheidung für das Diktat fiel durch Beschluss der Arbeitsgruppe Prüfmittel gegen meine vorgebrachten Einwände.*

Eine der Kernfragen des Experteninterviews beschäftigt sich mit den Prüfmittel Diktat an sich. Es geht dabei in erster Linie um die Begründung des Testkonstrukts, welches hinter dem Diktat steht.

Experte 1 geht davon aus, dass das Niederschreiben von Gehörtem zu den wesentlichsten kommunikativen Aufgaben eines Berufssoldaten zählt. Die Einführung des Diktats begründet der Experte dahingehend, dass aufgrund der hohen Probandenanzahl die Korrektur von Gesprächsnotizen nur mit großen zeitlichen und personellem Aufwand zu bewältigen wäre.

Angesichts dieser Begründung ist die Vorgangsweise der Arbeitsgruppe Prüfmittel durchaus nachvollziehbar. Ob das Diktat unter diesen Bedingungen eine Verbesserung gegenüber der alten Testversion, dem Lückendiktat darstellt, ist allerdings fraglich. Die Argumentation von Experte 1 lässt außerdem den Schluss zu, dass das Testformat Diktat nach wie vor einen hohen Status genießt.

Experte 2 betont hingegen, dass die Entscheidung für das Diktat ausschließlich aus ökonomischen Gründen fiel und nicht aus testtheoretischer Sicht.

- 1.6 In welchem Ausmaß wurden die beruflich sprachlich gebundene Handlungsfähigkeit und die streitkräfterelevante Fachsprache bei der Entwicklung der Testaufgabe mitberücksichtigt?  
*Wird in der Testaufgabe auf sprachliche Kompetenzen zur Befähigung und Bewältigung von berufsspezifischen Arbeitsprozessen überhaupt eingegangen?*

Experte 1:

*Die Texte für das Diktat wurden ausschließlich allgemeinsprachlich gehalten, weil die Diktate sowohl bei der Einstufungsprüfung als auch bei der Zulassungsprüfung eingesetzt werden sollten.*

Experte 2:

*Insoweit als bei der Erstellung stets auf Mitwirkung von Praktikern (vor allem erfahrene Unteroffiziere) Wert gelegt wurde, die von Anfang an ersucht wurden, Situationen und Textsorten zu berücksichtigen, die im beruflichen Alltag relevant sind. Inhalte wurden im Laufe mehrerer Jahre so festgelegt.*

Ein wesentlicher Aspekt bei der Untersuchung der schriftlichen Zulassungsprüfung ist die Art und Weise der Einbeziehung von erwartbaren berufsspezifischen Kompetenzen in die Testaufgabe. Interessanterweise treten gerade bei dieser Fragestellung unterschiedliche Blickwinkel der Experten zu Tage. Während Experte 1 das Diktat in erster Linie als allgemeinsprachlich gehaltenes Prüfmittel definiert, stehen für den Experten 2 beruflich relevante Kommunikationssituation und Textsorten im Vordergrund.

- 1.7 Inwieweit wurden die Folgen des Tests bei der Erstellung des Testformats in Betracht gezogen?

Experte 1:

*Es war uns bewusst, dass eine gewisse Benachteiligung von Personen mit Migrationshintergrund entstehen könnte, da diese in der Population aber sehr gering waren, wurde dies in Kauf genommen.*

*Wie sich in weiterer Folge herausstellte, „verschwanden“ diese während der militärischen Ausbildung und wurden nicht unmittelbar durch die Deutschprüfung ausgesiebt. Hier wären aber in weiterer Folge Forschungen anzusetzen und eine Grundsatzentscheidung durch BMLVS zu treffen, falls ein höherer Anteil an Migranten im UO-Kader der Zukunft gewünscht wäre.*

Experte 2:

*Insoweit als die Einführung einer Fördersystematik vorgesehen war. Test war nicht vordergründig als Selektionsmittel vorgesehen.*

Wenn man, wie in der Frage zuvor, von einer berufsspezifischen Ausrichtung eines Testformats wie dem Diktat spricht, so sollten die Argumente hinsichtlich der Testfolgen nicht außer Acht gelassen werden.

Experte 1 führt an, dass die Benachteiligung von Menschen mit nichtdeutscher Muttersprache durchaus wahrgenommen, allerdings wegen einer zu geringen Anzahl in Kauf genommen wurde. Bezug nehmend auf das Bestehen der Zulassungsprüfung führt Experte 1 aus, dass die meisten Migranten bereits vor der Prüfung aus unterschiedlichen Gründen aus dem Militärdienst ausscheiden. Experte 2 hingegen betont, dass es bei der Prüfung in erster Linie nicht um Selektion, sondern um Förderung geht.

Demzufolge kann gesagt werden, dass das Diktat nicht auf ein Ausscheiden von Soldaten abzielt, sondern ausschließlich zu Fördermaßnahmen führen sollte.

## **Teil 2: Güte- und Qualitätskriterien des Tests (Validität, Reliabilität)**

2.2 Passt das Niveau der Aufgabenstellung zum Niveau der Probanden?

Experte 1:

*Im Großen und Ganzen dürften diese übereinstimmen, da auch die Korrelation der Testergebnisse zwischen Einstufungsprüfung und Zulassungsprüfung dies zeigen.*

Experte 2:

*Weitgehend offensichtlich ja, da bis zum Ende Vorbereitung MilFü2 Niveau vom Großteil der Bewerber erreicht wird.*

Wie die Testergebnisse belegen, bestehen nur etwa 30 – 50% aller Pflichtschulabgänger das Diktat. Immerhin zeigt sich nach Ende der Fördermaßnahmen, dass etwa 90% die Zulassungsprüfung positiv bestehen. Ob allerdings das Niveau der Aufgabenstellung dem der Probanden und die 7% als Förderschwelle den realen beruflichen Bedingungen entsprechen, ist allerdings fraglich. Man darf dabei auch nicht übersehen, dass die Zulassungsprüfung, wie Experte 1 formuliert, lediglich eine **Mindestanforderung** darstellen soll.

2.5 Werden Testteilnehmer aufgrund des Aufgabentyps benachteiligt?  
(Teilnehmer mit Migrationshintergrund usw.)

#### Experte 1:

*Es war uns bewusst, dass Personen mit Migrationshintergrund wahrscheinlich durch das Diktat benachteiligt werden könnten. Wegen der relativ geringen Anzahl und mangels eines besseren Verfahrens gingen wir dieses Manko ein. Stressfaktor war nicht Absicht (Unter Stress richtig zu schreiben wird nicht benötigt!)*

#### Experte 2:

*Nach derzeitigem Wissensstand liegt diese Vermutung nahe. Das Diktat ist ein komplexes Prüfungsformat, das unterschiedliche Fertigkeiten auf eine spezielle Weise überprüft. Das heißt, dass sowohl beim Hörverstehen als auch bei der schriftlichen Übertragung (wortwörtlichen Übertragung) es zu Benachteiligung von Menschen mit Migrationshintergrund kommen kann.*

Bei dieser Fragestellung habe ich nochmals das Thema „Diskriminierung“ explizit aufgegriffen, um den Interviewpartnern die Möglichkeit zu geben, ausführlicher zu antworten. Die Argumentationen der beiden Experten zeigen deutlich, dass diese Problematik durchaus ernst genommen wird. Was nicht in Betracht gezogen wird, sind Alternativformen beim Diktat wie z. B., dass man für Personen mit nichtdeutscher Muttersprache den Diktattext statt dreimal viermal abspielt (zeitliche Konsequenz!).

- 2.11 Kann das Testergebnis als Indikator für voraussichtliche sprachliche Leistungen dienen (Vorhersagevalidität)?  
*Wie gut sagt das Testergebnis künftige Leistungen im Bereich der im Testkonstrukt verankerten Fertigkeiten voraus?*

#### Experte 1:

*Dies konnte mangels noch nicht durchgeführter Transferevaluierung nicht beantwortet werden. Hierzu wäre eine Befragung der Vorgesetzten von Absolventen der weiterführenden Unteroffizierskurse (MilFü) von 2009 und 2010 durchzuführen.*

#### Experte 2:

*Diktat eher nein.*

Hinsichtlich dieser Fragestellung gibt es offenbar noch keine aussagekräftigen Ergebnisse seitens der Heeresunteroffiziersakademie. Daten liegen nur insofern vor, dass immer wieder Texte von Unteroffizieren im Zuge von Ausbildungsaktivitäten verfasst werden. Allerdings werden solche Texte meist computerunterstützt, also mit Einschaltung des Rechtschreibprogramms produziert.

- 2.15 Lassen sich aus den Testergebnissen auch präzise Förderoptionen ableiten?

#### Experte 1:

*Grundsätzlich nicht.*

Experte 2:

*Die ZulP diene bis jetzt eher zur Erhebung eines Ist-Zustandes der Sprachbeherrschung insgesamt. (Mindestleistung!) Diktat kann Schwächen in bestimmten Teilbereichen der Orthographie aufzeigen (Klein- u. Großschreibung, s-Schreibung).*

Nicht unerwartet gehen beide Experten davon aus, dass das Diktat keine präzisen Förderoptionen zulässt. In Betracht ziehen sollte man dabei, dass es zahlreiche Maßnahmen im Bereich des ÖBH gibt, mit denen Sprachförderungen im großen Ausmaß durchgeführt werden. Es drängt sich daher die Frage auf, ob nicht mit präziseren Förderoptionen eine wesentlich höhere Durchkommensrate bei der Zulassungsprüfung bzw. ein geringerer Förderbedarf erreicht werden könnte?

2.18 Ist eine Objektivität der Auswertung gegeben?

*Ist das Testergebnis unabhängig vom Prüfer/Beurteiler?*

Experte 1:

*Diese ist relativ hoch; bei Unklarheiten wird durch akademisches Personal eine möglichst objektive Entscheidung, wenn nötig kommissionell, getroffen.*

Experte 2:

*Grundsätzlich ja. Es wird das Diktat von ausgebildeten Pädagogen ausgewertet.*

Laut Aussagen beider Experten scheint bei der Beurteilung des Diktats eine objektive Leistungsfeststellung gewährleistet zu sein.

2.21 Wurden beim Test Lern- und Gebrauchsstrategien mitberücksichtigt, die den Lernenden Vorteile bei der Lösung der Testaufgabe bringen?

Experte 1:

*Nein, meines Wissens nach nicht.*

Experte 2:

*„Teaching to the test“ im Rahmen von Förderseminaren bringt dem Lernenden Vorteile.*

Bei der inhaltlichen Ausrichtung dieser Fragestellung geht es vor allem um Rückkopplungseffekte von Sprachprüfungen, die sich insbesondere auf die Didaktik und Methodik bei Förderseminaren auswirken. Wie Experte 2 formuliert, scheint es tatsächlich so zu sein, dass bei den Förderseminaren im Bereich „Schriftlicher Gebrauch“ fast ausschließlich jene sprachlichen Phänomene trainiert werden, die beim Diktat vorkommen.

### **Teil 3: Nützlichkeiten des Tests (Wirkung, Authentizität, Interaktivität, Praktikabilität)**

- 3.1 Stellt der Test eine sinnvolle Lernaktivität dar und sind positive Auswirkungen auf das Lernen und die Motivation der Probanden zu erwarten?  
*Welche Auswirkungen könnte der Test auf das zukünftige Lernverhalten haben?*

Experte 1:

*Eine Lernaktivität ist mit dem Test nicht beabsichtigt; es handelt sich ausschließlich um eine Zulassungsprüfung. Indirekte Effekte der Lernmotivation bei Nichtbestehen sind aber nicht ausgeschlossen.*

Experte 2:

*Diktat scheint eher demotivierend zu sein.*

Laut den Aussagen beider Experten ist eine Lernaktivität der Teilnehmer nicht beabsichtigt. Interessant ist hier, dass auch beide Experten indirekte Effekte bei Nichtbestehen des Diktats nicht ausschließen können. Wenn man die Beurteilungen der Fragebogenteilnehmer zum Thema „Testangst“ heranzieht, bestätigt sich die Hypothese von Experte 2.

- 3.6 Besteht beim Testverfahren ein sinnvolles Verhältnis zum Aufwand (Personal, Material, Räume, Unterlagen, Zeit)?  
*Kann man beim vorliegenden Testverfahren von einer ökonomisch angemessenen Vorgangsweise sprechen?*

Experte 1:

*Ja.*

Experte 2:

*Testformat ist weitgehend mit Papier und Schreibgerät zu bewältigen, das Diktat erfordert zusätzlich ein CD-Spielgerät. Der Bedarf an Testpersonal ist ebenfalls nicht zu hoch. Die Ökonomie liegt daher eher im oberen Bereich.*

Abgesehen von der akademischen Ausbildung und Schulung des Auswertungspersonals kann man laut Aussage der Experten das Diktat durchaus als ökonomisches Verfahren bezeichnen.

### **Teil. 4: Transparenz**

- 4.1 Wissen die Lerner, was im Test verlangt wird und können sie sich entsprechend darauf vorbereiten?

Experte 1:

*Das Diktat ist zumeist bereits vor der Zulassungsüberprüfung bekannt.*

Experte 2:

*Die Teilnehmer beim Förderseminar ja, sonst nur bei aktiver Informationsbeschaffung.*

Beide Experten betätigen, dass das Testformat Diktat bereits vor der Zulassungsprüfung bekannt ist. Ein Grund dafür ist, dass bereits bei der Einstufungsüberprüfung das Diktat als Prüfmittel eingesetzt wird. Unabhängig davon besteht bei Prüfungsformaten wie dem Diktat die Gefahr, dass die Texte mittels Mobiltelefon im Zuge der Prüfung photographiert und ins Internet gestellt werden. Dass dies unbemerkt passieren kann, soll hier nicht weiter ausgeführt werden.

### **13 Interpretation der Ergebnisse**

#### Angemessenheit der Interpretation

Die angemessene Interpretation von Ergebnissen kann als zentraler Bestandteil in der empirischen Forschung begriffen werden. Dabei sollen die Interpretationen sich auf die Forschungsfrage beziehen und sich an möglichst methodisch gesicherten Untersuchungsergebnissen orientieren. Bei der Vorgangsweise gilt es, anhand von relevanten inhaltlichen Aspekten die Ergebnisse zu formulieren und daran Schlussfolgerungen zu knüpfen.<sup>121</sup>

#### **13.1 Interpretation der Testergebnisse**

Ein erster Schwerpunkt der Interpretation gilt den Testergebnissen der Stichproben 1 und 2, woraus sich einige forschungsrelevante Schlussfolgerungen ableiten lassen.

Im Einzelnen lassen sich in der Reihenfolge der im Kapitel: „Darstellung der Testergebnisse“ behandelten Aspekte folgende Aussagen festhalten:

Sieht man von der allgemeinen Problematik des Ansagens ab, zeigt sich, dass das Diktat vor allem bei schwächeren TeilnehmerInnen stark differenziert. Die Ursachen dafür liegen meiner Ansicht nach zum einen im nicht ausreichend vorhanden Welt- und Sachwissen und zum anderen in der Schwierigkeit des Diktattextes.

Bestätigt werden kann diese Aussage dadurch, dass ältere Teilnehmer trotz unterrichtsbedingter Rechtsschreibdefizite (Rechtsschreibreform 2006!) aufgrund ihrer Erfahrung mit Medien (Zeitungen, Bücher) und ihrer Prüfungsroutine zumeist besser abschneiden als jüngere. Weiters zeigt sich, dass Hauptschüler generell geringe Chancen beim

---

<sup>121</sup> vgl. dazu Stigler (2005), S. 215

Diktat haben. Das betrifft auch Personen ohne Fremdsprachenausbildung. Im Gegenzug stellt das Diktat für leistungsstarke Lerner kein großes Problem dar.

Die Daten bestätigen auch die Meinung von Bildungsexperten, dass Menschen mit hohem Bildungshintergrund zumeist bessere Chancen bei Sprachtests dieser Art haben.

Allerdings zeigt sich auch, dass auch Maturanten durchaus schlechter abschneiden können. Wenn auch Migranten insgesamt anteilmäßig schlechter abschneiden, machen solche mit Hauptschulabschluss kaum mehr Fehler als Muttersprachler mit schwachen Testleistungen. Dieser Umstand deutet darauf hin, dass bilinguale Unterrichtsformen auf das normgerechte Schreiben wie beim Diktat bereits positive Auswirkungen zeigen. Die Lücke zwischen Pflichtschülern und Personen mit nichtdeutscher Muttersprache scheint sich allmählich zu schließen.

Insgesamt kann aus den Ergebnissen geschlossen werden, dass die Prüfung für Pflichtschulabsolventen zu schwer war und daher zu hoch angesetzt scheint. Das belegt auch das Summenergebnis der Mittelwerte beider Stichproben, das über der Förderschwelle von 7% liegt. Darüber hinaus ist eine deutliche Korrelation zwischen den Testergebnissen und den sozialen Parametern wie Bildung und Alter nicht zu übersehen. Eine große Rolle spielt dabei die Art und Weise der Prüfung, wo nicht Kompetenz, sondern Wissen (Schulwissen, Weltwissen) den Erfolg ausmacht.

Bei den Selbsteinschätzungen von Sprachkenntnissen fällt auf, dass ein Großteil der Befragten zwar über ihre Stärken und Schwächen, aber nicht über die Lernmethoden und Lernstrategien Bescheid wusste. Insofern stellt sich die Frage, inwieweit Grammatik- und Orthographiewissen im Erwachsenenalter überhaupt erlernbar ist.

Wichtige Parameter für das Lernen von Erwachsenen sind Motivation und Interesse. Nur solche Lerninhalte, die als notwendig und sinnvoll anerkannt werden, ermöglichen einen nachhaltigen Lernprozess. Oft sind Erwachsene weniger bereit wie Kinder, sich auf Lernen als vermeintlich „sinnloses“ Spiel einzulassen. Erwachsenenlernen steht vielmehr im Kontext konkreter Lebenssituationen und -ereignisse Übergänge, Übernahme neuer Rollen, Krisen, veränderte Anforderung aus der Arbeitswelt und ähnliches.<sup>122</sup>

### Reliabilität und Objektivität der Testergebnisse

Anhand der Testergebnisse möchte ich auch auf mögliche Einschränkungen hinsichtlich der Messgenauigkeit näher eingehen.

---

<sup>122</sup> Szablewsky-Cavus, Petra (2000): Verstehen und Verständigung in Deutsch. Grundzüge einer berufsbereisübergreifenden Didaktik.“ In: Bildungsarbeit in der Zweitsprache Deutsch, 2/2000. S. 16.

## Fehlerkorrektur und Auswertung

Wenn man Messgenauigkeit bei schriftlichen Sprachtests spricht, so fällt der erste Blick auf die Korrektur bzw. Auswertung des geschriebenen Textes.

Wenn auch die Korrektur bei Diktaten erfahrungsgemäß keine großen Abweichungen in der Bewertung zulässt, so kann man keineswegs Fehlbeurteilungen ausschließen. Wenn man davon ausgeht, dass das Diktat für Menschen mit nichtdeutscher Muttersprache eine schwierige Hörverstehensaufgabe ist, dann ist es meiner Meinung kaum möglich, *Hörverstehensfehler von Grammatik- bzw. Orthographiefehlern zu unterscheiden.*

In unserem Fall wird der diktierte Text „*Wenn Schnee zur Gefahr wird*“ von einer ausgebildeten Auswertungsperson gemäß Korrekturmodus ausgewertet. Eine Gegenkontrolle erfolgt nur dann, wenn Textstellen unleserlich sind bzw. der Verdacht auf motorische Schwächen wie z. B. Legasthenie besteht. Demnach entscheidet eine Person über die weitere berufliche Laufbahn eines Probanden. Darüber hinaus ist auch für die Eingabe der Prüfungsergebnisse nur eine Person vorgesehen.

Geht man von der Tatsache aus, dass *ein Fehler über „Förderbedarf“ oder „kein Förderbedarf“ entscheidet*, zeigt sich hier eine deutliche Beeinträchtigung der Reliabilität und Objektivität im Bereich der Auswertung und der Interpretation.

Unter diesen Bedingungen ist daher nicht auszuschließen, dass es zu falschen Zuordnungen innerhalb der Zulassungsentscheidung kommen kann. Die Konsistenz im Bewertungsverhalten der Auswertungspersonen ist daher meiner Ansicht nach nicht ausreichend gegeben. Im Extremfall bedeutet das, dass eine Person, die Förderbedarf benötigen würde, nicht gefördert wird.

## Korrelation zwischen den Testvarianten

Um den Bekanntheitsgrad eines Diktats möglichst gering zu halten, wurden fünf annähernd gleich lange (163 – 165 Wörter), authentische Texte (Zeitungstexte) für die Zulassungsprüfung Deutsch verwendet. Es handelt sich dabei um Texte, die laut Testersteller annähernd gleiche orthographische und grammatische Schwierigkeiten aufweisen.

Dennoch stellt sich die Frage, ob unterschiedliche Testvarianten ein hohes Maß an Reliabilität und Objektivität überhaupt erreichen können.

Wie die Gesamtergebnisse der Zulassungsprüfungen Deutsch aus dem Jahr 2010 zeigen, lassen sich, gemessen an der Höchstfehleranzahl, deutliche Unterschiede zwischen den

Testvarianten ausmachen. Wenn sich beispielweise ein Kandidat bei einer Wiederholungsprüfung innerhalb von drei Monaten von 62 Fehlern auf 28 Fehler verbessert, so liegt der Verdacht nahe, dass dabei ein wesentlich leichter Text verwendet wurde. Im umgekehrten Fall, würde der Förderbedarf bei einem schwächeren Kandidaten im Extremfall gar nicht aufscheinen.

### **13.2 Interpretation der Ergebnisse aus den Fragebögen**

Wie bereits erwähnt, stellt der Fragebogen in empirischer Hinsicht den wichtigsten methodischen Zugang zum Forschungsobjekt dar. Insofern waren meine Erwartungen hinsichtlich der Ergebnisse entsprechend groß.

Bei der Beurteilung der äußeren Form des Testformates ist die Mehrheit aller Befragten der Meinung, dass die äußere Form des Testformats kein Missfallen auslöst. Allerdings steht die Beurteilung eines Testformates immer im unmittelbaren Zusammenhang mit der Authentizität von Texten. Das bedeutet, dass durch die Authentizität von Textmaterialien zumeist auch die Akzeptanz des Tests (*Augenscheinvalidität*) erreicht werden kann. Darüber hinaus handelt es sich, so auch Bachmann und Palmer, bei der Bewertung um kein Expertenurteil, sondern um subjektive Einschätzungen von Lernern. Dennoch scheint der Motivationsschub durch die Akzeptanz vorhanden zu sein. Das Diktieren wird überraschenderweise zumindest in seiner äußeren Form (Angabe, Diktatzettel, Tonbandaufnahme) als Testformat angenommen.

Wenn man davon ausgeht, dass allen schwächeren Teilnehmern (Personen mit deutscher und nichtdeutscher Muttersprache) das Niederschreiben des Diktats große Probleme bereitet, sollten möglichst viele der Feststellung, dass es durch das Diktat zu Benachteiligungen kommen kann, zustimmen. Insofern überrascht es, dass etwa 70% der Befragten *Diskriminierungen durch das Testformat Diktat* ausschließen.

Betrachtet man die Beurteilungsergebnisse bei der Gruppe der Migranten, so fällt das Resultat noch deutlicher zu Gunsten des Diktats aus, ähnlich wie bei der Personengruppe, die mehr als eine Fremdsprache gelernt haben. Gerade Menschen mit nichtdeutscher Muttersprache sollten aufgrund ihrer Sprachlernerfahrung wissen, dass das orthographisch einwandfreie Niederschreiben von Gehörtem unter Zeitdruck eine äußerst schwierige Aufgabe darstellt.

Dieses Antwortverhalten lässt die vage Beurteilung zu, dass jene Personen, trotz der Anonymität des Fragebogens, aus Angst vor Konsequenzen nicht auffallen wollen.

In den militärischen Ausbildungsgrundlagen geht man davon aus, dass Soldaten Testsituationen wie beim Diktat psychisch gewachsen sind. Wie die beiden Stichproben zeigen, scheint es SoldatInnen, trotz der Anonymität des Fragebogens, schwer zu fallen, über Angstzustände zu schreiben. Dennoch gibt es einige Personen, wie in der Stichprobe 1, die sich *bei Testsituationen wie der Zulassungsprüfung bedroht fühlen*. Gründe dafür könnten schlechte Erfahrungen aus vergangenen Schulzeiten sein, wo Grammatik- und Orthographiefehler oftmals in Zusammenhang mit schlechten Noten und Sanktionen standen bzw. die existenzielle Angst vor dem Ausscheiden aus dem Militärdienst. Dass Testangst unmittelbare Auswirkungen auf das Testergebnis haben kann, soll hier zusätzlich erläutert werden.

Einen Schwerpunkt meiner Untersuchungsarbeit stellt die Berufsrelevanz des Testformats dar. Es ist daher sehr aufschlussreich, dass vor allem bei der zweiten Stichprobe viele Teilnehmer dem analogen reproduzierenden Schreiben einen *hohen Grad an Berufsrelevanz* beimessen, obwohl beim Militär zumeist nur stichwortartig mitgeschrieben wird. Daraus kann geschlossen werden, dass das Mitschreiben von Diktierem nicht nur als Notwendigkeit für den Berufsalltag angesehen wird, sondern dass die Bewältigung der Testaufgabe, laut Meinung der Testteilnehmer, den zu erwartenden kommunikativen Anforderungen einer militärischen Sprachverwendungssituation entspricht. Das bedeutet, dass sowohl die *Authentizität* als auch die *Konstruktvalidität* von den Befragten nicht in Frage gestellt wird. Gemäß den Nützlichkeitskriterien von Bachmann und Palmer könnte man aus diesem Blickwinkel das *Diktat als nützliche Testaufgabe* verstehen.

Von großem Interesse für jeden Testersteller ist, *ob anhand der Testergebnisse zukünftige Leistungen voraussagbar sind*. Wenn, wie in Stichprobe 1, die Mehrzahl der Teilnehmer feststellen, dass das Lösen der Testaufgabe keine Gewissheit gibt, den zu erwartenden sprachlichen Leistungen zu entsprechen, dann kann gemäß Nützlichkeitskriterien die *Validität des Testformats* nicht als ausreichend angesehen werden (*Vorhersagevalidität*).

Im Gegensatz zur Berufsrelevanz stuft der überwiegende Teil der Befragten von Stichprobe 1 das *Diktat als eine unnütze Aufgabe* ein. Diese Feststellung lässt den Schluss zu, dass das Schreiben des Diktats aufgrund der ungewöhnlichen Aufgabenstellung bei vielen Testteilnehmern *nicht genügend Interaktivität* auslöst. Wenn die Testaufgabe als solche als sinnlos angesehen wird, so ist meiner Meinung nach aufgrund *mangelnder Akzeptanz* die

*Augenscheinvalidität* nicht gegeben. Darüber hinaus muss die Sinnhaftigkeit einer Testaufgabe immer im Zusammenhang mit den Testergebnissen gesehen werden. Wenn die Kursteilnehmer im Vorfeld wissen, dass die Prüfung ausschließlich mittels eines Fehlerprozentschlüssels bewertet wird, dann kann das durchaus Auswirkungen auf die Einstellung und die Motivation haben.

Einen sehr wichtiger Themenkomplex in Bezug auf die allgemeinen Nützlichkeitskriterien eines Sprachtests stellt *das Interesse und die Motivation der Teilnehmer bei der Erfüllung der Aufgabe*<sup>123</sup> dar. Ähnlich wie beim Thema „nützliche Aktivität“ überwiegen negative Einschätzungen der Befragten.

Die geringe Wertschätzung des Testformats kann als Beeinträchtigung der Validität (Augenscheinvalidität) verstanden werden. Der pädagogische Wert des Testkonstrukts wird von den Testteilnehmern nicht erkannt.

Widersprüchliches zeigt sich beim Vergleich dieser Ergebnisse mit jenen beim Themenkomplex „Die äußere Form des Testformats löst bei mir Missfallen aus“. Etwa 25% der Befragten finden zwar Gefallen am Test, sehen jedoch keinen Sinn dahinter.

Wie die Auswertungen zeigen, stellen etwa 50 % aller Stichprobenteilnehmer fest, dass sie durch das Diktat keine sprachlichen Mängel erkennen konnten, obwohl sie, wie die Testergebnisse zeigen durchwegs schlecht abgeschnitten haben. Demnach stellen normorientierte Fehler, für zumindest die Hälfte der Versuchsteilnehmer, keinen Mangel an sprachlichen Fähigkeiten dar. Bei der Gruppe mit Hauptschulabschluss zeigt sich hingegen ein völlig anderes Bild. Etwa 90 % der Befragten stellen Mängel bei ihren sprachlichen Fähigkeiten fest. Meiner Einschätzung nach scheinen die Fördermaßnahmen, die etwa 60% aller Hauptschüler betreffen, durchaus angebracht zu sein. Allerdings bleibt die Frage offen, welche Mängel festgestellt wurden. Eine vertiefende Frage hätte vermutlich genauere Aussagen diesbezüglich zugelassen.

Nicht zu übersehen ist, dass es sich bei den 50%, die keine sprachlichen Mängel erkennen konnten, auch um Teilnehmer der Spitzengruppe handelt, für die das Diktat kaum eine Herausforderung darstellt.

Von besonderem Untersuchungsinteresse ist für mich der persönliche Eindruck der Befragungsteilnehmer. Relevante Ergebnisse ergaben sich insbesondere bei der Frage: „*Welche sprachlichen Teilfertigkeiten werden Ihrer Einschätzung nach beim Diktat geprüft?*“

---

<sup>123</sup> vgl. dazu Dłaska (2009), S. 38.

Die Ergebnisse lassen für mich den Schluss zu, dass nahezu alle Stichprobenteilnehmer beurteilen können, welche Teilfertigkeiten beim Diktat geprüft und nicht geprüft werden.

Wenn man aber die Beurteilungsergebnisse aller anderen Themenkomplexe im Fragebogen heranzieht, wird das Diktat zumeist als Test verstanden, der zur Überprüfung der sprachlichen Korrektheit dient. Es ist daher anzunehmen, dass die Mehrheit der Befragten die Teilfertigkeit Hörverstehen ohne Vorgabe nicht angeführt hätte. Wie schon in den Auswertungen der Fragebögen erwähnt, wurde bei der Themenstellung auf die kognitive Leistung beim Diktat nicht Bezug genommen.

Wie aus den Interpretationen der Testergebnisse hervorgeht, zeigen sich deutliche Diskrepanzen bei jenen Themenkomplexen, die in unmittelbarer Abhängigkeit voneinander stehen. Beim Vergleich der Befragungsergebnisse zeigt sich, dass sowohl die berufsspezifische Ausrichtung der Testaufgabe als auch das Diktat als solches, je nach Thematik, ziemlich unterschiedlich eingeschätzt werden. Man könnte zum einen annehmen, dass für einige die Themenstellung zu komplex ist und zum anderen das Interesse an der Fragebogenerhebung nachlässt bzw. zunimmt. Interessant erscheint auch die Tatsache, dass die beiden Stichproben nicht immer korrelieren. Ein Grund dafür könnte sein, dass der Ort und der Zeitpunkt der Fragebogenerhebung einen ziemlichen Einfluss auf das Ergebnis ausüben.

Ein eindeutigeres Ergebnis hatte ich mir beim Thema „*Benachteiligung aufgrund der Testaufgabe*“ erwartet. Dass nur die Hälfte der Befragten das Diktat als diskriminierend einschätzt, ist für mich überraschend. Ebenso unerwartet ist das Ergebnis beim Thema „*Testangst*“, wo nur wenige Teilnehmer das Diktat als Bedrohung ansehen. Dabei sollte mitgedacht werden, dass es sich hier um angehende Berufssoldaten handelt, für die Angst und Diskriminierung keine wichtigen Themen darstellen.

Im Zusammenhang mit den Qualitätskriterien eines Sprachtests lassen sich deutliche Mängel am Diktat ausmachen. Trotz nicht immer eindeutiger Einschätzungsergebnisse der Befragten kann gesagt werden, dass sowohl *Validität* als auch *Objektivität*, *Authentizität* und *Interaktivität* nicht ausreichend gegeben sind.

So aufschlussreich die Ergebnisse einzelner Themenkreise waren, haben sich die Erwartungen jedoch nur zum Teil erfüllt. Gründe für die meiner Ansicht nach nicht immer aussagekräftigen Daten sind vermutlich die zu komplexen Formulierungen der zu beurteilenden Aussagen innerhalb des Fragebogens bzw. das Ausfüllen des Fragebogens unmittelbar nach dem Test.

Dennoch hat der Fragebogen insbesondere bei den Themenstellungen „*Berufsbezogenheit*“ und „*Benachteiligung*“ untersuchungsrelevante Ergebnisse geliefert.

### 13.3 Interpretation und Zusammenschau der Experteninterviews

*Experten finden sich häufig auch bei jenen Funktionsträgern, die Entscheidungen vorbereiten und durchsetzen. Ihr Spezialwissen bezieht sich auf begrenzte Wirklichkeiten und Sichtweisen, und nur darauf zielen die Interviews.*<sup>124</sup>

Aufgrund des berufsspezifischen praxisorientiertem Wissen der beiden Experten standen die Vergegenwärtigung der Testentwicklung und die Testsituation im Mittelpunkt der Befragung. Wie zahlreiche empirische Arbeiten zeigen, geben Experteninterviews zumeist detaillierte Auskünfte über den Untersuchungsgegenstand und seine Rahmenbedingungen.

Unter dieser Prämisse lassen sich folgende grundlegende Schlussfolgerungen ableiten:

- Diktat als Kompromisslösung

Versuche, freie Schreibaufgaben als Testformat einzusetzen, scheiterten an der Reliabilität und Objektivität der Auswertung. Eine zu große Anzahl an Probanden in Verbindung mit einer zu geringen Auswertungszeit ließ zum damaligen Zeitpunkt keine umfangreiche Diagnose zu. Darüber hinaus gab es durch inkonsistente Messungen Probleme bei der Bewertung von produktiven Schreibaufgaben. Ein entsprechendes Diagnosewerkzeug (Raster) stand zum damaligen Zeitpunkt nicht zur Verfügung.

- Diktat soll Defizite in den Bereich Orthographie und Grammatik aufdecken

Obwohl ein Experte bemerkt, dass beim Diktat durchaus auch andere Fertigkeiten wie Hörverstehen oder Schreiben in Verbindung mit kognitiven Fähigkeiten zum Tragen kommen, soll das Diktat lediglich zur Überprüfung von Orthographie und Grammatik dienen. Das beweist auch die Auswertung mittels Fehlerprozentsatz.

Auch die bisherigen Fördermaßnahmen scheinen, zumindest im schriftlichen Bereich, auf das Wiederholen, Vertiefen und Üben von orthographischen und grammatischen Phänomenen abzielen.

---

<sup>124</sup> vgl. dazu Meuser, Michael & Nagel, Ulrike (1991): ExpertInneninterviews vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In: Garz, Detlef & Kraimer, Klaus (Hrsg.): Qualitativ empirische Sozialforschung. Konzepte, Methoden, Analysen. Opladen (Westdt.-Verl.) S. 441 - 442.

- Zulassungsprüfung als Mindestanforderung für die Unteroffiziersanwärter

Einer der beiden Experten meint, dass das Aufgabeniveau insofern zu rechtfertigen ist, als nach entsprechenden Fördermaßnahmen 90% aller Teilnehmer den Test positiv bestehen. Wenn man aber bedenkt, dass zumindest 50% der Teilnehmer mit deutscher Muttersprache und Pflichtschulabschluss das Diktat beim ersten Antreten nicht positiv bestehen, scheint entweder das Aufgabenniveau zu hoch oder die Testaufgabe nicht zu entsprechen. Mitberücksichtigt werden sollte, dass bei den Fördermaßnahmen das Diktat zumeist ein Schwerpunkt der Lernaktivität darstellt und das zahlreiche Teilnehmer mehrmals zur Prüfung antreten. Daraus lässt sich schließen, dass die 90 % nur durch präzise Vorbereitung, insbesondere durch „*teaching to the test*“ und „*test-taking-Strategien*“, bzw. durch oftmalige Testwiederholung erreicht werden.

- Anforderungsniveau der Probanden versus Niveau der Testaufgabe

Da es sich beim Diktat, wie ein Experte meint, um eine Mindestanforderung handelt, sollte das Niveau des Testformats klar definiert sein. Insofern überrascht es, dass sich beide Experten nicht einig sind, ob vom Referenzniveau B1 bzw. von der 8. Schulstufe ausgegangen werden sollte. Da laut Expertenmeinung mit dieser Prüfung in erster Linie Orthographie- und Grammatikkenntnisse überprüft werden, betrifft das meiner Meinung nach hauptsächlich schulisches Sprachwissen bzw. kognitive Fähigkeiten, während beim Referenzrahmen für Sprachen von Sprachkönnen bzw. Sprachkompetenz grundsätzlich ausgegangen wird. Angesichts dieser Feststellung stellt sich die Frage, wie in der Schule bei der Überprüfung von Rechtsschreib- bzw. Grammatikwissen bei der Bewertung vorgegangen wird. Insgesamt scheint es Testteilnehmer zu motivieren, wenn die Aufgabe anspruchsvoll gestaltet und bewältigbar ist. Insgesamt sehe ich beim Diktat nicht nur Probleme in Bezug auf das Niveau, sondern auch in Bezug auf das *Testkonstrukt* im Sinne der *Validität*. Durchaus signifikant sind auch die unterschiedlichen Betrachtungsweisen der Experten bezüglich des Testgegenstands. Während für den einen Experten das Diktat allgemeinsprachliche Fähigkeiten überprüft, steht für den anderen das berufsrelevante Wissen im Vordergrund.

- Militärische Ausrichtung von Texten – Authentizität

Laut Expertenmeinung gehört das Niederschreiben von Gehörtem zu den wesentlichen kommunikativen Aufgaben eines Soldaten. Allerdings, so der Experte, passiert das

zumeist mittels Gesprächsnotizen. Aus wie bereits erwähnten, praktikablen Gründen wird das Diktat nicht stichwortartig, sondern Wort für Wort wiedergegeben. Demzufolge handelt es sich meiner Meinung nach nicht mehr um eine berufsspezifische Aktivität, besonders dann, wenn die Korrektheit des Schreibens im Vordergrund steht. So gesehen hält das Testkonstrukt nicht das, was es verspricht. Die Testaufgabe kann somit *weder als valide noch als authentisch gelten*, obwohl, wie ein Experte meint, durchaus authentische militärbezogene und niveaugerechte Textsorten verwendet werden.

Die authentische Ausrichtung von Textsorten dient in diesem Fall, meiner Meinung nach, lediglich dem Zweck des Augenscheins. Der Text „*Wenn Schnee zur Gefahr wird*“ steht weder vom Niveau noch von der fachlichen Ausrichtung in einem Zusammenhang mit dem Bundesheer!

- Diktat als ökonomisches Verfahren

Warum gerade das Diktat aufgrund seiner ökonomischen Handhabbarkeit das Lückendiktat ersetzen soll, geht aus den Interviews nicht hervor.

Noch ökonomischer wäre meiner Ansicht nach sicherlich ein C-Test, der wesentlich aussagekräftigere Ergebnisse liefern würde. Nicht zu unterschätzen ist beim Diktat auch der Zeitaufwand für die Auswahl des Sprechers und die Aufnahme bzw. die Auswahl von möglichst konsistenten Diktattexten.

- Förderung versus Diskriminierung

Ebenso wie bei der Fragebogenerhebung gilt diesem Thema mein besonderes Interesse, weil es unmittelbar mit der Forschungsfrage in Zusammenhang steht. Beide Experten meinen, dass eine Benachteiligung aufgrund der geringen Zahl der Migranten in Kauf genommen wird, weil es bei der Prüfung nicht um Selektion im Sinne von Ausscheiden, sondern ohnehin nur um Förderung geht. Angesichts dieser Formulierung stellt sich die Frage, ob ein nach einem Auswertungsschlüssel (Prozentanteil an Fehlern) bestimmter Förderbedarf nicht schon einer diskriminierenden Maßnahme gleichkommt. Nicht zu vergessen ist dabei, dass das Nichtbestehen von Prüfungen, besonders bei Migranten, mit existenziellen Ängsten verbunden ist.

Wenn, beispielsweise ein Testteilnehmer davon ausgeht, dass er eine Aufgabe wie das Diktat nicht ohne Fehler niederschreiben kann, dann kann es zur Resignation kommen.

Wie ich bei der Beobachtung feststellen konnte, kapitulierten einzelne Testteilnehmer bereits nach 5 Sätzen.

- Lernaktivität (Nützlichkeit) und Testformat

Warum, laut Aussagen beider Experten, eine Lerneraktivität bei der schriftlichen Zulassungsprüfung nicht beabsichtigt ist, erscheint ausreichend geklärt. Vor allem dann, wenn die Ergebnisse beim Diktat unmittelbar mit Förderoptionen verknüpft sind. *Interaktivität* wäre nur dann gegeben, wenn die Testperson durch das Interesse an der Aufgabe einen unmittelbaren Nutzen hätte.

- Vorhersagevalidität fraglich

Laut Experten liegen keine aussagekräftigen Ergebnisse auf. Von großem Interesse wäre es meiner Ansicht nach, ob Personen mit negativen Diktatergebnissen schlechter abschneiden würden als jene mit positiver Testleistung. Darüber hinaus müsste die Frage gestellt werden, ob die im Diktat abgebildeten Fähigkeiten in dem Maße überhaupt verlangt werden. Im Kompetenzkatalog für Einsatz- und Friedensfunktionen wird jedenfalls das wortwörtliche Reproduzieren von Texten nicht aufgelistet.

- Objektivität der Auswerter

Wenn lediglich 1 Fehler über Fördern bzw. Nichtfördern entscheidet, dann ist meiner Meinung nach die Objektivität nicht gegeben. Nicht zu vergessen sind hier nicht zu vermeidende Auswertungsfehler aufgrund von Müdigkeit oder anderen Einflüssen wie Störungen, bzw. Unterbrechungen. Dabei kann es trotz des abgestimmten Korrekturverhaltens durchaus zum Übersehen von Fehlern, zu einer falschen Diagnose bei unleserlichen Passagen, sowie zu Fehlern beim Zusammenzählen kommt, was erfahrungsgemäß immer vorkommen kann. Demnach sind die Ergebnisse abhängig vom Auswerter. Es ist daher sowohl die *Reliabilität* als auch die *Objektivität* nicht ausreichend gegeben.

Auch wenn die Befragung in Form eines Interviews laut empirischer Sozialwissenschaft, als bewährte Untersuchungsmethode gilt, so können sich durchaus Probleme ergeben.

Bei institutsinternen Experten besteht z. B. die Gefahr, dass sozial erwünschte Antworten gegeben bzw. Meinungen vertreten werden, die sich nach beruflichen Konventionen richten. Nicht zu vergessen sind die Einflüsse durch die Untersuchung selbst.

In inhaltlicher Hinsicht zeigt sich, dass beide Experten bei der *sprachspezifischen Fundierung des Testverfahrens* das Diktat als Kompromisslösung sehen, die lediglich aus pragmatischen Gründen gewählt wurde. Dennoch führen ebenfalls beide Experten an, dass bei der Einführung des Diktats die Überprüfung von orthographischen Fähigkeiten durchaus erwünscht war.

Hinsichtlich der *sprachlich gebundenen Handlungsfähigkeit im Beruf* argumentiert insbesondere Experte 1, dass das Diktat ausschließlich zur Überprüfung der allgemein sprachlichen Fähigkeiten dient und dass auf eine berufsspezifische Ausrichtung bewusst verzichtet wurde. In Bezug auf die *Benachteiligung von Migranten* betonen beide Experten, dass das Diktat vordergründig nicht als Selektionsmittel, sondern lediglich Fördermittel vorgesehen war. Eine Diskriminierung ist zwar gegeben, sollte aber keine unmittelbaren Auswirkungen auf den Verbleib als Berufssoldat haben.

Bei den *Güte- und Qualitätskriterien* sehen beide Experten deutliche Mängel, hauptsächlich im Bereich der Validität. Ebenso einig sind sich beide Experten, dass das Diktat *keine präzisen Förderoptionen zulässt*.

Durchaus bemerkenswert sind die Expertenaussagen zur *Nützlichkeit des Diktats*.

Während ein Experte meint, dass eine Lernaktivität gar nicht beabsichtigt sei, empfindet der andere das Diktat sogar als demotivierende Aufgabe.

Gleichermaßen positiv fällt das Urteil beider Experten hinsichtlich der *Praktikabilität* aus.

Sowohl Experte 1 als auch Experte 2 sehen das Diktat als ökonomisches Verfahren, das mit wenig Aufwand – vor allem im Bereich des Auswertens – zu bewältigen ist.

Insgesamt zeigen beide Expertenmeinungen, dass trotz systembedingter Herangehensweisen durchaus kritisch über heeresinterne Auswahlverfahren nachgedacht und die Möglichkeit für Evaluierungen geboten wird.

## **14 Nützlichkeit und Qualität des Diktats als Schreibaufgabe**

### **14.1 Überprüfung nach den Nützlichkeitskriterien**

Wenn man die Testergebnisse, das Teilnehmer- und Expertenurteil sowie das Beobachtungsurteil anhand der Nützlichkeitskriterien von Bachmann und Palmer (vgl. dazu Kapitel 5.3) vergleicht und zusammenfasst, lassen sich folgende Rückschlüsse ziehen:

*Die Konsistenz der Messungen (Reliabilität) ist beim Diktat nicht ausreichend gegeben!*

Aufgrund des Auswertungsverfahrens und der zumeist subjektiven Entscheidungsfindung (Fehlerauffindung, Fehlersummierung, Fehlerprozentsatz sowie Übertragungsfehler) kann es zu Einschränkungen sowohl in der *Reliabilität* als auch in der *Objektivität* kommen. Zusätzlich kann sich auch Milde und Strenge bei der Diktatkorrektur erheblich auswirken. Testangst, mangelnde Praxis mit der Handschrift und die Ermüdung beim Schreiben verfälschen das Ergebnis im Sinne der *Reliabilität*. Die Interpretation der Ergebnisse beschränkt sich ausschließlich auf das Zählen von Fehlern. Dabei sind Messfehler insbesondere durch die Gleichgewichtung von „fehlenden Wörtern“ und „fehlerhaften Wörtern“ gegeben (*Auswertungsreliabilität*).

Messfehler können aber auch aufgrund der unterschiedlichen Diktattexte bei Prüfungswiederholungen auftreten, insbesondere dann wenn die Diktattexte nicht den gleichen Schwierigkeitsgrad aufweisen (*Wiederholungsreliabilität*).

Hinsichtlich der *Objektivität* garantiert das Testformat keine fairen Bedingungen für alle Testteilnehmer. Die Testaufgabe und der Auswertungsmodus entscheiden größtenteils zugunsten bestimmter Herkunftsgruppen aus bildungsnahen Kreisen. Beim Diktat handelt es sich um eine Form des analytischen Hörverstehens in Verbindung mit einer künstlich geschaffenen Schreibsituation, dem normgerechten und wortgetreuen Aufschreiben von Wörtern. Das Wahrnehmungstempo, insbesondere beim Hörverstehen ist bei Personen mit nichtdeutscher Muttersprache zumeist langsamer und erfordert eine höhere Konzentration auf das Hören und lenkt vom Schreiben ab. Aufgrund der außersprachlichen Einflüsse kann das Testergebnis daher nur verzerrte und somit ungültige Aussagen liefern.

*Die Gültigkeit des Diktats (Validität) ist auf Basis der Interpretation der Forschungsergebnisse nicht gegeben!*

Der niedergeschriebene Text kann lediglich als Indiz für eine übergreifende, im Testkonstrukt nicht definierte Fähigkeit (z. B. Bildungslücken bei Menschen mit nichtdeutscher Muttersprache, fehlendem Grammatikwissen in Bezug auf Groß- und Kleinschreibung) begriffen werden. Die vorgenommene Definition des Testkonstrukts greift daher zu kurz, sie bezieht sich nur auf die Oberfläche des Geschriebenen und nicht auf das tatsächliche Wissen und Können. Ziel einer Testaufgabe wie des Diktats, darf es nicht sein, ausschließlich die Fähigkeit zur Lösung zu erfassen

(Orthographiewissen bzw. Strategiewissen und Formatwissen). Das Diktat gibt außerdem keine Auskunft über jenen aktiven Wortschatz, der einem Schreibenden aufgrund seiner schulischen Bildung und seiner beruflichen Erfahrungen zur Verfügung steht. Gleichermäßen problematisch ist, dass sich die Diktate nicht nur auf ein begrenztes Wortmaterial beziehen, sondern dass es sich dabei auch um nicht mehr aktuelle Texte handelt, die zumeist aus vergangenen Jahren stammen. Die zu testende Fähigkeit ist meiner Ansicht nach daher auch in sprachtheoretischer Sicht problematisch. Bemerkenswerterweise können selbst Orthographie- und Grammatikschwächen mit dem Verfahren nicht ausreichend diagnostiziert werden. Aus Sicht der Bedarfsträger (Augenscheininvalidität) sind der Sinn und der Zweck der Aufgabe nicht nachvollziehbar. Ebenso lässt das Diktat in Hinsicht auf das Testkonstrukt daher nur ungenaue und unvollständige Interpretationen zu, die *Konstruktvalidität* ist nicht gegeben.

*Das wortgetreue Reproduzieren von Texten spiegelt keine realen beruflichen Sprachsituationen im Sinne der Authentizität wieder!*

Trotz des von den Befragten zum Teil positiv eingeschätzten Textmaterials (Zeitungsartikel) handelt es sich beim Diktat um keine authentische Aufgabenstellung. Darüber hinaus entspricht weder das vollständige Niederschreiben von Gehörtem dem beruflich geforderten „*Schriftlichen Gebrauch von Sprache*“ noch das langsam betonte Sprechen des Ansagers der üblichen Berufs- bzw. Befehlssprache. Außerdem wird das Reproduzieren von Texten nicht nur in realen Sprachverwendungssituationen, sondern auch im Kompetenzkatalog für Einsatz- und Friedensfunktionen ausschließlich als stichwortartiges Mitschreiben verstanden.

*Das Ausmaß an Aktivierung und Einbeziehung der KandidatInnen in die Testsituation kann als nicht ausreichend bezeichnet werden!*<sup>125</sup>

Gemäß den Daten der Fragebogenuntersuchung ist das Interesse der Teilnehmer an der Erfüllung der Aufgabe gering. Die Aufgabe wird vom Großteil der Befragten in Bezug auf den Lernprozess als unnützlich und sinnlos eingeschätzt. Das Diktat wird lediglich als Notwendigkeit verstanden.

---

<sup>125</sup> vgl. dazu Dlaska (2009), S. 38.

*Das Testformat Diktat steht in einem sinnvollen Verhältnis zum Aufwand!*

Gemäß Experteninterviews kann festgestellt werden, dass die Erstellung, der Einsatz als Prüfmittel und die Ermittlung der Ergebnisse in einem sinnvollen Verhältnis zu den verfügbaren Mitteln stehen. Das Testverfahren kann sowohl in materieller als auch in personeller Hinsicht als praktikabel eingeschätzt werden.

## **14.2 Auswirkungen auf die Zulassungsentscheidung**

Im Zuge von Probetestungen bei Unteroffizieren hat sich herausgestellt, dass man trotz eines erfolgreich absolvierten Diktats nur vage Aussagen darüber machen kann, ob die Person über ausreichende Orthographie- und Grammatikkenntnisse verfügt. Das Diktat kann lediglich grobe Auskünfte darüber geben, welche gravierenden Rechtschreibprobleme eine Testperson zum Zeitpunkt des Tests hat. Bei der Bewertung, die auf sprachliche Korrektheit abzielt, wird auf direktem Wege versucht, nicht Sprachkönnen, sondern Mängel zu identifizieren. Außer der Textlänge stehen Bewertung und Textauswahl in keinem nachvollziehbaren Zusammenhang.

In diesem Zusammenhang stellt sich auch die Frage, inwieweit und in welcher Zeit Erwachsene ihre Orthographie- und Grammatik-Kenntnisse überhaupt verbessern können.

Wie die Praxis zeigt, können sich die zielgerichteten Vorbereitungsmaßnahmen auf die Prüfung (*teaching to the test*) auch negativ auf den Lernprozess auswirken. Durch das regelmäßige Üben eines Testformats wie dem Diktat ist eine Verkümmern der Textfähigkeit zu beobachten. Das bedeutet, dass sich die didaktischen und methodischen Maßnahmen der Lehrpersonen im Zuge der Sprachfördermaßnahmen zumeist an der Prüfung orientieren und nicht am Sprachstand der Personen.

Sieht man von der Tatsache ab, dass bei der computerunterstützten Textverarbeitung meist mit Hilfe von Rechtschreibprogrammen gearbeitet wird, stellt das korrekte Niederschreiben von Texten beim Militär eher eine Ausnahme dar (Funksprechverkehr).

Auch die Verwendung von Wörterbüchern ist bei der Erstellung von militärischen Textsorten nichts Ungewöhnliches. Das in der Praxis übliche Nachschlagen unbekannter Wörter im Wörterbuch ist beim Diktat nicht erlaubt bzw. nicht Zweck der Prüfung. Für die Kontrolle könnte z. B. ein Wörterbuch zur Verfügung stehen, dessen Gebrauch mitbewertet wird.

Unter diesen beruflichen Bedingungen, kann durch das Diktat die berufssprachliche Qualifizierung für den schriftlichen Gebrauch von Sprache nicht gewährleistet werden.

### **14.3 Folgenabschätzung in pädagogischer und gesellschaftlicher Hinsicht**

Dadurch, dass sich die Selektion bei der Zulassungsprüfung Deutsch sowohl auf eine oberflächliche Interpretation der Ergebnisse als auch auf eine nicht ausreichende Beschreibung des Testkonstrukts bezieht, kann es vorkommen, dass hinsichtlich des Förderbedarfs falsche Schlussfolgerungen gezogen werden. Das betrifft insbesondere jene Personen, die beim Diktat positiv abschneiden, aber Förderung benötigen würden. In gesellschaftlicher Hinsicht könnte das Ausscheiden von Menschen mit Migrationshintergrund vor allem dann größere Auswirkungen haben, wenn die Anzahl von Migranten in Zukunft ansteigt.

Nicht zu unterschätzen dabei sind auch die pädagogischen Konsequenzen („*Backwash-Effekt*“), im Bereich der Fördersystematik. Das Diktat hat sowohl Auswirkungen auf die Gestaltung und Zielsetzung des Förderprogramms als auch auf Einstellungen der LernerInnen zum Unterricht.<sup>126</sup>

## **15 Schlussfolgerungen**

### **15.1 Hinweise auf Mängel**

Bei den Erhebungsdaten der Stichproben 1 und 2 zeigen sich zum Teil große Abweichungen, obwohl bei den Testergebnissen durchaus konsistente Werte vorliegen.

Warum hier derartige Differenzen im Antwortverhalten aufgetreten sind, kann meiner Ansicht nach nur damit begründet werden, dass es sich bei der Stichprobe 2 um einen anderen Ort und einen anderen Zeitpunkt handelte. Interessant scheint hier auch der Umstand zu sein, dass bei der Erhebung 1 in Ebelsberg fast ausschließlich Hauptschüler an der Untersuchung teilnahmen.

### **15.2 Bestätigung der Hypothese**

Bei der vorliegenden Untersuchung galt es, den schriftlichen Teil der Zulassungsprüfung Deutsch im Bezug zu den erwartbaren sprachlichen Kompetenzen eines Berufsunteroffiziers bei der Produktion von Texten zu prüfen. Es musste daher geklärt werden, ob das Prüfungsformat Diktat den Zweck der Qualifikation für das berufsspezifische Schreiben und

---

<sup>126</sup> vgl. dazu Glaboniat (1998), S. 49 – 50.

den Zweck der Zulassung zu einer Sprachfördermaßnahme erfüllt. Dabei sollten außersprachliche Faktoren, die das Testergebnis beeinflussen können, untersucht werden.

Um meine Schlussfolgerungen aus den Untersuchungsergebnissen zu verdeutlichen, sollen im Folgenden die einzelnen Schritte meiner Untersuchung und die wichtigsten Erkenntnisse im Hinblick auf die Hypothesenstellung zusammengefasst werden.

Im ersten, theoretischen Teil meiner Arbeit steht die Einordnung und Bestimmung des Testformats „Diktat“ im Vordergrund. Dabei galt es festzustellen, warum und in welchem institutionellen Rahmen es eine Deutschprüfung beim ÖBH gibt. Wie aus meinen Ausführungen über die Deutschausbildung beim Bundesheer hervorgeht, handelt es sich dabei in erster Linie um die Überprüfung und Förderung von Fähigkeiten zur Zusammenarbeit in sprachlicher Hinsicht.

Wie im Kapitel 1 beschrieben, richtet sich der Förderbedarf nach zwei Prüfungen, der Einstufungsprüfung Deutsch beim Heerespersonalamt und der Zulassungsprüfung Deutsch.

Neben einer Wortlückenaufgabe stellt das Diktat mit etwa 163 Wörtern den zentralen schriftlichen Teil der Zulassungsprüfung Deutsch dar. Bei der Auswertung wird mittels Fehlerschlüssel das Ausmaß der korrekten Schreibung, insbesondere der Rechtschreibung, festgestellt. Wie die Entwicklungsgeschichte der Zulassungsprüfung Deutsch zeigt, gab es seitens der Arbeitsgruppe Prüfmittel zahlreiche Bemühungen, vor dem Entschluss zum Diktat ein „freies Schreiben“ einzusetzen. Aufgrund ökonomischer Überlegungen bzw. qualitativer Einschränkungen scheiterten jedoch diese Vorhaben.

Im zweiten Teil der theoretischen Untersuchung geht es im Wesentlichen um problembezogene Forschungsliteratur, die als Hypothesengerüst für meine Forschung dient. Dabei galt es, eine Annäherung an den Forschungsgegenstand anhand von aktuellen wissenschaftlichen Arbeiten der FÖRMIG-Gruppe (Sprachprofilmessungen) bzw. eines norddeutschen Forschungsprojekts (Hamburger Schreibprobe) vorzunehmen (siehe Kapitel 4.1 – 4.4). Um einen Bezug vom Sprachniveau der Testteilnehmer zu den schriftlichen Anforderungen des Diktats herstellen zu können, habe ich mich sowohl mit dem Fehlerbegriff innerhalb der Forschung als auch mit den bildungspolitischen Vorgaben (Bildungsstandards, Europäischer Referenzrahmen) auseinandergesetzt.

Neben grundlegenden sprachtesttheoretischen Überlegungen wurden Qualitätsanforderungen an Sprachtests bzw. Gütekriterien vorgestellt und diskutiert.

Um den Ansprüchen einer qualitativen Sprachtest-Untersuchung gerecht zu werden, bezog ich mich bei der Untersuchung des Diktats in erster Linie auf die bewährten Nützlichkeitskriterien von Bachmann und Palmer. Im Vordergrund stand dabei die Validität,

das für mich entscheidende Nützlichkeits- und Qualitätskriterium. Im Bereich des Messens und Prüfens zeigte sich insbesondere, dass die Validierung sich oft als sehr umfangreiches und komplexes Verfahren darstellt, wo die Berücksichtigung unterschiedlicher Gesichtspunkte und Meinungen (Testteilnehmer, Experten) notwendig erscheint.

Noch vor dem empirischen Teil meiner Untersuchung habe ich mich mit der Problematik der berufsspezifischen Ausrichtung von schriftlichen Aufgabenstellungen wie dem Diktat auseinandergesetzt, um hinsichtlich des zu entwickelnden Berufskompetenzmodells entsprechende Grundlagen zu haben. Den Abschluss des theoretischen Teils meiner Arbeit bildete eine zusammengefasste Expertise über das Prüfmittel „Diktat“.

Unter Einbeziehung des im theoretischen Teil erlangten Wissens über die institutionellen und fachlichen Rahmenbedingungen und Problematik des Testverfahrens wurde im empirischen Teil das Diktat einer Untersuchung unterzogen. Um das Validitätsproblem, aber auch die Generalisierung qualitativer Methodik keineswegs unberücksichtigt zu lassen, stützte ich mich auf eine Kombination unterschiedlicher Datenerhebungsmethoden. Dabei handelt es sich um eine analytische Interpretation der Testergebnisse, Fragebogenerhebungen, Expertenleitfadeninterviews sowie eine Datenerhebung durch teilnehmende Beobachtung. Im Rahmen dieser Untersuchung wurde anhand von zwei Stichproben in Ebelsberg und Enns von Dezember 2010 bis April 2011 eine Evaluierung des Diktats durchgeführt, wo ich sowohl die Testergebnisse ausgewertet und interpretiert, als auch eine Fragebogenuntersuchung in Verbindung mit einer Beobachtung durchgeführt habe. Im Zuge der Evaluierung galt es, nicht nur das Testverfahren anhand der Nützlichkeitskriterien, sondern insbesondere auch außersprachliche Einflüsse zu untersuchen. Die Auswahl und Darstellung der Ergebnisse beschränkte sich auf untersuchungsrelevante Themenkomplexe. Bei der Interpretation der Untersuchungsergebnisse wurden die Messwerte anhand von Themenkomplexen strukturiert und in Beziehung gebracht.

Aufgrund der Untersuchungsergebnisse kann festgestellt werden, dass sich die Hypothese bestätigt hat.

*Die Ergebnisse der schriftlichen Zulassungsprüfung Deutsch liefern nur unzureichende Informationen über jene schriftsprachlichen Qualifikationen, die den militärischen Anforderungsprofilen entsprechen. Insbesondere lassen die Ergebnisse nur unzureichend Rückschlüsse auf sprachliche Mängel zu und erfüllen den Zweck der Zulassung zu Sprachfördermaßnahmen nicht. Insbesondere verfälschen außersprachliche Faktoren die Prüfungsergebnisse.*

*In der Untersuchung wurde deutlich, dass sich mit dem Diktat weder die Schreibfähigkeit im militärischen Berufsfeld noch allgemeinsprachliche Fähigkeiten überprüfen lassen. Demnach überprüft der Test nicht das, was er überprüfen soll.*

Dieser Faktor zeigt sich insbesondere bei der dargestellten Analyse nach den Nützlichkeitskriterien, wo die zentralen Kriterien, wie Validität, Reliabilität und Objektivität, nicht erfüllt werden.

Darüber hinaus lässt die Testaufgabe keine gezielten Förderoptionen zu.

Insgesamt kann gesagt werden, dass das Niederschreiben von Angesagtem im Gegensatz zum freien Schreiben einen unnatürlichen, von außen gesteuerten Prozess darstellt. Das Schreiben wird ausschließlich unter dem Aspekt des Normbezugs ohne Berücksichtigung von inhaltlichen und kommunikativen Kriterien bewertet. Statt Sprachkönnen werden kognitive Fähigkeiten wie Weltwissen und Sachwissen vorausgesetzt. Bewertet wird lediglich das, was von der Norm abweicht. Für das Bestehen der Prüfung sind daher zumeist soziale und sprachbiographische Faktoren ausschlaggebend; Mehrsprachigkeit wird zum Ausschlussgrund. Für Nichtmuttersprachler stellt das Diktat deshalb eine schwierige und problematische Aufgabenstellung dar, denn nicht das bereits Gelernte, sondern das nicht vorhandene Wissen wird gewertet.

### **15.3 Vergleich mit anderen Prüfverfahren**

Im Vergleich zu anderen empirischen Untersuchungen von Sprachtests (z. B. Test-DaF, ÖSD) kann gesagt werden, dass ein idealer Aufgabentyp, der einen eindeutigen Bezug zur Sprachrealität liefert, meiner Ansicht nach in der Praxis kaum realisierbar ist. Jegliche definierte sprachliche Fähigkeit kann nur im Zusammenhang mit dem Lernprozess interpretiert werden. Das bedeutet, dass die Textoberfläche, insbesondere beim Diktat, nur einen kleinen Ausschnitt des Sprachstands hervorbringt. Es ist daher meiner Meinung nach ein naives Unterfangen, eine völlige Validität bei einer Sprachprüfung erreichen zu wollen. Ein hohes Maß an Eindeutigkeit der zu messenden Fähigkeit(en) kann ausschließlich Sprachprofilmessungen wie z. B. *Bumerang* (siehe Kap. 4.1) zugeschrieben werden, wo der Lernprozess ein wesentliches Kriterium der Messung darstellt. Interessanterweise kommt auch die „Hamburger Schreibprobe“, wenngleich es dabei wie beim Diktat vordergründig nur um Orthographie geht, einem idealen Aufgabentyp ziemlich nahe. Dieses meiner Ansicht nach sehr präzise Diagnoseinstrument führt anhand der unterschiedlichen Schreibaufgaben

vor, welche unterschiedlichen Phänomene beim schriftsprachlichen Ausdruck zu berücksichtigen sind. Fehler treten oft nur in orthographischer Ummantelung nach außen. Dahinter verbergen sich zumeist völlig andere Phänomene wie Schwächen im Hörverstehen, Grammatikschwächen (hauptsächlich Syntax), Konzentrationsschwächen und Flüchtigkeitsfehler, aber auch Probleme in der Aussprache.

Die „Hamburger Schreibprobe“ (siehe Kap. 4.4) zeigt aber auch, dass es durchaus möglich ist, Orthographieschwächen zu diagnostizieren bzw. daraus Fördermaßnahmen abzuleiten. Dabei darf man aber nicht übersehen, dass es aus wissenschaftlicher Sicht nahezu aussichtslos ist, all jene Wörter in einem Fundus zusammenfassen, die ein Mensch mit Pflichtschulabschluss korrekt zu schreiben imstande ist.

Das Diktat kann daher im Vergleich mit anderen Sprachmessinstrumenten lediglich als institutionelles Kontrollverfahren verstanden werden. Es überprüft keinesfalls den aktuellen Sprachstand und schon gar nicht eine Lernmethode.

## **16 Erprobungs- und Evaluierungsschritte bei der Einführung des Testformats NEU**

### **16.1 Berufskompetenzmodell als Basis zur Erstellung der neuen Testaufgabe**

Um den Untersuchungsgegenstand, die schriftsprachlichen Kompetenzen eines Unteroffiziers, in Hinsicht auf die neue Testaufgabe erfassen zu können, gilt es für mich vorerst den zu erwartenden Schriftverkehr des Unteroffiziers zu untersuchen. Dabei soll eine Grundlage für den Sprachbedarf und die Sprachbedürfnisse eines Unteroffiziers im Rahmen von berufsorientierten und befehlsorientierten Handlungsketten geschaffen werden. Auf Basis der Bildungsstandards der 8. Schulstufe und dem Personalstrukturmodell Unteroffizier habe ich ein Berufskompetenzmodell konzipiert, das konkrete Aspekte der tatsächlichen Arbeits- und Sprachwirklichkeit (sprachliches Handlungsfeld) erfassen soll, um daraus resultierende Prüfverfahren ableiten zu können.

Es handelt sich dabei insbesondere um folgende Kriterien:

- Erfassung konkreter linguistischer Anforderungen
- Einbeziehung der operativen funktionalen Dimension von Sprache
- Erfassung von adressaten-, situations- und berufsbezogenen Kommunikationsketten
- Erfassung berufstypischer Fähigkeiten
- Erfassung von zu erwartenden kommunikativen Handlungsketten
- Berücksichtigung von außersprachlichen, innersprachlichen und soziolinguistischen Merkmalen (siehe Kap. 3.1)

Entscheidend dabei waren genormte militärische Textsorten, die bei berufstypischen Aktivitäten eine relevante Rolle spielen.

Die Merkmale dieser berufsrelevanten Textsorten sind zumeist recht unterschiedlich. Während bei normierten, linearen Textsorten, wie z. B. einer militärischen Niederschrift ein ganzes Bündel von Kriterien zum Tragen kommt (Inhalt, Textkohärenz, Lexik, verständliches und korrektes Niederschreiben usw.) spielen bei der schriftlichen Wiedergabe von Befehlen und Vorschriften (Notizen, Mitschriften) zumeist nur inhaltliche und fachspezifische Faktoren eine Rolle. Das Ausfüllen von genormten Formularen hingegen (z. B. Anforderungsformulare für Versorgungsgüter) verlangt in erster Linie Knappheit und Verständlichkeit in Verbindung mit sprachlicher Korrektheit.

Der folgende Ausschnitt aus dem Berufskompetenzmodell soll zeigen, welche unterschiedlichen Testformate sich aus den berufstypischen Schreibaktivitäten ableiten lassen:

**Kompetenzbereich: SCHREIBEN**

Bildungsstandards/ Teilkompetenzen	Personalstrukturmodell Unteroffizier Qualifizierungsstufe 2 (GrpKdt)	Konkretisierung der Fähigkeiten „berufstypische Schreibaktivitäten“	Geeignete Prüfverfahren
<b>2. Texte verfassen</b>			
2.2. Sachverhalte und Inhalte nachvollziehbar, logisch richtig und zusammenhängend formulieren	Sichere Anwendung der Grundkenntnisse der deutschen Sprache einschließlich des sinnerfassenden Lesens auf Pflichtschulniveau in Bezug auf Lesbarkeit und Verständlichkeit in Wort und Schrift einschließlich militärischer Fachsprache <b>Sozial-kommunikative Kompetenz</b> <b>Kognitiver Aspekt</b> Nutzung des Wissen aus ziviler Schul- und/oder Berufsausbildung <b>Fach- und Methodenkompetenz</b> <b>Kognitiver Bereich</b>	beim Schreiben von Berichten und Niederschriften grundlegende textspezifische Mittel anwenden	<b>Reproduzieren</b> Diktat, Gedächtnisprotokoll <b>Bezeichnen</b> Beschriften Gegenstände, Phänomene benennen Liste, Verzeichnis, Merkzettel anlegen (Kofferpacken) <b>Ergänzen</b> Lückentext füllen Informationen in einen vorgegebenen Text einfügen Textteile ergänzen <b>Umformen</b> Textsorte, Textperspektive wechseln <b>Ausformulieren</b> <b>Strukturieren</b> <b>Produzieren</b> Briefe beantworten Gegenstands-Beschreibungen Betriebsanleitungen Formulare ausfüllen Bildgeschichten Leerstellen eines Textes füllen
2.3. für ein Thema relevante Argumente und Gegenargumente formulieren und sie sprachlich verknüpfen bzw. gegenüberstellen		aus Mitschriften und Notizen lineare und nichtlineare Texte verfassen	
2.4. formalisierte lineare Texte/nicht-lineare Texte verfassen (Lebenslauf, Bewerbungsschreiben, Formulare usw.)		Schreiben von Befehlen auf Basis des allgemeinen Befehlsschemas Ausarbeiten von schriftlichen Ausbildungsaufträgen, Formulare insbesondere militärspezifische Textsorten wie Meldungen, Niederschriften	
2.5. unter Einhaltung wesentlicher Kommunikationsregeln an einer medialen Kommunikation teilnehmen (E-Mail usw.)		Berichte, Anträge unter Berücksichtigung von Exaktheit, Eindeutigkeit (Einbeziehung von fachspezifischen Terminologien) und Ökonomie verfassen nicht lineare Texte wie Wünsche und Beschwerden verfassen	
2.6. Schreiben als Hilfsmittel für eigenes Lernen einsetzen (Stichwortzettel, Zusammenfassung usw.)		im E-Mailverkehr relevante Kommunikationsregeln anwenden  Zusammenfassungen und Stichwortzettel für Befehle Anweisungen, Belehrungen und Einweisungen verfassen	
<b>4. Schreiben als Hilfsmittel einsetzen</b>			
4.1. Informationen strukturiert schriftlich wiedergeben	Umsetzung relevanter Vorschriften <b>Fach- und Methodenkompetenz</b> <b>Kognitiver Bereich</b> Engagement zutreffende Rechtsgrundlagen und berufsethische Prinzipien anzuwenden <b>Fach- und Methodenkompetenz</b> <b>Affektiver Aspekt</b>	Befehle, Anweisungen, Verlautbarungen sowie Anleitungen und Beschreibungen schriftlich wiedergeben, zusammenfassen Mitschriften, Exzerpte, Stichwortzettel und Gesprächsnotizen verfassen	
4.2. Relevante Informationen markieren und kommentieren			
4.3. Mitschriften verfassen			

**Kompetenzbereich: SPRACHBEWUSSTSEIN**

Bildungsstandards/ Teilkompetenzen	Personalstrukturmodell Unteroffizier Qualifizierungsstufe 2 (GrpKdt)	Konkretisierung der Fähigkeiten „berufstypische Aktivitäten“	Geeignete Prüfverfahren
<b>4. über Rechtschreibbewusstsein verfügen</b>			
4.1. grundlegende Regeln der Dehnung, Schärfung, des Stammprinzips, der Groß- und Kleinschreibung beherrschen und diese beim Schreiben anwenden können			Lücken- und Volldiktate
4.2. die Rechtschreibung des Gebrauchswortschatzes einschließlich gängiger Fremdwörter beherrschen			Unbekannte Wörter umschreiben
4.3. Arbeitshilfen zur Rechtschreibung einsetzen		mit Wörterbüchern, Grammatiken umgehen	

(Der Anforderungskatalog – Kompetenzen für Einsatz- und Friedensfunktionen im ÖBH 2010 wurde durch die Spalten „Konkretisierung der Fähigkeiten“ und „Geeignete Prüfverfahren“ von mir ergänzt!)

Im Besonderen soll anhand dieses Berufskompetenzmodells gezeigt werden, welche Möglichkeiten sich für die Erstellung von schriftlichen Sprachtests wie der Zulassungsprüfung Deutsch bieten, aber auch welche Entscheidungen aufgrund der großen Auswahl der zur Verfügung stehenden Testformate getroffen werden müssen. Insofern überrascht es nicht, dass aus pragmatischen Überlegungen heraus das Diktat als Prüfungsformat ausgewählt wurde.

## 16.2 Vom Diktat zur Bildschreibung – Erprobung einer Alternative zum Diktat

Unter Berücksichtigung der Untersuchungsergebnisse beim Diktat, soll der Focus bei der Entwicklung eines neuen Prüfmittels in erster Linie auf Eigeninitiative und Kreativität gelegt werden. Dabei soll eine Öffnung der Aufgabenstellung unter Berücksichtigung der kommunikativen Anforderung im Berufsleben vorgenommen werden, um eine möglichst chancengleiche Sprachüberprüfung zu gewährleisten. Darüber hinaus soll aus den Testergebnissen eine möglichst gezielte Sprachförderung ableitbar sein.

Angesichts dieser Überlegungen darf nicht vergessen werden, dass Tests keine begleitenden Maßnahmen darstellen, sondern lediglich als Momentaufnahmen eines Sprachstandes verstanden werden dürfen.

Wenn ein neues Testformat anstatt des Diktats entwickelt werden soll, stehen meiner Ansicht nach zwei zentrale Fragen im Vordergrund:

*Was soll bzw. kann eine Sprachprüfung, wie die schriftliche Zulassungsprüfung Deutsch beim Bundesheer überhaupt leisten? Welche Interessen und Rahmenbedingungen sollten dabei berücksichtigt werden?*

Wie aus den Experteninterviews und den heeresinternen Grundlagen hervorgeht, soll die Zulassungsprüfung Deutsch als Erhebung eines Istzustandes der Sprachbeherrschung begriffen werden. Dabei gilt es festzustellen, ob die sprachliche Leistung der Mindestanforderung an Deutschkenntnissen (8. Schulstufe) entspricht. Insbesondere liegt es im Interesse aller Verantwortlichen (Arbeitsgruppe Prüfmittel), dass die Ergebnisse weniger der Selektion von Teilnehmern, sondern vielmehr als Entscheidungsgrundlage über Förderung oder Nichtförderung dienen. Im Hinblick auf die Förderung soll sich die Sprachfeststellung auf eine möglichst produktive Schreibhandlung der Teilnehmer beziehen.

Laut dem *Anforderungskatalog Kompetenzen für Einsatz- und Friedensfunktionen* (wie in der Tabelle auf S. 125/126 ersichtlich) soll der Teilnehmer nicht nur seine Sprachfähigkeit, sondern auch seine kognitive Leistungsfähigkeit beim Erschließen und Umsetzen von Aufgabenstellungen, abgestimmt auf den militärischen Kontext, unter Beweis stellen.

Mitentscheidend für die Auswahl eines Testformats sind die Rahmenbedingungen bei der Durchführung der Prüfung. Dazu meint Glaboniat:

*„Auch wenn die TN mit Migrationshintergrund bereits sehr hohe Kompetenzen in Deutsch haben, liegt ihr Problem meist im Grad der Automatisierung bzw. der Schnelligkeit. Ihr Nachteil liegt v.a. darin, dass sie sowohl zur Erschließung als auch für die Produktion eines Textes einfach mehr Zeit brauchen würden als TN mit Erstsprache Deutsch. Hier wäre es also empfehlenswert, für ALLE insgesamt immer etwas mehr Bearbeitungszeit vorzugeben, als dies für die TN mit Erstsprache Deutsch notwendig wäre. Eine weitere Frage in diesem Zusammenhang betrifft den Gebrauch von Wörterbüchern (einsprachig, zweisprachig etc.).“<sup>127</sup>*

Wenn man von der Erstellung eines neuen Testverfahrens spricht, so müssen die ökonomischen Bedingungen dabei unbedingt berücksichtigt werden. Im Fall der Zulassungsprüfung handelt es sich um ein Mengengerüst von etwa 20 – 120 Teilnehmern pro Prüfungstermin. Was die Zeitvorgabe für die Auswertung betrifft, sind je nach Termin zwischen 3 und 7 Tagen vorgesehen. Bei den drei vorgesehenen Personen handelt es sich um fachlich geschultes, akademisches Lehrpersonal. In Anbetracht der Tatsache, dass deren vornehmliche Aufgabe nicht die Auswertung dieses Tests ist, dürfen erfahrungsgemäß max. zwei Stunden pro Tag für diese Tätigkeit gerechnet werden.

Das bedeutet, dass im schlechtesten Fall maximal 18 Stunden Zeit für die Bearbeitung von 120 Personen zur Verfügung stehen, also knapp sieben Minuten pro Testblatt. Im Umkehrschluss sind daher all jene Testverfahren zu verwerfen, die mehr Zeit beanspruchen.

---

<sup>127</sup> Ausschnitt aus einem E-Mail von Glaboniat vom 30. 05. 2011

Daraus ergibt sich die Notwendigkeit, ein Testformat zu entwickeln, das zum einen den angeführten Interessen und Rahmenbedingungen weitgehend entspricht und zum anderen auch die ökonomischen Umstände nicht außer Acht lässt. Es geht dabei nicht nur um den Sprachbedarf, sondern auch um die Sprachbedürfnisse in einer beruflichen Zielsituation.

Angelehnt an die förderdiagnostischen Mitteln der FÖRMIG-Gruppe (Auswertungsraster – Sprachhandlung Beschreiben, Bumerang-Verfahren) (siehe Kap. 4.1 – 4.2) habe ich mich für eine freie Schreibaufgabe mit unmittelbarem Berufsbezug entschieden.

Um alle Problembereiche, die bei der Erstellung eines neuen Testformats und dessen Auswertung anfallen entsprechend zu berücksichtigen, sah ich es als notwendig, an einen Anforderungskatalog anhand der Schlussfolgerungen aus den bisherigen Untersuchungsergebnissen zu erstellen. Auf Basis von Konrad Ehlichs „*Anforderungen an Verfahren regelmäßiger Sprachstandsfeststellung*“ habe ich mich auf folgende Anforderungskriterien festgelegt:

- Abbildung berufstypischer Aktivitäten in der Testaufgabe in Anlehnung an das Berufskompetenzmodell
- Niveau der Aufgabenstellung in sprachlicher Hinsicht orientiert sich an den Bildungsstandards der achten Schulstufe
- Niveau der Aufgabenstellung in inhaltlicher Hinsicht (Weltwissen, Sachwissen) orientiert sich an dem Lehrstoff der 5. – 8. Schulstufe
- Niveau der Aufgabenstellung in fachlicher Hinsicht (Militär) orientiert sich an den Ausbildungszielen des Vorbereitungslehrgangs an der Heeresunteroffiziersakademie
- Kommunikative Handlungsfähigkeit soll bei der Aufgabenstellung im Vordergrund stehen
- Ausrichtung der Aufgabenstellung an den aktuellen militärischen Grundsatzdokumenten
- Berücksichtigung von Personen mit nichtdeutscher Muttersprache in der Zeitvorgabe (für alle Testteilnehmer wird mehr Zeit gegeben!)
- Ausgleich von sozialem und schulischem Hintergrund bei der Aufgabenstellung (hoher Bildungshintergrund sollte so wenig Vorteile wie möglich bringen!)
- Testaufgabe soll ohne Verwendung eines Wörterbuchs, insbesondere für Personen mit nichtdeutscher Muttersprache möglich sein
- Erfüllen der Testaufgabe sollte ohne großen materiellen Aufwand gewährleistet sein (Benutzerfreundlichkeit)

- Ökonomische Rahmenbedingungen sollten weitgehend berücksichtigt werden
- Einhaltung der Qualitätskriterien wie Validität, Reliabilität, Objektivität, Interaktivität und Authentizität sowie Praktikabilität
- Sprachwissenschaftliche Fundierung des Auswertungsrasters
- Auswertung im Sinne einer gesamtsprachlichen Leistung <sup>128</sup>

Ähnlich wie bei dem sprachdiagnostischen Verfahren „Bumerang“ soll ein Bildimpuls die Schreibhandlung einleiten. Es geht dabei aber nicht um die Erstellung einer Bauanleitung, sondern um eine Situationsbeschreibung anhand einer photographischen Abbildung. Der Vorteil eines Bildes als Schreibimpuls ist die große Auswahlmöglichkeit an Vorlagen.

Insgesamt erfordert die Bildbeschreibung mehrere Schreibhandlungen. Berücksichtigt wird dabei auch die handschriftliche Kompetenz bzw. die Schreibgeschwindigkeit. Es gibt genügend Zeit für die Aufgabe.

Um die kommunikative Handlungsfähigkeit bzw. die Kreativität des Testteilnehmers entsprechend zu aktivieren, stehen Bilder zur Wahl, die nicht nur genügend Anhaltspunkte (Personen, Fahrzeuge, Gebäude, Landschaften usw.) liefern, sondern auch einen möglichst großen Spielraum an Interpretationsmöglichkeiten zulassen. Aufgrund des großen vorhandenen Bildmaterials ist der Fokus in erster Linie auf militärische Katastrophen- und Friedenseinsätze gerichtet.

Gemäß dem Kompetenzmodell für Berufssoldaten entspricht die Bildbeschreibung etwa einem Beobachtungsprotokoll in Form einer „*Militärischen Meldung*“. Die universell einsetzbare Textsorte „Meldung“ stellt eine berufstypische reale Sprachhandlung dar, die sowohl in schriftlicher als auch in mündlicher Form vorkommt. Soldaten wissen, dass jede Meldung klar, eindeutig und vollständig zu erfolgen hat.

Innerhalb der ersten Erprobungsphase galt es, vorerst ohne Leitpunkte auszukommen. Allerdings stellte sich heraus, dass die Umsetzung der Aufgabenstellung derart unterschiedlich erfolgt, dass eine gerechte bzw. vergleichbare Auswertung kaum möglich schien. Aufgrund dieser Schlussfolgerung verfasste ich drei zentrale Leitfragen anstelle der allgemein formulierten Fragestellung. Bei der Testaufgabe soll das Bild anhand der Leitfragen in Form eines Kurztextes von mindestens 100 Wörtern (etwa 10 Sätze) beschrieben werden.

Wie in der Aufgabenstellung steht, sollen dabei auch, insbesondere bei der Beschreibung von Waffen, Ausrüstung, Gerät und Fahrzeugen, *militärische Fachbegriffe* verwendet werden.

---

<sup>128</sup> vgl. dazu Ehlich (2004), S. 46 – 50.

Insgesamt geht es bei der Bildbeschreibung um die kommunikative Wirksamkeit bei der Bewältigung der Aufgabe. Wesentlich ist, dass sowohl die drei Leitfragen behandelt, als auch die Textlänge eingehalten werden. Ein Mindestausmaß von 100 Wörtern wurde deshalb festgelegt, um von jedem Testteilnehmer ausreichend Textmaterial für die Auswertung zu haben. Für die Bewältigung der Aufgabe stehen 25 Minuten zur Verfügung.

### Verfassen eines Beobachtungsergebnisses anhand eines authentischen Bildes

Freies Schreiben nach Anleitung

Länge des Kurztextes: mindestens 100 Wörter (etwa 10 – 12 Sätze)

Zeit für die Aufgabe: 25 Minuten

#### Aufgabe

Sie sind als Beobachtungsposten eingeteilt und haben den Auftrag, Ihrem Vorgesetzten Ihr Beobachtungsergebnis zu melden.

Die Bildbeschreibung soll insbesondere folgende drei Leitfragen behandeln:

- 1) Welche Personen und Gegenstände (Waffen, Gerät, Ausrüstung) erkennen Sie?**
- 2) Welche für das Beobachtungsergebnis wichtigen Gelände- und Witterungsbegebenheiten (Verkehrswege, Gebäude, Bodenbedeckung usw.) erkennen Sie?**
- 3) Stellen Sie Vermutungen über die Situation an (z. B. Einsatzraum, mögliche Bedrohungen, Konfliktparteien)!**

*Verwenden Sie ganze Sätze und keine Stichworte!*

*Nützen Sie Ihre bisherigen militärischen Erfahrungen und benützen Sie, wenn erforderlich, militärische Fachbegriffe!*

Achten Sie darauf, dass die Beschreibung des Bildes auch ohne Bild nachvollziehbar sein sollte. Ihr Beobachtungsergebnis bildet eine wesentliche Grundlage und Voraussetzung für die Entscheidung Ihres Kommandanten, der die Situation (siehe Bild) nicht kennt!

Beschreiben Sie die auf dem Bild dargestellte Situation möglichst genau und vollständig. Beschreiben Sie **alle erkennbaren Einzelheiten** wie Soldaten, Zivilpersonen, Bewaffnungen, Geräte **sowie für das Beobachtungsergebnis relevante Objekte und Gegenstände**.

*Achten Sie vor allem auf jene Details, die für das Beobachtungsergebnis wichtig sein könnten.*



(Aufgabenblatt – Stand November 2011)

### 16.3 Gewichtung, Bewertung und Punktevergabe mittels Auswertungsraster

Die Auswertung der Schreibaufgabe erfolgt auf Basis einer Diagnose mit dem Ziel, Fördermaßnahmen zu planen. Die Kriterien des Auswertungsinstruments orientieren sich an den Anforderungen, der Funktion und dem Zweck der Bildbeschreibung.

Die Kompetenzstufen richten sich nach dem Niveau der 8. Schulstufe.

Um auf den unterschiedlichen Ausprägungsgrad von Kompetenzen und auf mögliche Abweichung Rücksicht zu nehmen sind Bewertungsstufen vorgesehen.

Der Auswertungsraster ist in Bewertungskategorien gegliedert, mit denen der Einsatz und die Organisation von sprachlichen Mitteln gewichtet und bewertet werden sollen. Anders als bei realen militärischen Meldungen wurde auf Formalkriterien bewusst verzichtet.

Im Zuge eines Workshops bei der „Grazer Tagung 2011“<sup>129</sup> wurden die Kriterien für den Auswertungsraster ausgewählt, gewichtet und an realen Texten ausprobiert. Beim „Benchmarking“<sup>130</sup> wurde festgestellt, dass es Überschneidungen bei Kriterien wie Lexik, Syntax und Morphologie, aber auch Unklarheiten bei der Beschreibung der Kompetenzstufen gibt.

Dabei wurde insbesondere die Problematik hinsichtlich der Gewichtung der Kriterien erkannt und diskutiert. Die Konsequenzen daraus waren die Streichung der Kategorie Textkohärenz und die Zusammenfassung von Syntax und Morphologie.

Das Endprodukt des Auswertungsrasters enthält folgende Kategorien:

- *Aufgabenbewältigung*  
Dabei geht es vor allem um die inhaltliche Erfüllung der Schreibaufgabe. Bei Nicht oder nicht ausreichender Behandlung einzelner Leitpunkte gibt es Abzugspunkte. Zusätzlich fließt die Textlänge in die Beurteilung ein. Bei Nichterfüllen der 100 Wörter gibt es ebenfalls Abzugspunkte.
- *Ausdrucksfähigkeit*  
Bei der Kategorie geht es sowohl um die allgemeine, der Textsorte angemessene Wortwahl, als auch um die Verwendung von Fachausdrücken. Unangemessene bzw. falsche Wortwahl, aber auch fehlende Fachausdrücke (Gewehr statt Sturmgewehr) wirken sich auf die Niveaustufe aus.
- *Normorientierung 1*  
Die Kategorie Normorientierung 1 beinhaltet sämtliche Bereiche der Grammatik wie Syntax oder Morphologie. Es geht dabei in erster Linie nicht um eine korrekte Normeinhaltung, sondern um die Eindeutigkeit und Verständlichkeit der Aussagen.

---

<sup>129</sup> 14. Grazer Tagung DaF/DaZ – 17./18. Juni 2011 mit dem Thema „Sprachstandsfeststellung“

<sup>130</sup> Benchmarking bezeichnet die vergleichende Analyse von Ergebnissen oder Prozessen mit einem festgelegten Bezugswert.

- *Normorientierung 2*  
Bei Normorientierung 2 stehen Orthographiefehler, die das Verständnis des Textes beeinträchtigen im Vordergrund. Wenige Rechtschreibfehler werden auch in der höchsten Niveaustufe toleriert (Es geht hier um das Niveau der 8. Schulstufe!).

Bei der Punktevergabe soll ein ausgewogenes Verhältnis zwischen Ausdrucksfähigkeit und Normorientierung herrschen. Die höchste Kompetenzstufe innerhalb der Kriterien soll jeweils das Sollniveau abbilden (8. Schulstufe). Um bei großen Mängeln innerhalb einer Kategorie unmittelbar mit Fördermaßnahmen reagieren zu können (z. B. Lexik) wurde eine Spalte mit „nicht bestanden“ hinzugefügt. Ohne den Zusatzpunkt in der Kategorie Aufgabenbewältigung kann ein Kandidat insgesamt 11 Punkte erreichen.

Die Förderschwelle wurde gemäß folgender Annahme festgelegt: Wenn ein Kandidat in Aufgabenbewältigung 0, in Lexik 3, in Orthographie und Grammatik jeweils 2 Punkte erreicht, dann gilt die Prüfung als bestanden. Förderbedarf ist damit für diejenigen Personen vorgesehen, die weniger als 7 Punkte erreicht bzw. in einer Kategorie nicht bestanden haben. Mitgedacht dabei werden soll, dass bei der Punktevergabe in Orthographie in erster Linie die Verständlichkeit der Aussage bewertet wird.

### Auswertungswerkzeug für die „Bildbeschreibung“

	nicht bestanden	- 2 Punkte	- 1 Punkt	0 Punkte	1 Punkt
<b>Aufgabenbewältigung</b> inhaltliche Themenbezogenheit und Angemessenheit; Behandlung der drei Leitpunkte; Einhaltung der Textlänge	Thema <i>verfehlt</i> ; Aufgabe inhaltlich unvollständig erfüllt;  kein Leitpunkt inhaltlich und über das Zitierende hinausgehend behandelt  -59 Wörter	Aufgabe erfolgt <i>unvollständig</i> , an vielen Stellen unklare Vorgangsweise;  1 Leitpunkt inhaltlich und über das Zitierende hinausgehend behandelt  und/oder 60 – 79 Wörter	Aufgabe <i>inhaltlich weitgehend</i> erfüllt; Beschreibung erfolgt wirklichkeitsentsprechend, aber noch lückenhaft;  2 Leitpunkte inhaltlich und über das Zitierende hinausgehend behandelt  und/oder 80 – 99 Wörter	Aufgabe <i>inhaltlich zur Gänze</i> erfüllt; Beschreibung erfolgt wirklichkeitsentsprechend und vollständig;  alle Leitpunkte inhaltlich und über das Zitierende hinausgehend behandelt  und/oder 100 – 119 Wörter	Aufgabe inhaltlich <i>über das erwartete Maß</i> erfüllt; Beschreibung erfolgt wirklichkeitsentsprechend und vollständig; situative Einbettung der Personen und Gegenstände; alle Leitpunkte inhaltlich und über das Zitierende hinausgehend behandelt  und/oder 120 Wörter und mehr
	nicht bestanden	1 – 2 Punkte	3 – 4 Punkte	5 Punkte	
<b>Ausdrucksfähigkeit</b> präzise Formulierung, Sachzusammenhang, Wortwahl (Lexik), Stil, Fachsprache	<i>umgangssprachliche</i> Ausdrücke; Wortwahl verständnisstörend und nicht angemessen; Formulierung erfolgt nicht in vollständigen Sätzen Zeigewörter, Näherungsbegriffe <i>Fachbegriffe werden nicht oder falsch verwendet</i>	meist <i>verständliche Alltagsprache</i> ; Wortwahl manchmal unpassend; meist allgemeine Begriffe und annähernde Bezeichnungen;  <i>Fachbegriffe werden nur teilweise richtig verwendet</i>	<i>einfacher</i> , geläufiger Wortschatz, Wortwahl passend und angemessen; treffende Umschreibungen  <i>Fachbegriffe werden großteils richtig verwendet</i>	<i>differenzierter</i> Wortschatz; variantenreiche, sachangemessene Wortwahl,  richtiger und gezielter Einsatz von <i>Fachbegriffen</i>	
	nicht bestanden	1 Punkt	2 Punkte	3 Punkte	
<b>Normorientierung 1 sprachliche Richtigkeit</b> Grammatik (Syntax, Morphologie)	Übertragung aus der <i>Mündlichkeit</i> ; Formulierung erfolgt nicht in korrekten Sätzen; Vorkommen von Äußerungen ohne Verb;  viele grammatische Fehler, die das Verständnis des Textes erheblich beeinträchtigen;  wichtige Aussagen sind unverständlich	<i>einfache</i> Sätze mit Unsicherheiten in Wortstellungen und Wortformen; Unsicherheiten bei der Flexion von Verben  häufig grammatische Fehler, die das Verständnis des Textes teilweise beeinträchtigen; wichtige Aussagen sind großteils verständlich	<i>erweiterte</i> Sätze mit einfachen Attributen und Relativsätzen; meist schematische Verbindung von Sätzen; einfache Nebensätze  wenige grammatische Fehler, die das Verständnis des Textes aber nicht beeinträchtigen; alle Aussagen sind eindeutig und verständlich	<i>komplexe</i> Attribute und adverbiale Bestimmungen; <i>spezifische</i> Nebensätze; gezielter Einsatz des Konjunktivs  nahezu fehlerlose Grammatik, morphologische und syntaktisch großteils korrekt	
	nicht bestanden	1 Punkt	2 Punkte	3 Punkte	
<b>Normorientierung 2 sprachliche Richtigkeit</b> Orthographie	viele Rechtschreibfehler, die das Verständnis des Textes erheblich beeinträchtigen	durchgängig Rechtschreibfehler, die das Verständnis des Textes zum Teil beeinträchtigen	einige Rechtschreibfehler, die das Verständnis des Textes aber nicht beeinträchtigen	nahezu fehlerlose Rechtschreibung,	

(Auswertungsraster – Stand: November 2011)

Nachdem sich bei einer Erprobungsphase mit zwei Gruppen von jeweils zehn institutionsinternen Grundwehrdienern einige Probleme bei der Erfüllung der Testaufgabe ergaben, wurden einige Korrekturen im Aufgabenblatt vorgenommen (Einführung der Leitpunkte) bzw. die Beschreibung der Kompetenzstufen geringfügig verändert. Aus dem Umstand heraus, dass nahezu die Hälfte aller Probestestteilnehmer weniger als 100 Wörter lieferte, ergab sich die Notwendigkeit, neben der Behandlung der Leitpunkte auch die Textlänge in die Kategorie Aufgabenbewältigung einzubeziehen. Im Zuge einer Arbeitsgruppensitzung wurde das Testformat einer fachlichen Diskussion unterzogen, bei der alle in Betracht kommenden Vor- und Nachteile abgewogen wurden. Schließlich wurde das Testformat NEU für eine weitere Evaluierung, insbesondere für die Vergleichsstudie freigegeben. Dabei galt es auch, die *Auswertungsreliabilität* anhand der Diagnosewerkzeuge zu prüfen, um unterschiedliche Meinungen bei der Bewertung so weit wie möglich zu objektivieren. Mittels einer repräsentativen Auswahl der Stichprobe 3 (Vergleichsstudie) wurden von allen drei Auswertern anhand des Diagnoserasters Auswertungen vorgenommen und besprochen, damit eine Übereinstimmung bei der Fehlermarkierung, insbesondere bei der Punktevergabe gewährleistet werden konnte.

## **17 Vergleichsstudie – Einsatz des neuen Testformats**

### **17.1 Untersuchungsmethodik**

Aus forschungsrelevanter Sicht können Testformate nur dann verglichen werden, wenn ausreichend messbare Vergleichsebenen vorhanden sind. Das erscheint beim Diktat und bei der Bildbeschreibung auf den ersten Blick problematisch, weil es sich sowohl aus testtheoretischer als auch aus traditioneller Hinsicht um völlig unterschiedliche Aufgabenstellungen handelt. Während es sich beim Diktat um eine vielfach erprobte und standardisierte Schreibaufgabe handelt, geht es bei der Bildbeschreibung um ein völlig neu konzipiertes Prüfmittel, dessen Aufgabenstellung und Schwierigkeitsgrad noch nicht empirisch abgesichert ist. Darüber hinaus werden die Ergebnisse beim Diktat und bei der Bildbeschreibung auf Basis völlig unterschiedlicher Auswertungskriterien bestimmt.

Das bedeutet, dass ein direkter Vergleich von Gesamt- und Einzelergebnissen unter Einbeziehung unterschiedlicher Aspekte wie Alter oder Geschlecht aus meiner Sicht empirisch kaum haltbar wäre. Aufgrund dieser Überlegungen beschränke ich mich bei meiner Vergleichsstudie auf die Ergebnisse einer Fragebogenerhebung bzw. auf die Meinungen zweier Experten, die ich zum Diktat und zur Bildgeschichte kurz befragt habe.

Wenngleich beim Fragebogen der Vergleichsstudie im Gegensatz zum Fragebogen Diktat mehr Feststellungen zu bewerten sind (30) benötigten Befragungsteilnehmer nicht mehr Zeit, da es sich vielfach um Vergleichsfragen handelt. Zu berücksichtigen ist allerdings, dass die Teilnehmer zuvor zwei Testaufgaben statt einer zu erledigen hatten.

## **17.2 Beschreibung der Stichprobe**

Ganz anders dagegen verlief die Auswahl der Versuchspersonen bei der Vergleichsstudie, die aus zeitlichen und prüfungstechnischen Gründen im Rahmen eines Förderseminars stattfand. Es handelte sich dabei um keine wirkliche Prüfungssituation, sondern um eine im Bezug zur realen Prüfungssituation stehende Probetestung, die als Lernfortschrittsprüfung verstanden werden sollte. Um auch direkte Vergleiche zu der zweiten Stichprobe ziehen zu können, habe ich die gleiche Testvariante (Diktat) wie bei der zweiten Stichprobe für die Studie herangezogen.

Da es sich dabei um TeilnehmerInnen eines Förderseminars handelte, standen ausschließlich jene Personen zur Verfügung, die aufgrund der Einstufungsprüfung Deutsch beim Heerespersonalamt einen Förderbedarf vorzuweisen hatten.

Angesichts der Tatsache, dass bei Prüfungen im Zuge von Förderseminaren annähernd gleiche Bedingungen vorherrschen wie am Ende der Vorbereitungslehrgänge in Ebelsberg und Enns, lassen sich auch untersuchungsrelevante Vergleiche zu den beiden vorangegangenen Stichproben anstellen.

In Anbetracht der wesentlich komplexeren Problemstellung hat sich die Wahl der Stichprobe auch insofern bewährt, als dass das Ausfüllen des Fragebogens nicht unter Zeitdruck erfolgen musste (15 Minuten).

### Ort/Stärke - Auswahl der gültigen Testpersonen - Gründe

Die dritte Stichprobe fand am 15 07 2011 um 1000 Uhr in der Andreas-Hofer-Kaserne in Absam/Tirol statt. Beim Untersuchungsort handelte es im Gegensatz zu Ebelsberg und Enns um keine schulische Ausbildungseinrichtung, sondern um den Sitz der Gebirgsbrigade des Österreichischen Bundesheeres. Der Zeitpunkt der Vergleichsstudie wurde deshalb so spät gewählt, weil zuvor das neue Testformat erst entwickelt und erprobt werden musste.

Die Gesamtstichprobe umfasste 25 Versuchspersonen, wobei 2 davon ausgeschlossen wurden. Gründe dafür waren ausschließlich mangelhaft ausgefüllte Fragebögen in

Verbindung mit einer unterdurchschnittlichen Ausfüllzeit (unter 5 Minuten). Nach Abzug verblieben bei der dritten Stichprobe 23 gültige Versuchsteilnehmer.

### Testvarianten

Diktatvariante: „Wenn Schnee zur Gefahr wird“ – 163 Wörter, wie in Stichprobe 1 und 2.

Bildbeschreibung – Variante: „Österreichische Patrouille im Tschad-Einsatz“, wo mindestens 100 Wörter zu drei Leitfragen geschrieben werden mussten. Es durfte kein Wörterbuch verwendet werden.

### Bearbeitungsdauer

Im Gegensatz zum Diktat, wo der zeitliche Ablauf durch das Abspielen der Tonbandaufnahme strikt vorgegeben ist, wird bei der Bildbeschreibung eine maximale Bearbeitungszeit von 25 Minuten vorgegeben. Geht man davon aus, dass Testteilnehmer für die Ausarbeitung auch weniger Zeit in Anspruch nehmen, stellt die Bearbeitungsdauer einen forschungsrelevanten Faktor dar. Insbesondere dann, wenn daraus Rückschlüsse auf das Testergebnis und die Testperson gezogen werden können.

Bei den 23 gültigen Versuchsteilnehmern betrug die durchschnittliche Bearbeitungsdauer etwa 18 Minuten. Während bei der Spitzengruppe Werte von 11 bis 15 Minuten gemessen wurden, schöpften die beiden Personen mit Migrationshintergrund die vollen 25 Minuten aus.

### Alter

Die vorliegende Stichprobe der Vergleichsstudie setzt sich aus folgenden Altersgruppen zusammen:

Altersgruppe	Häufigkeit	Prozent (gerundet)
unter 20-Jährige	-	-
20 – 25-Jährige	20	86,5
25 – 30-Jährige	2	9,0
über 30-Jährige	1	4,5
Gesamt	23	100

Das Durchschnittsalter liegt bei 24 Jahren, was der angestrebten Altersschnittmenge entspricht. Ähnlich wie bei der ersten Stichprobe dominiert die Altersgruppe der 20 – 25-Jährigen.

### Geschlecht

Bei der dritten Stichprobe nahm eine weibliche Testperson teil. Das entspricht den Werten der ersten und zweiten Stichprobe.

### Muttersprachen – Deutsch als wievielte Sprache?

Bei den erworbenen bzw. erlernten Sprachen der Testpersonen zeigen sich ähnliche Werte wie bei den ersten beiden Stichproben. Allerdings handelt sich nicht ausschließlich um Bosnisch, sondern auch um Polnisch als Erstsprache.

Sprachen	Häufigkeit	Prozent (gerundet)
Deutsch als Muttersprache	21	91,0
Deutsch als Zweitsprache	2	9,0
Polnisch	1	
BKS	1	
Gesamt	23	100

### Erlernete Fremdsprachen

Die Tendenz, dass Englisch zumeist als erste Fremdsprache gelernt wird, bestätigt sich auch in der Stichprobe der Vergleichsstudie.

Fremdsprachen	Häufigkeit	Prozent (gerundet)
keine Fremdsprache	-	-
eine Fremdsprache Englisch	21	91,0
zwei Fremdsprachen- Englisch	2	9,0
+ Französisch	1	
+ Spanisch	1	
Gesamt	23	100

### Schulabschluss

Laut Angaben der Versuchspersonen ergaben sich folgende Werte:

positiver Schulabschluss	Häufigkeit	Prozent (gerundet)
Hauptschulabschluss	21	91,0
Unterstufe Gymnasium	-	
Matura	-	
andere Schulabschlüsse	2	9,0
Gesamt	23	100

Das Kriterium Schulabschluss zeigt, dass die Stichprobe sich überwiegend aus Personen mit Hauptschulabschluss zusammensetzte und dass keine Person mit Matura an der Stichprobe

teilnahm. Da es sich hier größtenteils um Teilnehmer eines Förderseminars handelt, überraschen diese Werte nicht.

Fasst man die vorhandenen Personaldaten zusammen, kann man meiner Ansicht nach trotz einiger Abweichungen durchaus von repräsentativen Gruppen sprechen. Dass bei der Auswahl von Stichproben, insbesondere bei Untersuchungen von Sprachtests, Abweichungen vom Idealdurchschnitt kaum zu vermeiden sind, steht für mich außer Zweifel.

### 17.3 Testergebnisse der Vergleichsstudie (Diktat, Bildbeschreibung)

Trotz der erwähnten empirischen Bedenken will ich aus Untersuchungsinteresse zumindest einen groben Vergleich der Fehlerhäufigkeit im Bereich der Orthographie und Grammatik anstellen. Mit zu berücksichtigen dabei ist, dass es sich beim Diktat um ein Textausmaß von 163 Wörtern handelt und bei der Bildbeschreibung durchschnittlich etwa 100 – 120 Wörter bewertet werden. Allerdings zeigen sich höchst signifikante Auffälligkeiten im Bereich von Orthographie und Grammatik, die weder mit der Bewertung, noch mit der Länge des Textes in Zusammenhang stehen können. Es geht dabei in erster Linie um die Fehleranfälligkeit beim Diktat und in der Bildbeschreibung. Während die Teilnehmer der Spitzengruppe beim Diktat vergleichsweise gleich viele Orthographie- und Grammatikfehler machen wie bei der Bildbeschreibung, präsentiert sich das Ergebnis bei den schwächeren Teilnehmern in umgekehrter Richtung. Die meisten Teilnehmer dieser Gruppe machen bei Berücksichtigung der Textlänge bei der Bildbeschreibung weniger Fehler. Dieser Umstand fällt besonders bei der Klein- und Großschreibung auf. Interessant ist auch das Phänomen, dass Wörter, die beim Diktat fehlerhaft reproduziert wurden, bei der Bildbeschreibung plötzlich richtig geschrieben werden.

Noch größere Abweichungen als bei den Muttersprachlern zeigen sich bei den Teilnehmern mit nichtdeutscher Muttersprache:

Absam 15 07 2011	Diktat		Bildbeschreibung		Abweichung Prozentwerte
	Häufigkeit	Prozentwert zu 163 Wörter	Häufigkeit	Prozentwert zur Textlänge	
Teilnehmer mit deutscher Muttersprache	24	14,72%	15	11,90%	2,82%
Teilnehmer mit nichtdeutscher Muttersprache	25	15,34%	5	6,25%	9,09%

Von zusätzlichem Interesse ist hier auch der Faktor Zeit. Während Muttersprachler für die Bildbeschreibung im Durchschnitt 15 Minuten brauchen, schöpfen Personen mit nichtdeutscher Muttersprache in den meisten Fällen die vollen 25 Minuten aus.

Beim Vergleich der Ergebnisse entsteht für mich der Eindruck, dass hier sowohl die Testaufgabe als auch die Testsituation, insbesondere bei den Migranten, eine große Rolle spielen. Nicht zu vergessen dabei ist, dass die Testteilnehmer bei der Bildbeschreibung meist nur das niederschreiben, was sie bereits können. Diese deutlichen Leistungsunterschiede deuten auf ein Phänomen hin, das wie die Praxis zeigt, zumeist im Bereich Prüfen und Testen zu beobachten ist. Desto normorientierter eine Prüfung angelegt ist, desto mehr Lerner mit Migrationshintergrund scheitern an der Prüfung.

#### **17.4 Fragebogenergebnisse aus der Vergleichsstudie**

Um die Eignung des in Erprobung befindlichen neuen Testformats zu prüfen, gilt es, eine Auswahl an zukünftigen Bedarfsträgern zu befragen. Bei den Ergebnissen der Fragebogenerhebung steht daher, ähnlich wie bei der Evaluierung des aktuellen Testformats (Diktat), die Einschätzung und Beurteilung im Vordergrund. Gegenstand der Erhebung ist die Gegenüberstellung zweier Testformate, wo das Abschneiden der beiden Testformate bei den Bedarfsträgern dargestellt werden soll.

Wie schon bei der Stichprobe 1 und 2 stellt der befragte Teilnehmer anhand einer Ratingskala fest, ob Aussagen zutreffen, oder nicht. Im Unterschied zu den ersten beiden Stichproben beschränkte ich mich bei der Fragebogenerstellung auf jene sechs Themenblöcke, die eine Gegenüberstellung beider Testformate zulassen.

- A) Testdesign und Testlänge
  - B) Niveau, Qualität und Konsistenz der Testaufgaben
  - C) Störende Aspekte/Testangst
  - D) Sprachverwendung beim Bundesheer
  - E) Testaktivität als nützliche Ausbildungszeit
  - F) Persönlicher Eindruck
- Fragen zur Person

Die graphische Darstellung der Daten erfolgt mittels Balkendiagrammen auf Basis einer Datenmatrix-Tabelle, die sowohl die befragten Personen als auch die Variablen (z. B. 1-4) repräsentiert. Bei der Darstellung habe ich statistisch auffällige und untersuchungsrelevante Ergebnisse ausgewählt. Die Untersuchungsdaten über die beiden Testformate werden im direkten Vergleich gezeigt.

Die Darstellung der Auswertungen erfolgt auf Basis der bereits erprobten Ratingskala.

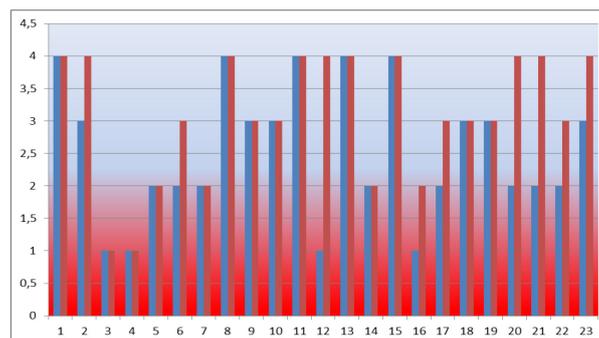
trifft zu	trifft überwiegend zu	trifft eher nicht zu	trifft nicht zu
1	2	3	4

Die blauen Säulen stellen das Testformat Diktat, die roten die Bildbeschreibung dar.

Mit einer roten Hintergrundfarbe wurde jener Bereich gefüllt, wo aufgrund der Feststellungen der Befragten negative Auswirkungen auf die Qualität des Testformats zu erwarten sind.

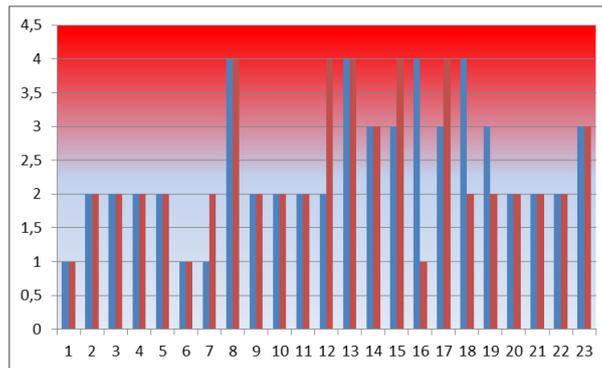
Sowohl bei der Beurteilung des Testdesigns und der Testlänge als auch bei beim Themenblock „Niveaus und der Qualität der Testaufgabe“ zeigen sich keine signifikanten Unterschiede. Beide Testformate werden vom Großteil der Befragten als ansprechend und niveaugerecht eingestuft.

Auffälligkeiten zeigen sich dagegen beim Thema „störende Aspekte/Testangst“.



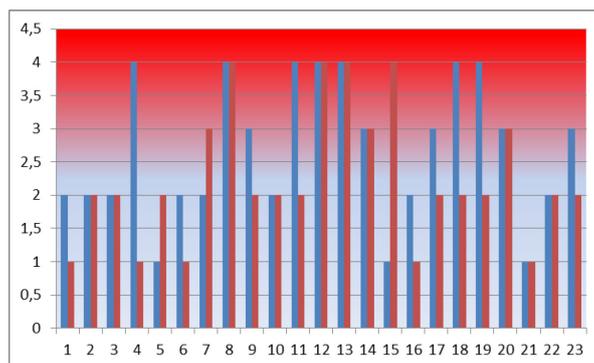
Mehr als die Hälfte der Teilnehmer stellen fest, dass das Diktat Angst auslösen kann. Bei der Bildbeschreibung hingegen gibt es nur sechs Teilnehmer, die dieser Aussage zustimmen. Der Großteil der Befragten sieht in der Bildbeschreibung keine störenden Aspekte.

Wie schon bei den beiden ersten Stichproben schneidet das Diktat beim Thema „Sprachverwendung beim Bundesheer“ überraschend gut ab.



Während 11 Befragte das Diktat als berufsrelevant einstufen, sind es bei der Bildbeschreibung 17 Befragte, also deutlich mehr als beim Diktat. Interessant ist hier, dass 14 Befragte das Diktat und die Bildbeschreibung gleich einschätzen.

Ganz andere, unterschiedlichere Ergebnisse zeigen sich insbesondere beim Thema „*Testaktivität als nützliche Ausbildungszeit*“.



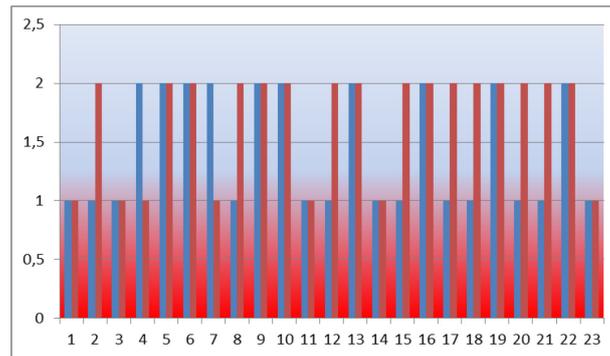
Die Bildbeschreibung schneidet kaum besser ab als das Diktat. Insgesamt sind deutliche Korrelationen bei den Befragungsergebnissen festzustellen. Ähnlich wie bei den Ergebnissen zuvor, schätzen 16 der Teilnehmer das Diktat gleich wie die Bildbeschreibung ein.

Wenn man einen Vergleich zwischen zwei Testformaten anstellen will, so ist der „*Persönliche Eindruck*“ des Befragten ein wesentlicher Aspekt zur Beurteilung.

Anders als bei den Themen zuvor entscheiden hier die Befragungsteilnehmer mit ja oder nein. Die Ziffer 1 steht für „*ja*“, und 2 steht für „*nein*“ im Balkendiagramm.

Besonders zu erwähnen ist dabei der fünfte Punkt der Frage 12 des Fragebogens:

*„Kann mangelnde Übung mit der Schreibschrift Ihrer Einschätzung nach beim Niederschreiben des Textes (Diktats) bzw. der Bildbeschreibung zu Schwierigkeiten führen?“*



Während 13 von 23 Befragten der Meinung sind, dass diese Feststellung auf das Diktat zutrifft, gibt es lediglich 7 Befragte, die auch bei der Bildbeschreibung Probleme aufgrund mangelnder Übung befürchten. Mitgedacht werden sollte, dass das Diktat 163 Wörter, also wesentlich mehr Wörter als die Bildbeschreibung umfasst und unter Zeitdruck geschrieben wird.

## 18 Interpretation der Vergleichsstudie

Mit der Vergleichsstudie habe ich einige Erwartungen verknüpft, die sich nur zum Teil erfüllt haben. Dass der Vergleich von zwei völlig unterschiedlichen Testformaten mittels einer Fragebogenuntersuchung nicht unproblematisch sein wird, stand für mich außer Zweifel. Dennoch lieferte die Vergleichsstudie aussagekräftige Daten zur Beantwortung offener Fragen aus den Stichproben von Ebelsberg und Enns.

Beim Vergleich der Ergebnisse entsteht für mich der Eindruck, dass hier sowohl die Testaufgabe als auch die Testsituation, insbesondere bei den Migranten, eine große Rolle spielen. Nicht zu vergessen dabei ist, dass die Testteilnehmer bei der Bildbeschreibung meist nur das niederschreiben, was sie bereits können. Diese deutlichen Leistungsunterschiede deuten auf ein Phänomen hin, das wie die Praxis zeigt, zumeist im Bereich Prüfen und Testen zu beobachten ist. Desto normorientierter eine Prüfung angelegt ist, desto mehr Lerner mit Migrationshintergrund scheitern an der Prüfung. Inwieweit die kognitive Leistung bei beiden Testformaten eine Rolle spielt, wurde nicht untersucht.

Beim Thema „*Störende Aspekte und Testangst*“ zeigt sich, dass die Einschätzungen bei der Bildbeschreibung im Vergleich zum Diktat deutlich abweichen. Während beim Diktat, so wie schon bei den Stichproben 1 und 2, mehr als die Hälfte der Teilnehmer Testangst wahrnehmen, sind es bei der Bildbeschreibung deutlich weniger.

Das bedeutet, dass sich aufgrund des Datenvergleiches die Annahme bestätigt, dass das Diktat eher Testangst auslöst.

Angesichts der Tatsache, dass das Diktat beim Thema „*Sprachverwendung beim Bundesheer*“ bereits bei den ersten beiden Stichproben gut abschnitt, erwartete ich eine noch größere Zustimmung bei der Bildbeschreibung. Wie sich zeigt, wurde die Bildbeschreibung aufgrund ihrer berufsspezifischeren Ausrichtung wesentlich besser eingestuft als das Diktat. Wenn auch nicht alle diese Einschätzung teilen, scheint der unmittelbare Berufsbezug in der Aufgabenstellung doch Auswirkungen auf die Bedarfsträger zu haben.

Unerwartetes zeigen die Daten beim Thema „*Testaktivität als nützliche Ausbildungszeit*“. Aus der Vergleichserhebung geht hervor, dass die produktivere Aufgabenstellung kaum besser abschneidet als das Diktat. Gerade bei diesem Thema hätte ich mir weit mehr Stimmen für die Bildbeschreibung erwartet. So gesehen, scheint die Sinnhaftigkeit und Nützlichkeit einer Testaufgabe keinen hohen Stellenwert bei den Stichprobenteilnehmern einzunehmen.

Hinsichtlich der Aussage bezüglich der *mangelnden Übung mit der Schreibschrift* zeigt sich im Gegensatz zur Testaktivität ein eindeutiges Bild. Durch den geringen Zeitdruck sind kaum Schwierigkeiten beim Schreiben zu erwarten.

## **19 Schlussfolgerungen und Auswirkungen auf die DaM- und DaZ-Didaktik**

Bei der Vergleichsstudie wurde ein neues Testformat vorgestellt, das sowohl den Ansprüchen einer Zweitsprachendidaktik gerecht werden als auch die sprachlichen Kompetenzen von einsprachig aufwachsenden MuttersprachlerInnen messen sollte.

Um dem Anspruch der Testteilnehmer auf faire Bewertungen ihrer Leistungen gerecht zu werden, wurde bei der Normorientierung der Grad der Verständlichkeit der Aussagen (die Kommunikation beeinträchtigen) sowohl in Grammatik als auch in Orthographie als Hauptkriterium angenommen. Darüber hinaus wurde unter Rücksichtnahme auf Personen mit nichtdeutscher Muttersprache auch der Zeitrahmen für die Erfüllung der Aufgabe um fünf Minuten erweitert.

In diesem Zusammenhang meint Drorit Lengyel:

*„Kein Verfahren ist wirklich für beide Gruppen (DaM, DaZ) in allen Facetten geeignet.“*

*Gerade bei Testverfahren, die den psychometrischen Gütekriterien und testtheoretischen Ansprüchen entsprechen, steht dann häufig im Manual: Die Ergebnisse müssen bei Personen mit Migrationshintergrund sensibel interpretiert werden, da sie in der Stichprobe nur unzureichend berücksichtigt wurden“.<sup>131</sup>*

Auch wenn mit dieser Untersuchung die Grundlage für den Einsatz des neuen Testformats gelegt wurde, bleibt abzuwarten, ob dieses Testformat genügend Aussagekraft über die Schreibfähigkeit im militärischen Kontext hat. Schriftliche Textprodukte sind meist das Ergebnis sehr komplexer Spracherwerbsprozesse, die zumeist nur eine eingeschränkte Überprüfung zulassen. Sie lassen sich schon gar nicht, wie beim Diktat gerne interpretiert, einer Normkategorie wie der Orthographie zuordnen.

Ob in Zukunft computergenerierte Testverfahren das Schreiben mit der Hand gänzlich ersetzen werden, bleibt offen.

Trotz dieses eher ernüchternden Resümees werden die Auswirkungen des Testformats NEU bei der Fördersystematik Deutsch deutlich spürbar sein.

Statt „*teaching to the test*“ werden individuelle Lernwege von Menschen mit nichtdeutscher Muttersprache, aber auch von Menschen mit unterschiedlichem Bildungsniveau ins Zentrum der Deutschdidaktik innerhalb des ÖBH rücken. Demzufolge können Förderpläne individuell abgestimmt und Lernfortschritte dokumentiert werden. Ein wiederholtes Besuchen des Förderseminars wird daher zum Ausnahmefall, Einsparungen bei den Ausbildungskosten sind zu erwarten.

Was bleibt ist die Erkenntnis, dass die sinnvolle Funktion von Prüfungen, die dem Menschen das Erreichen von Zielen und das Vorhandensein von sprachlicher Kompetenz bestätigen, dann verloren geht, wenn die Prüfung und deren Bewertung sich lediglich an der Oberfläche der Schriftsprache orientieren.

---

<sup>131</sup> Ausschnitt aus einem E-Mail von Drorit Lenyel vom 16.03.2011.

## Literaturverzeichnis

---

Albert, Ruth/Marx, Nicole [Hrsg.]: Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht. Tübingen: Narr 2010.

Augst, Gerhard (2004): Orthografie/Orthography. In: Ammon Ulrich et al. [Hrsg.]: Soziolinguistik: Ein internationales Handbuch zur Wissenschaft und Sprache und Gesellschaft. 2., vollständig neu bearbeitete und erweiterte Auflage. Berlin/New York: de Gruyter S. 646 - 653.

Bolton, Sybille (2005): Probleme der Bewertung bei den Produktiven Fertigkeiten in den Deutschprüfungen des Goethe-Instituts. In: Kühn, Ingrid / Lehker, Marianne / Timmermann, Waltraud [Hrsg.]: Sprachtests in der Diskussion. Reihe: Wittenberger Beiträge zur deutschen Sprache und Kultur - Band 4. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Lang S. 53 – 59.

Bachmann, L. F., Palmer, A. S. (1996): Language testing in practice; Designing and developing useful language tests (Oxford Applied Linguistics). Oxford: Oxford University Press.

Chapelle, C. A. (2001): Computer applications in second language acquisition. Foundations for teaching, testing and research (Cambridge Applied Linguistics). Cambridge: Cambridge University Press.

Dannerer, Monika (2008): „Beschreibungsmöglichkeiten der Fach- und Berufskommunikation im Deutschen“. In: ÖDaF-Mitteilungen, Heft 1/2008, S. 22 - 36.

Demme, Silke (2007): „... wo Fehler sind, da ist auch Erfahrung“ (Tschechov) – auch beim Lehren und Lernen von Fremdsprachen. In: Eßer/Krumm [Hrsg.]: Bausteine für Babylon: Sprache, Kultur, Unterricht ... Festschrift zum 60. Geburtstag von Hans Barkowski. Iudicium, München. S. 206 – 216.

Ehlich, Konrad (2005): eine Expertise zu „Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuellen Sprachförderung von Kindern mit und ohne Migrationshintergrund“. In: Gogolin, Ingrid u. a. [Hrsg.]: Sprachdiagnostik von Kindern und Jugendlichen mit Migrationshintergrund. Dokumentation einer Fachtagung am 14. Juli 2004 in Hamburg. Münster: Waxmann S. 33 – 50.

Glaboniat, Manuela (1998): Kommunikatives Testen im Bereich Deutsch als Fremdsprache. Eine Untersuchung am Beispiel des Österreichischen Sprachdiploms. Innsbruck & Wien: Studienverlag.

Glaboniat, Manuela (2006): Sprachkenntnisse beschreibbar, vergleichbar und messbar gemacht!? Über den Nutzen und die Grenzen der Europaratsskalen im Bereich des Prüfens und Zertifizierens. In: Krumm Hans-Jürgen, Portmann-Tselikas Paus [Hrsg.]: Begegnungssprache Deutsch - Motivation, Herausforderung, Perspektiven, Innsbruck-Wien. Bozen: Studienverlag S. 81 – 94.

Glaboniat, Manuela (2009): „Wie viel Standardisierung braucht ein Test?“ Vom Nutzen der Standardisierungstendenzen im Bereich des Sprachenlernens. In: ÖDaF-Mitteilungen, Heft 2/2009, S. 41 - 52.

Kecker, Gabriele (2009): Validität und Validierung von Sprachprüfungen. In: Berndt Annette / Kleppin Karin [Hrsg.]: Sprachlehrforschung: Theorie und Empirie Festschrift für Ruediger Grotjahn. Frankfurt: Lang S. 129 - 143.

Kleppin, Karin (1998): Fehler und Fehlerkorrektur. Fernstudieneinheit 19, München: Langenscheidt.

Krumm, Hans-Jürgen (1990): „Ein Glück, dass Schüler Fehler machen!“ Anmerkungen zu Umgang mit Fehlern im lernerorientierten Fremdsprachenunterricht. In: Leupold, Einar/Yvonne Petter [Hrsg.]: Interdisziplinäre Sprachforschung und Sprachlehre, Festschrift für Albert Raasch, Tübingen S. 99 - 105.

Krumm, Hans-Jürgen (2009): Sprachenpolitik und Mehrsprachigkeit: Von der Mehrsprachigkeitsrhetorik zur (nicht mehr ganz so traurigen?) Realität. In: Visionen. Gegenwart und Zukunft von DaF/DaZ in Österreich. ÖDaF - Mitteilungen Sonderheft zur IDT 2009. S. 6 – 15.

Krumm, Hans-Jürgen (2006): Prüfungen und Bildungsstandards weltweit: die Globalisierung erreicht den Sprachunterricht. In: ÖDaF – Mitteilungen 2/2006. S. 67 – 73.

Krumm, Hans-Jürgen (2005): Was kann eine Sprachdiagnostik bei Kindern und Jugendlichen mit Migrationshintergrund leisten? In: Gogolin, Ingrid/Neumann, Ursula/Roth, Hans-Joachim [Hrsg.]: Sprachdiagnostik bei Kindern und Jugendlichen mit Migrationshintergrund. Münster: Waxmann S. 97 - 107.

Krumm, Hans-Jürgen (2005): Sprachtests: Politische und ethische Dimension. In: Bolton, Sybille: Probleme der Bewertung bei den Produktiven Fertigkeiten in den Deutschprüfungen des Goethe-Instituts. In: Kühn, Ingrid / Lehker, Marianne / Timmermann, Waltraud [Hrsg.]: Sprachtests in der Diskussion. Reihe: Wittenberger Beiträge zur deutschen Sprache und Kultur - Band 4. Frankfurt am Main u. a.: Lang S. 9 – 27.

Lado, Robert (1971): Testen im Sprachunterricht. Handbuch für die Erstellung und den Gebrauch von Leistungstests im Fremdsprachenunterricht. Aus den Amerikanischen von R. Freudenstein. München. (Original: Lado, Robert (1961) Language testing: The construction and use of foreign language tests. London: Longmans.)

Laska, Andrea/Krekeler, Christian (2009): Sprachtest. Leistungsbeurteilungen im Fremdsprachenunterricht evaluieren und verbessern. Baltmannsweiler: Schneider Verlag Hohengehren.

Lengyel, Drorit, Reich, Hans H., Roth, Hans-Hoachim, Döll, Marion [Hrsg.] Von der Sprachdiagnose zur Sprachförderung. Waxmann: Münster/New York 2009.

Lienert G. A./Raatz, U. (1994): Testaufbau und Testanalyse, Weinheim: Beltz.

Meuser, Michael & Nagel, Ulrike Nagel (1991): Experteninterviews – vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In.: Qualitative- empirische Sozialforschung: Konzepte, Methoden, Analysen. Hrsg. Von Detlef Garz. Klaus Kraimer. Opladen S. 141 – 144.

Oakeshott-Taylor, Angela (1977): „Dictation as a test of listening comprehension“. In: Dirven, René [Hrsg.]: Hörverständnis im Fremdsprachenunterricht/Listening comprehension in foreign language teaching. Kronberg, Ts.: Scriptor, 131 - 154.

Oller, John W., Jr.: Language tests at school: A pragmatic approach. London: Longman 1979.

Perlmann-Balme, Michaela (2010): Formen und Funktionen von Leistungsmessung und Kontrolle. In: Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. Herausgegeben von Hans-Jürgen Krumm, Christian Fandrych, Britta Hufeisen und Claudia Riemer. 2 Bände. De Gruyter Verlag: Berlin S. 994 – 1006.

Popp, Ulrike (2002): Geschlechtersozialisation und schulische Gewalt. Weinheim: Beltz.

Roelcke, Thorsten (2005): *Fachsprachen*. Berlin: Erich Schmidt Verlag.

Schelten, Andreas (2010): Einführung in die Berufspädagogik. 4., überarbeitete und aktualisierte Auflage, Stuttgart: Steiner Verlag.

Schifko, Manfred (2001): „Prüfungen, Zertifikate, Abschlüsse als Planungskategorien für den Unterricht“. In: Helbig, Gerhard et al. [Hrsg.]: Deutsch als Fremdsprache. Ein internationales Handbuch. 2. Halbband, Berlin: de Gruyter S. 827 - 834.

Solmecke, Gert (2000): „Faktoren der Schwierigkeit von Hörtests“. In: Bolton, Sybylle [Hrsg.]: TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar. Köln: VUB Gilde, S. 63.

Stigler, Hubert/Reicher, Hannelore [Hrsg.]: Praxisbuch Empirische Sozialforschung in den Erziehungs- und Bildungswissenschaften. Innsbruck-Wien-Bozen: Studienverlag 2005.

Szablewsky-Cavus, Petra (2000): Verstehen und Verständigung in Deutsch. Grundzüge einer berufsbereisübergreifenden Didaktik.“ In: Bildungsarbeit in der Zweitsprache Deutsch, 2/2000, S. 16.

## **Internetverweise**

---

Homepage des Bundesinstituts für Bildungsforschung, Innovation & Entwicklung des Österreichischen Schulwesens (BIFIE): Internet-Zugriff 04. 12. 2011

Homepage der FÖRMIG Gruppe - Ein Instrument zur Beobachtung bildungssprachlicher Entwicklungen in der Sekundarstufe: Internet-Zugriff 04. 12. 2011

Homepage von Peter May - Hamburger Schreibprobe: Internet-Zugriff 04. 12. 2011

DWDS (Wörterbuch der deutschen Gegenwartssprache): Internet-Zugriff 04. 12. 2011

## **E-Mail -Verkehr mit Expertinnen**

---

E-Mail von Drorit Lengyel vom 16.03.2011

E-Mail von Martina Glaboniat vom 30.05.2011

## Grundlagen im Bereich des Österreichischen Bundesheeres

---

Anforderungskatalog Sprachwesen 2008

Anforderungskatalog – Kompetenzen für Einsatz- und Friedensfunktionen im ÖBH 2010

Prüfungsordnung für Sprachprüfungen im ÖBH 2010

Durchführungsbestimmungen für die Sprachausbildung 2010

Curriculum für die Unteroffiziersausbildung Militärische Führung 2 (MilFü2) 2009

Ausbildungsgrundlage Förderseminar Deutsch 2011

Besprechungsprotokoll der AG Prüfmittel vom 13.03.2007

## Abkürzungsverzeichnis

---

### *ÖBH*

Österreichisches Bundesheer

### *STANAG 6001*

Standardization Agreement 6001 (Language Proficiency Levels) ist ein von Österreich im Rahmen der Partnerschaft für den Frieden ratifiziertes Abkommen mit den NATO-Vertragsstaaten, welches die Ebenen der Sprachkenntnisse (Sprachprüfungen) innerhalb der NATO und PfP regelt und diese damit gegenseitig anrechenbar macht.

### *MilFü2*

Lehrgang für militärische Führung 2 an der Heeresunteroffiziersakademie

### *SIB*

Sprachinstitut des Bundesheeres

### *UO*

Unteroffizier

### *GrpKdt*

Gruppenkommandant

## Abbildungsverzeichnis

---

<i>Abbildung 1:</i>	Auszug aus dem Anforderungskatalog für Sprachwesen	S. 11
<i>Abbildung 2:</i>	Bildungsstandards der 8. Schulstufe	S. 13
<i>Abbildung 3:</i>	Diagnoseinstrumente zur informellen Kompetenzmessung (IKM)	S. 14
<i>Abbildung 4:</i>	Monika Dannerer – Modell fachlicher und beruflicher Kommunikation aus ÖDaF-Mitteilungen, Heft 1/2008, S. 23	S. 30
<i>Abbildung 5:</i>	Kennzeichen von Fachsprache	S. 33
<i>Abbildung 6:</i>	Bumerang – Bildimpuls	S. 34
<i>Abbildung 7:</i>	Auswertungskriterien Deutsch	S. 34
<i>Abbildung 8:</i>	Auswertungsraster für die Sprachhandlung Beschreiben	S. 36
<i>Abbildung 9:</i>	Christians Schreibungen in der HSP 5-9	S. 40
<i>Abbildung 10:</i>	Lehrplan 1. – 4. Klasse Hauptschule	S. 60
<i>Abbildung 11:</i>	Schwierigkeitsfaktoren der akustischen Komponente	S. 63

## Abstract

---

Bei der vorliegenden Arbeit geht es um die Weiterentwicklung der schriftlichen Zulassungsprüfung Deutsch beim Österreichischen Bundesheer. Ziel dieser Arbeit ist es, die Grundlagen für ein neues, berufsorientiertes Prüfmittel zu schaffen, das auch faire Chancen für Menschen mit nichtdeutscher Muttersprache und Menschen mit motorischen Schwächen bietet. Insbesondere soll dabei geklärt werden, ob das aktuelle Prüfungsformat Diktat den Zweck der Qualifikation für das berufsspezifische Schreiben bzw. den Zweck der Zulassung zu gezielten Fördermaßnahmen erfüllt.

Im ersten, theoretischen Teil steht die Einordnung und Bestimmung des Testformats „Diktat“ im Mittelpunkt. Um das prüfungsrelevante Bezugssystem festzulegen, wird insbesondere eine Bestimmung des kommunikativen Handlungsfeldes „Militär“ vorgenommen. Zusätzlich werden relevante problembezogene wissenschaftliche Untersuchungen herangezogen, um den empirischen Teil der Untersuchung aus testtheoretischer Sicht abzusichern. Ein besonderes Augenmerk gilt dabei den Güte- und Nützlichkeitskriterien von Sprachtests. Den Abschluss des theoretischen Teils bildet eine zusammengefasste Expertise über die „Rechtschreibfähigkeit“ als Prüfungsgegenstand.

Im zweiten, empirischen Teil der Arbeit wird das Prüfmittel „Diktat“ einer umfangreichen Evaluierung unterzogen. Anhand von zwei Stichprobenerhebungen in Ebelsberg und Enns wird eine Untersuchung des Diktats vorgenommen, wo eine Kombination von vier unterschiedlichen Untersuchungsmethoden zur Anwendung kommt. Zum einen werden die Testergebnisse ausgewertet und interpretiert bzw. eine Beobachtung der Teilnehmer während des Tests durchgeführt. Zum anderen erschließt sich die Datenerhebung aus Fragebögen und Experteninterviews. Bei der Interpretation und Zusammenschau der Untersuchungsergebnisse werden die Messwerte anhand von Themenkomplexen strukturiert und in Beziehung gebracht.

Wie die Untersuchungsergebnisse zeigen, lassen sich mit dem Diktat weder die Schreibfähigkeit im militärischen Berufsfeld noch allgemeinsprachliche Fähigkeiten ausreichend überprüfen. Zentrale Qualitätskriterien wie Validität, Reliabilität und Objektivität werden nicht erfüllt.

Abschließend wird mittels einer Vergleichsstudie ein neues Testformat vorgestellt, das sowohl den Ansprüchen einer Zweitsprachendidaktik gerecht werden als auch die sprachlichen Kompetenzen von einsprachig aufwachsenden MuttersprachlerInnen messen soll.

## **Lebenslauf**

---

*Name:* Robert Rozinsky  
*Geburtsdatum:* 14. 10. 1962

*Vater:* Werner Rozinsky  
*Mutter:* Hannelore Rozinsky

### Schulbesuchszeiten:

1969 - 1973 Volksschule in Wien  
1973 - 1977 Hauptschule in Wien  
1977 - 1978 Oberstufenrealgymnasium in Wien  
1978 - 1980 Lehrberuf Industriekaufmann – Handelskammerprüfung  
  
2003 Beamtenaufstiegsprüfung

### Studium:

2004 Studienberechtigungsprüfung für das Fach „Deutsche Philologie“  
an der Universität Wien  
2004 - 2012 Studium „Deutsche Philologie“  
Diplomarbeit: „163 Wörter als Berufseinstiegshürde“  
Die Entwicklung von schriftlichen Zulassungsprüfungen in Deutsch  
beim Österreichischen Bundesheer

### Berufstätigkeit:

1981 - 2004 Berufsunteroffizier beim Österreichischen Bundesheer  
ab 2004 Beamter der allgemeinen Verwaltung beim Bundesheer  
2004 - 2005 Ausbildung zum Sprachtrainer Deutsch  
ab 2010 Lehrtätigkeit im Bereich Deutsch als Fremdsprache und Deutsch als  
Zweitsprache am Sprachinstitut des Bundesheeres