

UNIVERSITÄT ORAN, ES-SENIA

Fakultät für Literatur, Sprachen und Künste  
Abteilung der angelsächsischen Sprachen

Germanistikabteilung  
Fachbereich: DaF-Didaktik

*MAGISTERARBEIT*

THEMA:

**Zur Problematik der Leistungsmessung - und beurteilung  
im algerischen Daf-Unterricht**

Vorgelegt von **Fatiha Chalabi**  
geb. am 17.10.1985 in Akbou

**Vorsitzender**      **Prof. Dr.**      **Kamal EL KORSO**  
**(Président)**

**Betreuer**          **Prof. Dr.**      **Boualem Yettou**  
**(Rapporteur)**

**Gutachterin**  
**(examineur)**      **MC (A)**      **Rafiaa BELBACHIR**

Oran, 2011

## ***Danksagung***

---

Mein Dank gilt vor allem meinem Betreuer Prof, Dr. YETTOU Boualem für dessen Betreuung und für die konstruktive Kritik bei der Erstellung dieser Arbeit.

Mein Dank gilt auch den Jurymitgliedern: Frau Dozentin Dr. Belbachir Rafiaa, und Herrn Prof. Dr. El Korso Kamel .

Dankbar bin ich den Lehrern des Korrekturzentrums „ Mecki el Mahi“ (Constantine, 2011) für ihre wertvollen Antworten auf die Befragung.

Ich bedanke mich herzlich bei den Schülern der Gymnasien Cherif Lerguett und Rebah (2010-2011) in Bordj Bou Arreridj für ihre Antworten auf die Befragung.

Ein besonderes Wort des Dankens gilt meiner Familie für die Hilfe und Ermutigung.

## **Inhaltsverzeichnis**

### **Abkürzungsverzeichnis**

### **Tabellenverzeichnis**

	<b>Einleitung</b>	<b>3</b>
<b>1.</b>	<b>Zu den theoretische Grundlagen von „Testen und Evaluation“</b>	<b>5</b>
<b>1.1</b>	<b>Zur Leistungsmessung</b>	<b>5</b>
1.1.2	Zum Verhältnis von Testen und Evaluation	7
1.1.3	Funktionen der Leistungsbeurteilung	8
<b>1.2</b>	<b>Zur Testtheorie</b>	<b>10</b>
1.2.1	Der Strukturalistische Ansatz	13
1.2.1.1	Testen von isolierten Elementen und Fertigkeiten	13
1.2.1.2	Testen von integrierten Fertigkeiten	14
1.2.2	Die kriteriumsorientierte Leistungsbeurteilung	14
1.2.3	Die zielerreichende Leistungsbeurteilung	15
1.2.4	Der Gemeinsame Europäische Referenzrahmen und Sprachzertifikate	18
<b>1.3</b>	<b>Zu den Testgütekriterien</b>	<b>19</b>
<b>1.4</b>	<b>Theoretische Grundlagen zum Testen</b>	<b>27</b>
1.4.1	Sprachtesten und Spracherwerbforschung- und Lehren	27
1.4.2.	Sprachliche Fertigkeiten in der kompetenzorientierten Testtheorie	28
1.4.2.1	Kommunikative Sprachkompetenz und Sprachtestforschung	29
1.4.2.2	Bachmans Modell der Sprachkompetenz	30
1.4.2.4	Zu den Testfacetten	38
1.4.2.5	Zu den Arbeitsanweisungen	41
1.4.3	Vorgehensweisen bei der Erstellung standardisierter Tests	44
1.4.3.1	Testimpact auf die Testkandidaten	49
1.4.3.2	Affektive Schemata und Testperformanz	51
1.4.3.3	Die Beurteilungsmethode	52
<b>2.</b>	<b>Empirische Untersuchungen und Befunde</b>	<b>55</b>

2.1	Curriculumanalyse	55
2.1.1	Ergebnisse und Auswertung der Curriculumanalyse	62
2.2	Analyse der Prüfungsanleitung	64
2.3	Ergebnisse und Auswertung der Lernerbefragung	91
2.4	Ergebnisse und Auswertung der Lehrerbefragung	93
2.5	Analyse und Auswertung der Schülerkopien	95
<b>Zusammenfassung und Ausblick</b>		<b>99</b>
<b>Literaturverzeichnis</b>		<b>A-C</b>
<b>Anhang</b>		<b>a-f</b>
<b>Eidesstattliche Erklärung</b>		

**Tabellenverzeichnis:**

- Tabelle 1** Ergebnisse der Testanalyse im Hinblick auf die Konstruktvalidität.
- Tabelle 2** Ergebnisse der Evaluation der Hauptschulprüfungsreliabilität.
- Tabelle 3** Ergebnisse der Testanalyse im Hinblick auf die Authentizitätskriterien.
- Tabelle 4** Ergebnisse der Testanalyse im Hinblick auf die Interaktivitätskriterien.

## **Abkürzungsverzeichnis**

<b>L. M</b>	Leistungsmessung
<b>L. B</b>	Leistungsbeurteilung
<b>N.R</b>	Norm Referenced Testing ( Normorientiertes Testen.)
<b>TLU</b>	Target Language Use ( Sprachgebrauchsbereich)
<b>ONEC</b>	Office Nationale des Examens et Concours
<b>ITE</b>	Institut Technologique d'education
<b>D. I</b>	Diskriminierungsindex
<b>F.V</b>	Facility Value ( Einfachheitsgrad)

## **Einleitung**

Beweggrund für diese Magisterarbeit ist die zunehmende Bedeutung, die der Leistungsbeurteilung (*L.B*) im algerischen DaF-Unterricht zukommt, insbesondere nach der vom erzieherischen Ministerium 2006 eingeführten Reform an algerischen Gymnasien. Die Leistungsbeurteilung wurde im Rahmen dieser Reform nicht vernachlässigt, denn die Schüler werden formellen Prüfungen und Tests unterzogen. Darüber hinaus werden sie im Unterricht kontinuierlich kontrolliert und die Hauptschulprüfung wird in einer einheitlichen Form eingesetzt. Trotzdem stellt das Thema „Leistungsbeurteilung bzw. Evaluation“ im algerischen Erziehungssystem immer noch Fragen, z.B. die Frage, ob diese Prüfungsreform einwandfrei ist und ob sie dem didaktischen Ansatz der fremdsprachlichen Reform im algerischen Gymnasium entspricht?

Auf der akademischen Seite bzw. an der Deutschabteilung der Universität Oran liegen meines Wissens keine wissenschaftlichen Arbeiten im Bereich der Leistungsmessung- und -beurteilung vor. So kann man feststellen, dass die Leistungsbeurteilung sowohl für die Lerner als auch für die Lehrer ein weitgehend stressverursachendes Thema ist. In diesem Gedankengang will die vorliegende Magisterarbeit die Problematik der Leistungsbeurteilung im algerischen DaF-Unterricht untersuchen. Im Zentrum der Untersuchung steht die Fragestellung: Inwieweit kann behauptet werden, dass die Testergebnisse im algerischen DaF-gymnasialen Unterricht den tatsächlichen sprachlichen Kompetenzen der Lerner entsprechen. Um diese Frage zu beantworten, gehe ich von den folgenden Arbeitshypothesen aus:

- Die im algerischen DaF-Unterricht verwendete Testpraxis verhilft nicht zur Verbesserung der Performanz der Lerner.
- Die Testform erfüllt aber die Gütekriterien und hilft somit den Lernern, bessere Leistungen zu erbringen.
- Die Prüfungsformen im algerischen DaF-Unterricht verfolgen die neuen Anforderungen der Lehrpläne nicht.
- Der Akzent wird mehr auf die Sprachanalyse gesetzt als auf die Sprachkompetenz.

Methodisch basiert die vorliegende Arbeit sowohl auf theoretischen Erkenntnissen als auf empirischen Untersuchungen. Deshalb wird sie in zwei Komplexe aufgeteilt. Im ersten Teil werden die theoretischen Grundlagen zur Problematik „Testen“ dargestellt. Hier wird die Wichtigkeit der Leistungsmessung für das Sprachenlernen gezeigt, eine Testtypologie

beschrieben und auf die Theorie der englischen Testwissenschaft eingegangen, nämlich auf die Konzepte von „Validität, Reliabilität und Gültigkeit“. Dieser theoretische Rahmen gilt als Basis für meine analytischen und experimentellen Untersuchungen im empirischen Teil. In diesem Teil werden zuerst einmal das Curriculum, sowie die Prüfungsanleitung der Hauptschulprüfung im Zusammenhang mit der Leistungsbeurteilung eingeführt. Danach werden die Ergebnisse der Schüler- und Lehrerfragebögen analysiert. Im letzten Kapitel werden die Lernerkopien analysiert, um eventuelle Einflüsse der Testmethode auf ihre Performanz zu erforschen.



## **1. Zu den theoretischen Grundlagen von „Testen und Evaluation“**

Der vorliegende Teil gliedert sich in drei Kapitel. Zuerst werden die Begriffe Testen und Evaluation definiert; es werden darüber hinaus die kriteriumsorientierte und die zielerreichende Leistungsbeurteilung offen gelegt. Dabei werde ich den Zusammenhang zwischen Testen und Evaluation erklären und den Akzent auf den Stellenwert der Leistungsbeurteilung im Sprachunterricht setzen, indem die verschiedenen Auffassungen zur Legitimierung der Leistungsbeurteilung in der Schule erläutert werden. Das zweite Kapitel wird den Beitrag der strukturalistischen Schule umreißen, denn die Anfänge der Testtheorie waren von den strukturalistischen Prinzipien geprägt. Es wird hauptsächlich auf die Testverfahren und auf die praktischen Prozeduren bei der Erstellung der Sprachtests bzw. auf die Design- und Operationalisierungsetappe und Durchführung eingegangen. Schwerpunkt des theoretischen Teils stellt das dritte Kapitel dar. Hier werden die Hauptbegriffe in der neuen Testtheorie skizziert, sowie die Komponenten der Arbeitsanweisungen und die praktischen Prozeduren zur Beurteilung eines Sprachtests.

### **1.1 Zur Leistungsmessung**

Das Wort „Leistungsmessung“ wird in vielen Bereichen verwendet. ZIEGENSPECK verweist auf die semantische Mehrdeutigkeit des Wortes; der Ursprung des Wortes „Leistung“ wird zum Gotischen „laistjahn“ zurückgeführt und bedeutet folgen oder nachfolgen.<sup>1</sup> In Lexika wird der Begriff Leistung rein physikalisch definiert. „... für die in der Zeiteinheit geleistete Arbeit“<sup>2</sup>. Theoretiker sehen, dass die allgemeine Definition von Leistung pauschal und der Fremdsprachendidaktik nicht entspricht. Zu diesem Zweck wird auf SACHER verwiesen: „...Wir (können) Leistungen im Humanbereich definieren als die zur Erlangung eines Zieles aufgewandte und auf einen Gütemaßstab bezogene Anstrengung.“<sup>3</sup> Zu humaner Leistung gehört also: Anstrengung, Zielgerichtetheit und ein Gütemaßstab.“<sup>4</sup>

Unter dem Begriff „Test“ werden nicht nur Klassenarbeiten, sondern alle Maßnahmen verstanden, mit denen auf methodisch kontrollierte Weise Informationen über das Sprachvermögen der Lerner erhoben werden. Das umfasst auch standardisierte allgemeine Sprachtests, die von Einrichtungen wie dem Goethe-Institut in weltweit einheitlicher Form angeboten werden.

---

<sup>1</sup> ZIEGENSPECK, J. ( 1973, S13-14): Zensur und Zeugnis in der Schule. Hannover: Schrödel .

<sup>2</sup> FELDBAUM, M. ( 1999, S 501): Neues universal Lexikon in Farbe von A-Z , Compact Verlag: München

<sup>3</sup> Vgl. KLAFKI. (1996, S 228): Neue Studien zur Bildungstheorie und Didaktik. Beltz

<sup>4</sup> SACHER, W. (2009, S 1) Leistungen entwickeln, überprüfen und beurteilen. Bewährte und neue Wege für die Primar- und Sekundarstufe. Julius Klinkhardt.

„Testen bedeutet (...) das Herstellen einer standardisierten Beobachtungssituation, in der alle Probanden unter gleichen Bedingungen die gleiche Aufgabenstellung zu lösen suchen. Die Lösungen der Probanden und die Beobachtungen der Testleiters werden dann noch in einer festgelegten standardisierten Weise protokolliert, so dass die Protokolle von verschiedenen Personen in genau der gleichen Weise und inhaltlich gleich ausgewertet und interpretiert werden können.“<sup>5</sup>

CARROL definiert Tests wie folgt: „Psychological or educational test is a procedure designed to elicit certain behaviour from which one can make inferences about certain characteristics of an individual.“<sup>6</sup> Andererseits unterscheidet BACHMAN (1990, 20) zwischen Evaluation, Tests und Messung. Ein Test sei ein Messinstrument zur Erhebung einer Stichprobe des individuellen Verhaltens eines Individuums. Laut ihm misst ein Test individuelle Charakteristika mithilfe expliziter Prozeduren und Verfahren; was ein Test von anderen Messungsarten unterscheidet, ist, dass er durchgeführt wird, um eine spezifische Art des Verhaltens zu erheben.<sup>7</sup>

Beim Testen werden die Begriffe „Messung“ und „Beurteilung“ nicht einheitlich verwendet. Mit Messung wird die Zuweisung eines Punktwerts bzw. einer Zahl, während die Beurteilung den Punktwert oder das Testergebnis interpretiert.

Die Beurteilung ist nach Meinung von NAMARA (2000, 55) ein umfassender Prozess, der zwei Phasen einschließt:

- die Messung, in der der Prüfer oder die Lehrkraft eine Punktzahl der Lernerleistung zuweist.
- die Beurteilung der geleisteten Arbeit anhand einer Checkliste in Bezug auf das Testergebnis.<sup>8</sup>

Bei WEIR (1990, 41) besagt der Punktwert 5 nichts, wenn er einem bestimmten Standard nicht verglichen wird. In diesem Zusammenhang ist für ihn die Messung der gesprochenen Sprache potenziell mehr problematisch als die Messung und Beurteilung geschriebener Lernerproduktion.<sup>9</sup>

---

<sup>5</sup> „KLEBER. (1979, 12) zitiert nach BOLTON, S (1985, S 60) : Die Gütebestimmung kommunikativer Tests, Tübingen: Günter Narr

<sup>6</sup> CARROLL, J. (1992, S 46) The ACQUILEX Lexical Database System: system description and user manual. In A. Sanfilippo (ed.) *The (Other) Cambridge ACQUILEX Papers*, Technical Report No. 253, Computer Laboratory, University of Cambridge.

<sup>7</sup> BACHMAN, L.(1990): *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

<sup>8</sup> Mc NAMARA, T. (2000) : *Language Testing*. Oxford University Press.

<sup>9</sup>WEIR, C. J. (1990, 41): *Communicative Language Testing*, New York: Prentice-Hall

### 1.1.2 Zum Zusammenhang zwischen Testen und Evaluation

Entscheidungen im erzieherischen Bereich sind mehr oder weniger zutreffend. Daher wird großen Wert auf die Evaluation gelegt. Hierzu schrieb BACHMAN (1990, 22), dass die Ergebnisse der Lehrerbeurteilungen nicht nur von der Kompetenz des Lehrers abhängen. Die getroffene Entscheidung hängt auch von der Qualität der gesammelten Daten über die getestete Kompetenz des Lerners. "The probability of making the current decision in any given situation is a function not only of the ability of the decision maker, but also of the quality of the information upon which the decision is based."<sup>10</sup>

Ziel der Evaluation ist folglich die Sammlung von reliablen und relevanten Informationen, um eine bildungsbezogene Entscheidung zu treffen. Tests sind an sich nicht evaluativ, sondern nur wenn die erbrachten Ergebnisse mithilfe bestimmter Bildungsstandards und Referenzen interpretiert und beurteilt werden.<sup>11</sup>

Evaluation wird vielmehr eingesetzt, um pädagogische Zwecke zu erreichen, Motivation bei den Lernenden zu stimulieren oder für rein deskriptive Zwecke. Nur wenn Testergebnisse bei Entscheidungen mitwirken, kann es sich um Evaluation handeln. Daher ist es laut BACHMANN (1990, 22) wichtig, zwischen der informationenbietenden Funktion der Messung und der Entscheidungsfunktion der Evaluation zu unterscheiden.<sup>12</sup> Hierzu versteht man, dass sich die Messung auf die Informationensammlung bezieht, während die Evaluation mit den Entscheidungen über die Lerner zu tun hat. Evaluation schließt zwei Komponenten ein: die Sammlung der Information und die Entscheidungen, die aufgrund dieser Information getroffen werden.<sup>13</sup>

Eine Information wie „helping Fanny improve her accent“ ist eine qualitative Information. Die Information: „Ein Schüler hat den Punktwert 4“ ist eine quantitative Information.<sup>14</sup> Die Unterscheidung zwischen qualitativen und quantitativen Informationen ist im Bereich des Sprachtestens üblich. Die Evaluation im Fremdsprachenunterricht kann beide Formen einsetzen. Hierzu meinen UPSCHUR und GENESE(1996, 55), dass die Vielfalt der gesammelten Informationen über das Lehren und Lernen die Reliabilität der entsprechenden Entscheidungen verbessern kann.<sup>15</sup> Die im erzieherischen Kontext gesammelten Informationen beziehen sich auf Lehreraktivität, Lernerperformanz, das didaktische Material,

---

<sup>10</sup> BACHMAN. (1990, S. 22): Fundamental Considerations in Language Testing, Oxford; Oxford university press.

<sup>11</sup> Ebd., S. 22

<sup>12</sup> Ebd., S. 23

<sup>13</sup> Vgl. BACHMAN/PALMER. (1996, S. 21): Language Testing in practice. Oxford University Press.

<sup>14</sup> UPSHUR, J/ GENESE, G. (1996, S. 55): Classroom based Evaluation in Second Language Education. Cambridge: Cambridge University press.

<sup>15</sup> Ebd., S. 55

die Lehrpläne usw. Deshalb sollten diese Informationen so genau wie möglich gesammelt und so klar wie möglich präsentiert werden. Tests und Prüfungen werden auch eingesetzt, um Stärke und Schwäche der Lerner zu erkennen. Prinzipiell ist jeder Sprachtest ein diagnostischer Test. Es wird aufgrund dieser Informationen auch entschieden, welche Inhalte zu vermitteln sind. Feedback ist wichtig sowohl für Lerner als auch für Lehrer, besonders bei der formativen *L.B.* Darum werden Tests eingesetzt um zu überprüfen ob die gelernten Inhalte assimiliert sind oder nicht. In diesem Zusammenhang schreiben BACHMAN/ Palmer (1996, 61), dass die Noten in den meisten Fällen auf der Basis der Lernerperformanz in den Tests und der Lernerperformanz innerhalb des Klassenzimmers zugewiesen werden.<sup>16</sup>

### **1.1.3 Funktionen der Leistungsbeurteilung**

Die Leistungsmessung hat im Fremdsprachenunterricht allgemein eine zentrale Rolle; sie liefert den Lernenden Informationen über den Stand ihrer Sprachentwicklung und fungiert als Rückmeldung für Lehrer und Eltern.

Die Ergebnisse der Leistungsmessung regulieren darüber hinaus die Lernanstrengungen und helfen den Lernern, den Lernstand und die Lücken zu erkennen. Dies verstärkt das strategische Lernverhalten der Lernenden, indem sie nur auf die unassimilierten Inhalte konzentrieren. Tests und Prüfungen werden im erzieherischen Bereich also hauptsächlich als Informationsquelle eingesetzt. Sie gelten als unabdingbare Bewertungsgrundlagen für die Selektionsfunktion der Schule. Deshalb sind Tests wichtig, um gültige Informationen zu liefern. Hierzu sei auf BACHMAN verwiesen: "In order to justify the use of tests for educational evaluation, certain assumptions and considerations must be made regarding the usefulness and quality of the information which tests provide." BACHMAN/PALMER (1996)<sup>17</sup>

Einen pädagogischen Beitrag bietet die *L.B.* wenn die Testergebnisse auf die Lernenden ermutigend einwirken und das Leistungsgefühl verstärken. Für die Lehrkräfte sind Tests ebenso wichtig; die *L.M.* liefert den Lehrern Informationen über ihre eigenen Lehrverhalten bzw. Prüfungsverhalten. Wenn die vom Lehrer erstellten oder ausgewählten Tests an Gütekriterien mangeln, werden die Ergebnisse divergieren. Wenn z.B. die ganze Klasse unterdurchschnittliche Noten erhält, eventuell bei mangelnder Inhaltsvalidität, oder unklaren Arbeitsanweisungen. ALDERSON/ DIANE und WALL (1995) gehen von der Annahme aus, dass ein Test an sich den Unterschied macht. Ein guter Test produziert einen positiven

---

<sup>16</sup> BACHMAN/PALMER. (1996, S. 61)

<sup>17</sup> Ebd., S. 54

Washback, das Gegenteil gilt für einen ungültigen Test. Für sie funktioniert der positive Washback nur wenn der eingesetzte Test gültig sei und den Zielstellungen im Design-Statement und im Blue-Print entspricht.<sup>18</sup> In Anlehnung ALDERSON et al., (1995) meint MESSICK(1996, S: 251), dass ein negativer Impact des Tests nicht von irgendeiner mangelnden Validität des Tests entstammen sollte.<sup>19</sup>

Über das für die Lehrkraft gelieferte Feedback über die Effektivität des Lehrens und des Lernprogramms hinaus, widmet das Sprachtesten Informationen, die bei der Auswahl der Lernmaterialien und Lerninhalte sowie für die Bildungspolitik und Verwaltung notwendig sind<sup>20</sup>. Darum enthalten virtuell alle Sprachcurricula Hinweise und theoretische Grundlagen zum Sprachtesten, um die Lehrer zu verhelfen, die angemessene Testform zu wählen und eventuell selbst Tests zu entwerfen.<sup>21</sup>

In Deutschland kam die Ablehnung der Ziffernnoten nach der von Erziehungspädagogen starken Kritik der frühen Selektion in deutschen Volksschulen.<sup>22</sup> WEINERT (2002, 18) sieht ein, dass die Forschung zum Thema Leistungsbeurteilung an der noch uneinheitlichen Meinung über Notenpraxis leide. Es haben sich folglich zwischen den Befürwortern der Zensurgebung und ihrer Gegnern feste Fronten gebildet.“

WEINERT setzt sich vor allem gegen diese „Schwarz-Weiß“ Diskussion bezüglich Notenpraxis.“

Fatalerweise ist nach ihm die vergleichende Leistungsmessung (...) in die bewertende und pauschalisierende Pädagogie „Entweder- Oder –Klassifikation“ geraten”.<sup>23</sup>

Die formative Leistungsbeurteilung bezieht sich auf die Beurteilung, die während des Unterrichts ausgeübt wird als Rückmeldung für den Lehrer, auch Lernfortschrittskontrollen genannt.<sup>24</sup> Der formativen Leistungsbeurteilung kommt laut BACHMANN (1990, 55) viel

---

<sup>18</sup> WALL/ DIANNE/ ALDERSON. (1995): “Examining washback: The Sri Lankan impact study in Alister Cumming and Richard Berwick (eds.): Validation in language Testing .Multilingual Matters,

<sup>19</sup> . MESSIK, S. (1996): Validity and washback in language testing, in Language Testing V 13, 1996.

<sup>20</sup> INGENKAMP/LISSMANN (2008, S:13) :Lehrbuch der pädagogischen Diagnostik , Beltz.

<sup>21</sup> BACHMAN, L . (1990, S. 8)

<sup>22</sup> LEDERER, A. (2008): Prüfungen kritisch überprüft: Probleme der schulischen Prüfungs- und Beurteilungspraxis untersucht an schriftlichen Prüfungen und Prüfungsaufgaben in ausgewählten Fächern der Realschule. Julius Klinkhardt: Bad Heilbrunn

<sup>23</sup> WEINERT, F. (2002): Leistungsmessungen in Schulen, Weinheim.

<sup>24</sup> TRIM / NORTH/ COSTE und SHEILS. (2001): Europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Europarat: Straßburg.

Lob zu, da sie durch Feedback auf eine Verbesserung des Unterrichts zielt. Vielmehr ist Feedback ein unabdingbarer Mechanismus für die Effektivität jedes erzieherischen Systems.<sup>25</sup>

Laut MAIER (2010) ist die formative Leistungsbeurteilung die alternative Form für die traditionelle Leistungsmessung in der Schule, die grundlegend auf die Notengebung und Selektion fokussiert. Er plädiert für eine Reform der schulischen Leistungsmessung, und für die Optimierung von Lehr- und Lernprozessen durch den Einsatz neuer Beurteilungsformen wie Portfolios und Lerntagebücher.<sup>26</sup>

Laut dem Europäischen Referenzrahmen für Fremdsprachen funktioniert Feedback nur, wenn der Rezipient in der Lage ist:

- etwas zu bemerken, d. h. aufmerksam und motiviert ist und vertraut mit der Form, in der er die Information eintrifft (...) zu interpretieren
- und ein entsprechendes Sprachbewusstsein hat, um den springenden Punkt zu verstehen und nicht zu kontraproduktiven Handlungen zu greifen.“<sup>27</sup>

Eine formative Evaluation, durch ihr bewusstseinbildendes Üben wird unter „evaluation formatrice“ bekannt. Grundlegend für diese Evaluationsform ist der subjektive Eindruck, der der Lerner über seine eigenen Leistungen anhand der Scheckliste hat und dies mit der Realität vergleicht. Formative Lerndiagnose wird meist detailliert eingesetzt besonders um frisch gelehrt Inhalte zu prüfen.

Die summative Leistungsbeurteilung bezieht sich dagegen auf normorientierte, am Kursende eingesetzte Beurteilung. Es geht nicht notwendigerweise um einen Performanztest oder ein „proficiency test“, da sie unbedingt kursbezogen sind. Der Interaktion wird allgemein in der Sprachverwendung und beim Sprachenlernen große Bedeutung zugeschrieben, weil sie eine zentrale Rolle bei der Kommunikation spielt. Üblich werden die Lernerleistungen einer vorgegebenen Skala eingeordnet. Die Skala kann eine Notenskala sein und Stufen von sehr gut bis ungenügend enthalten oder eine Punkteskala, z. B. von 0 bis 100 Punkte.

INGENKAMP( 1983, S. 495) glaubt, dass die Lerndiagnose vor allem in modernen Schulen zur Kontrolle des Lernerfolgs durch Folgendes dienen sollte:

- Korrektur falscher Lernergebnisse,
- Ausgleich von Lerndefiziten,
- Bestätigung erfolgreicher Lernschritte,
- Planung der nachfolgenden Lernschritte,

---

<sup>25</sup> Ebd., S .55

<sup>26</sup> MAIER,U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? **In:** Zeitschrift für Erziehungswissenschaft, 13(2), 293-308.

<sup>27</sup> TRIM et al.:(2001).

- Motivierung durch Steuerung der Aufgabengeschwindigkeit,
- Optimierung der Lernbedingungen.<sup>28</sup>

Problematisch bleibt, dass Prüfungen im täglichen Unterricht diese Funktionen erfüllen müssen. Hierzu schreibt INGENKAMP ( 1983, S. 30) :

“Die Zensurgebung genügt weder den Mindestanforderungen an die Objektivität der Auswertung, an die Gültigkeit und Zuverlässigkeit der Erfassung des Schülerverhaltens, noch liefert sie ausreichend differenzierte Informationen oder über den Rahmen der Klasse hinausreichende Vergleichsmaßstäbe (...) Weder dem Beurteiler noch den Adressaten ist klar, in welchem Masse Feststellungen, Meinungen, pädagogische Intentionen in die Zensur eingegangen sind. Der Vorgang des Zensierens und die Zensur sind eine viel zu unsichere Basis um daraus pädagogische Maßnahmen abzuleiten.”<sup>29</sup>

Lehrkräfte beklagen sich, dass sich die Schüler für die Fächer, in denen sie nicht benotet werden, unmotiviert sind. Es lohnt sich zu erwähnen, dass die Schüler für die Prüfung lernen und die Lehrer für das Lernen prüfen. Daher hört man in der pädagogischen Diskussion, dass die Zensurgebung den Schülern beim Lernen helfe. GÖBLER stellt fest, dass das Prüfungsbezogene Lernen der Schüler nicht negativ betrachtet werden soll.“ Besser arbeitet der Schüler wegen der Zensuren als überhaupt nicht “<sup>30</sup> GÖBLER (1966, S.143). Pädagogen versichern auch, dass Noten durch das Selbstwertgefühl und Leistungsgefühl den Lernprozess bestätigen und die Lernmotivation erhöhen. Hingegen begründen die Befürworter der Notenabschaffung ihre Haltung dadurch, dass die Schüler ausschließlich für und wegen der Prüfungen lernen, und so werden die Schüler sich nur für den Punktwert interessieren, was die Rückmeldung ausschließt, da sie auf ihre Fehler kaum achten. Die Fragen der Leistungsmessung -und beurteilung in der Schule sind überdies mit denen der Qualitätssicherung und der Lehrplanproblematik eng verbunden. Die Leistungsmessung hat als oberstes Ziel die Überprüfung der im Curriculum festgelegten Lernziele. Vielmehr liefern Tests Informationen über den Erfolg bzw. Misserfolg des Unterrichts, der Lehr- und Lernmethode, des Lehrwerks usw. Informationen von einem Sprachtest können nach BACHMANN zur Entwicklung neuer Programme eingesetzt werden, sie können auch indizieren, inwieweit die Lernziele, die das Curriculum fördert, erreicht wurden.<sup>31</sup>

---

<sup>28</sup> INGENKAMP, K. (1983): Die Fragwürdigkeit der Zensurgebung. Texte und Untersuchungsberichte. Weinheim: Basel

<sup>29</sup> Ebd., S. 30

<sup>30</sup> GÖBLER, A. (1966) : Zensuren und Zeugnisse. Stuttgart: Klett

<sup>31</sup> BACHMANN, L. ( 1990, S. 62)

## **1.2 Zu den Testtheorien**

Im Zusammenhang mit der Testtheorie unterscheidet SPOLSKY drei Etappen:

- „die vorwissenschaftliche Zeit der psychometrischen Testtheorie,
- die strukturalistische Testtheorie,
- und die integrativ-soziolinguistische Zeit.“<sup>32</sup>

Die vorwissenschaftliche Zeit, die von den psychometrischen und strukturalistischen Ansätzen beeinflusst war, war auch durch mangelndes Interesse für die statistischen Methoden und Gütekriterien wie Objektivität und Reliabilität, sowie durch die Beurteilungen der Lehrkraft charakterisiert. In der integrativen Etappe konzentrierte sich die Forschung auf die Realisierung objektiver Testverfahren. Zu diesen Etappen kamen die großen Entwicklungen in der Testtheorie, die die Testentwicklung mit der kommunikativen Didaktik binden und statistisch validieren. Seit den Anfängen der testtheoretischen Forschung in den 60er Jahren waren Tests von dem strukturalistischen Ansatz beeinflusst. Dies zeigte sich in den Forschungen von LADO.

### **1.2.1 Der strukturalistische Ansatz**

#### **1.2.1.1 Testen von isolierten Elementen**

Die Anhänger der strukturalistischen Testtheorie legten die sprachlichen Elemente fest, die überprüft werden sollen, in denen Unterschiede zwischen Muttersprache und Fremdsprache entstehen. Zunächst bezog sich diese Auffassung auf die Aussprache bzw. auf rein phonetischer Ebene. LADO hatte beobachtet, dass die Strukturen, in denen sich die beiden Sprachen übereinstimmen, für die Lernenden keine Schwierigkeiten darstellten. Da aber wo zwischen der Muttersprache und Fremdsprache Divergenzen bestehen und darum zu Interferenzen führen, sollte die Lehrkraft überprüfen, ob die Lerninhalte assimiliert sind oder nicht.<sup>33</sup>

---

<sup>32</sup> SPOLSKY, B. (1978): Educational linguistics: An introduction. Rowly, MA: Newbury House.

<sup>33</sup> LADO, R. (1971): Testen im Sprachunterricht. München: Max Hueber



LADO(1971) stellte fest, dass die Lerner die von der Muttersprache her erworbenen Gewohnheiten auf die Fremdsprache zu übertragen neigen.<sup>34</sup> Er ging allgemeiner ein; er fand heraus, dass diese Fehler nicht nur in der Aussprache festgestellt werden können, sondern auch auf der lexikalischen und syntaktischen Ebene. Grundlage für diese Untersuchungen waren die Befunde der kontrastiven Linguistik, die in den 40er und 50 er Jahren aufblühte. Da sie einen linguistischen Vergleich der sprachlichen Ebenen durchführt. Zusammenfassend sei hier LADO (1971, S. 178) zitiert:”Als Arbeitshypothese können wir darum formulieren: Das Erlernen der Probleme ist gleichbedeutend mit dem Erlernen der Sprache, und wir können weiterhin feststellen: das Testen der Probleme ist gleichbedeutend mit dem Testen der Sprache.”<sup>35</sup> Die Wichtigkeit dieser Verfahren besteht nach BOLTONs (1985, S. 21) Auffassung darin, dass die Testinhalte und Testschwierigkeiten vorhergesagt werden können.<sup>36</sup>

Die Überprüfung der integrierten Fertigkeiten beruht auf dem Standpunkt, dass im Sprachgebrauch die Fertigkeiten nicht isoliert, sondern zusammenhängend eintreten. Bei LADO ist die Fertigkeit nur rein linguistisch aufgefasst. „Die Fertigkeit ist die in ihr enthaltenen sprachlichen Elemente und Strukturen“.<sup>37</sup> OLLER kritisierte, dass LADO den Spracherwerb in das Erlernen dieser Elemente und Strukturen einschränkt und damit die Rolle des kommunikativen Kontexts weitgehend ausschließt<sup>38</sup>, was für die Testpraxis nicht ohne Folgen bleibt. Nach BOLTON geht der strukturalistische Ansatz auch bei dem Verfahren zur Überprüfung der integrierten Fertigkeiten nicht von Kommunikationssituationen und/oder Themen aus, sondern - wie auch bei den Testaufgaben zu den sprachlichen Elementen - von einer Zusammenstellung der Lernprobleme<sup>39</sup>

### **1.2.1.2 Testen von integrierten Fertigkeiten**

Die sprachlichen Mittel treten bei der Kommunikation nicht isoliert auf. Die Lernenden erwerben sie oft isoliert. Bei der Kommunikation d.h. im konkreten Sprachgebrauch erfahren wir die sprachlichen Mittel in einem einzigen, integrierten Redefluss. LADO meint, dass wir bei den integrierten Fertigkeiten nicht nur die Elemente einer Sprache, die für eine bestimmte Fertigkeit vorausgesetzt werden müssen, testen. Im gleichen Teil eines Tests kann eine Aufgabe das Vokabular betreffen, eine Andere die Struktur, eine Dritte die Aussprache und

---

<sup>34</sup> Vgl. BOLTON, S. (1985, S.20): Die Gütebestimmung kommunikativer Tests, Tübingen: Günter Narr

<sup>35</sup>LADO, R. (1971):Testen im Fremdsprachenunterricht.München: Max Hueber Verlag.

<sup>36</sup> BOLTON , S. (1985, S. 21)

<sup>37</sup> Ebd., S.21

<sup>38</sup> OLLER, J. (1979): Language Tests at School. Longman .

<sup>39</sup> BOLTON, S. (1985, S. 21)

eine vierte wiederum das Vokabular, ohne dass der Schüler durch den Test erführe, dass die betreffende Aufgabe sich mit Grammatik oder Vokabular befasst.“<sup>40</sup>

CAROLL(1968, 58) schrieb, dass „diese Tests, die Elemente der Sprache isoliert testen, dem eigentlichen Gebrauch der Sprache nicht entsprechen. Sie testen nicht die kommunikative Kompetenz, die kombinierte Anwendung von sprachlichen Mitteln im Alltag.“<sup>41</sup> Man bemängelt deshalb am strukturalistischen Ansatz hinsichtlich der Leistungsmessung im Sprachunterricht die Konzentration auf die linguistischen Elemente der Sprache und das Übersehen der Sprache als Kommunikationsmittel.

### **1.2.1.3 Die kriteriumsorientierte Leistungsbeurteilung**

In den letzten fünfundzwanzig Jahren hat der normorientierte Test „Norm referenced testing“ (NR) die Testverfahren dominiert, in dem die individuelle Leistung mit den anderen Personen der Lerngruppe verglichen wird.. Hingegen liefert Kriteriumsorientiertes Testen Informationen im Vergleich zu einem bestimmten Standard. Die normorientierten Prüfungen orientieren sich an der klassischen Testtheorie. Dieses Kriterium besagt, dass die Bewertung sich auf der jeweiligen Lerngruppe orientieren sollte. Folglich werden die sprachlichen Leistungen des Lernenden mit denen der anderen verglichen und in einer Reihenfolge eingeordnet. Hier ist zu erwähnen, dass diese Beurteilung die Lehrkraft zu regelwidrigen Benotung und zwar zur Scheinobjektivität verleitet.

Die klassische Testtheorie, an der die normorientierte L.B orientiert, legt fest, dass die Testaufgaben eine mittlere Schwierigkeit besitzen, so dass sie prinzipiell von 50 Prozent der Lernenden gelöst werden können.<sup>42</sup> INGENKAMP erläutert im Zusammenhang mit dem Vergleich von Leistungen: „Schülerleistungen aus verschiedenen Klassen miteinander zu vergleichen bedeutet, dass das Schulschicksal eines Kindes in erheblichem Masse von der Situation der Klasse abhängt, in die es mehr oder weniger zufällig hineingerät und, dass Zensuren für überregionale Selektion absolut untauglich sind.“<sup>43</sup>

Der kriteriumsorientierte Ansatz dagegen misst die Leistung der Lernenden in Bezug auf ein vorher festgelegtes Kriterium wie Bildungsstandards und Lernziele. Er orientiert sich an einem Curriculum oder an einem im Lehrbuch festgelegten Lernziel und legt fest, dass möglichst alle Schüler das Lernziel erreichen. Dieser Ansatz wurde heftiger Kritik ausgesetzt.

---

<sup>40</sup> LADO, R. ( 1971, S. 178)

<sup>41</sup> CAROLL, J. (1968, S. 58)

<sup>42</sup> LEDERER, A. ( 2008)

<sup>43</sup> INGENKAMP, K. (1983, S. 59)

Besonders im Zusammenhang mit der Curriculumsdiskussion gewann die kriteriumsorientierte Leistungsmessung mehr an Bedeutung.

#### **1.2.1.4 Die zielerreichende Leistungsbeurteilung**

In der zielerreichenden L.B. wird eine einzige „Minimalkompetenz“ festgelegt und werden die Lernenden danach beurteilt, also nach „Lernziel erreicht – Lernziel nicht erreicht“ eingeordnet. Der zielerreichende Ansatz bewertet Leistungen bezogen auf die Inhalte eines Kurses oder Moduls im Kompetenzkontinuum. Die Alternative zu diesem Ansatz ist die Ergebnisse eines jeden Tests auf das relevante Kompetenzkontinuum zu beziehen. Gewöhnlich mit einer Reihe von Qualitätsabstufungen. Hier ist das Kontinuum das Kriterium „äußere Wirklichkeit“, die sicherstellt, dass Testergebnisse etwas bedeuten.“<sup>44</sup> Seit der Orientierung der Schulnoten an den kriterialen Normen hat sich das Forschungsinteresse mit dem lernzielorientierten Test beschäftigt. GLÄSER erwähnte zum ersten Mal den Begriff „criterion referenced measurement“<sup>45</sup> Auch in der deutschsprachigen Literatur erschienen Ende der 60er Jahre Forschungen zur kriteriumsorientierten Leistungsbeurteilung. Es sei hier auf KLAUER verwiesen, er nannte Tests, die prüfen wie gut ein Schüler das Lehrziel erreicht „lernzielorientierte Tests.“<sup>46</sup> Das erste umfassende deutsche Werk über Tests und Leistungsbeurteilung ist das von dem Autorenkollektiv herausgegebene Buch „Lernzielorientierte Tests“. Das Buch erschien 1972 von KLAUER, FRICKE, HERBIG, RUPPRECHT und SCHOTT.<sup>47</sup> HERBIG schrieb als Ergänzung zu diesem Werk die Konstruktionsprinzipien dieser Tests. Die Grundsätze und die theoretischen Prinzipien wurden auch von KLAUER (1987) und FRICKE (1974) detailliert dargelegt. Die Notenpraxis und die Probleme der Zensurgebung und ihre Einsetzbarkeit wurden auch, wenn nicht ebenso intensiv wie in der englischen Literatur, in der deutschsprachigen Testtheorie diskutiert. Hierzu wird besonders auf „Die Fragwürdigkeit der Zensurgebung“ von INGENKAMP (1983) hingewiesen.

Das Werk gilt noch heute als Standardwerk deutschsprachiger Testforschung. Wegen der frühen Selektion an deutschen Schulen und dem Mangel an Objektivität plädiert INGENKAMP für die Abschaffung der Noten in der Schule und diese durch andere Instrumente und nämlich die formative Leistungsbeurteilung zu ersetzen. Dies führte zu einer Welle von Bemühungen, die Noten aus der Schule zu verbannen einschließlich methodischer

---

<sup>44</sup> TRIM et al., (2001, S. 179-180)

<sup>45</sup> LEDERER, A. (2008)

<sup>46</sup> KLAUER, K. (1972): Das Experiment in der pädagogischen Forschung. Düsseldorf: Schwann.

<sup>47</sup> LEDERER, A. (2008)

Ansätze, die die kriteriumsorientierte Leistungsbeurteilung im Vordergrund rückten. Diese neuen Diagnoseverfahren waren auch der Resultat der neuen Unterrichtsformen wie Freiarbeit.

INGENKAMP stellt in seinem Klassiker „Die Fragwürdigkeit der Zensurengebung“ fest, dass die Zensuren „keine Vergleichsfunktionen bei schulexternen Adressaten erfüllen können und, dass damit unser gesamtes schulisches Berechtigungswesen auf einer „Fiktion“ beruht.“<sup>48</sup> Diesen theoretischen Einsichten über die Fragwürdigkeit der Zensuren stand in der Schulpraxis ein Notenglaube zu. Die Befürworter der Notenabschaffung vertrauen der Notenpraxis nicht, aufgrund der mangelnden Objektivität bzw. der zahlreichen Fehler der Zensurengebung.

In Anlehnung an INGENKAMP meint BECKER (1983, 24), dass die Notenpraxis sowohl in der Schule als auch an der Universität alle Eigenschaften einer Lebenslüge aufweist.<sup>49</sup>

Die Notenpraxisgegner begründen ihre Haltung dadurch, dass die Zensurenpraxis die Qualitätsmerkmale von Testaufgaben und nämlich Objektivität, Reliabilität, und Validität nicht erfüllt. Die vergleichende Leistungsbeurteilung mit Ziffernnoten und individueller Forderung sind ein Widerspruch in sich, da die Orientierung an dem Niveau der Lerngruppe oder an festgelegten Kriterien als Bewertungsreferenz erfolgt. Nennenswert ist auch CUMMINS<sup>50</sup> Modell der kommunikativen Kompetenz. Das die Testformen und die Interpretation der Testergebnisse beeinflusst hat.

OLLER hat in den 70er Jahren eine neue Sicht des Sprachtestens eingeführt; sein Theorem fokussierte auf die psycholinguistischen Prozesse, die im Sprachgebrauch eingebettet sind. Seine Testtheorie orientiert sich nicht an der Überprüfung des Lernaltersprachwissens.

OLLER verwendet den Begriff „Unitary Competence Hypothesis“<sup>51</sup>. Dieser Begriff besagt, dass die Performanz in vielen Testtypen von der selben Kompetenz der Lerner abhängt, deshalb verteidigte OLLER Tests wie der Cloze-Test, der die gleichen Kompetenzen testet wie die produktiven offenen Testtypen. OLLER stellte weiterhin fest, dass der Cloze-Test eine angemessene Alternative für Tests ist, die produktive Sprachkompetenzen messen. Er begründete seine Haltung damit, dass die Lehrkraft bei der Erstellung eines Cloze-Tests grammatisches, lexikalisches und textuelles sowie pragmatisches Wissen integriert, um in der

---

<sup>48</sup> INGENKAMP, K. (1983, S. 192)

<sup>49</sup> BECKER, H. ( 1983) : Zensuren als Lebenslüge und Notwendigkeit, in : Becker, Helmut; v. Hentig, Hartmut : Zensuren. Lüge –Notwendigkeit-Alternativen, Frankfurt, Ullstein .

<sup>50</sup> CUMMINS, J. (1979): Cognitive/academic-language proficiency, linguistic interdependence, the optimum age question and some other matters, in: Working papers on Bilingualism, No. 19, 121-129

<sup>51</sup> OLLER, J.W.JR. 1972. ' Scoring methods and difficulty Levels for close tests of proficiency in ESL. Modern Language Journal 56: 151-8.

Lage zu sein die bestehenden Lücken zu füllen. Zum Erklären ist ein Cloze-Test ein von 400 Wörtern langer Text, der mit einem vollständigen Satz eingeführt wird, nach dem ersten vollständigen Satz werden Wörter systematisch bei jedem fünften oder sechsten Wort getilgt. Weitere Arbeiten danach haben demonstriert, dass der Cloze-Test zur Überprüfung der produktiven Kompetenzen nicht viel geeignet ist. Vielmehr testet der Cloze-Test die sprachlichen Elemente wie Grammatik und Wortschatz. Es wurde auch bewiesen, dass der Cloze-Test zur Überprüfung der kommunikativen Kompetenzen nicht entsprechend ist.<sup>52</sup>

DAVIES (1978) hat ein sehr interessanter Artikel über Sprachtesten veröffentlicht. Seitdem hat sich vieles im „State –Of –The –Art“ des Sprachtestens verändert, zahlreiche Artikel und Bücher wurden herausgegeben, u.a die Beiträge von BACHMAM( 1990/ 1996), ALDERSON (1995) und WEIR (1990) in der theoretischen und empirischen Testforschung.

Wir können feststellen, dass jede Testtheorie an eine bestimmte linguistische Theorie festhält, was meiner Meinung nach zur Vernachlässigung der Objektivität und des Ziels des Testens führt.

#### **1.2.1.5 Der Gemeinsame Europäische Referenzrahmen und Sprachzertifikate**

Der Gemeinsame Europäische Referenzrahmen ist ein umfassendes Standardwerk für Sprachtester. Er bietet Skalen, Checklisten und alles Notwendige für das Erfassen der Sprachkompetenz und legt Beherrschungsniveau der Sprachen fest. Der Referenzrahmen versucht ein Sprachbeherrchungsstandard zu liefern, der für jede unterrichtliche und kommunikative Situation gültig sein kann. Darüber hinaus gibt er Hinweise auf die inhaltliche Beschreibung von Tests und Prüfungen. Er stellt eine gemeinsame Basis dar für die Entwicklung von Zielsprachlichen Lehrplänen, curricularen Richtlinien, Prüfungen, Lehrwerken usw. Er beschreibt umfassend, was Lernende lernen müssen, um eine Sprache für kommunikative Zwecke zu benutzen und welche Kenntnisse und Fertigkeiten sie entwickeln müssen, um in der Lage zu sein, kommunikativ erfolgreich zu handeln.<sup>53</sup>

Er bietet genauere Beschreibungen und Test- Vorgehensweisen in der Praxis, hat auch die Testterminologie aktualisiert und ausführlich beschrieben. Der Referenzrahmen benutzt den Begriff „Beurteilung“ im Sinne von Beurteilung der Kompetenz eines Sprachverwenders, d.h. die Standardsprache kommt in Frage. Er legt großen Wert auf die drei Gütekriterien und ihre

---

<sup>52</sup> NAMARA, T.( 2000, S. 16)

<sup>53</sup> [http://de.wikipedia.org/wiki/Gemeinsamer\\_Europ%C3%A4ischer\\_Referenzrahmen](http://de.wikipedia.org/wiki/Gemeinsamer_Europ%C3%A4ischer_Referenzrahmen) besucht am 09/08/ 2009 um 14h20

Beziehungen zueinander. Darüber hinaus bildet heute der Rahmen die Referenz für die Schulen, die Berufsschulen und Fremdsprachenerwachsenenbildung.

Im Referenzrahmen geht es vor allem um zwei Fragen:

- Inhalt von Tests,
- Kriterien von Validität und Objektivität, kurz um Messung und Beurteilung.<sup>54</sup>

Im Kontext der Sprachtests spielen Sprachzertifikate auch eine Rolle. Diese Prüfungskategorien messen die fremdsprachlichen Sprachkenntnisse. Beispiele dafür sind die europäischen Sprachzertifikate wie „Test Of Englisch als Foreign Language (TOEFEL)“. Sehr gängig werden die Lernerleistungen einer vorgegebenen Skala eingeordnet. Die Skala kann eine Notenskala sein und Stufen von „sehr gut“ bis „ungenügend“ enthalten oder eine Punkteskala, z.B. von 0 bis 100 Punkte. Im algerischen Erziehungssystem sind Skalen von 0 bis 20 gängig. Die Leistung wird dann einer Notenskala assoziiert von „ungenügend“ bis „sehr gut“ oder „hervorragend“.

Ein weiteres Beispiel für Sprachzertifikate ist das „Zertifikat Deutsch“. Die international anerkannte Prüfung zum Zertifikat Deutsch ist das Ergebnis einer Kooperation zwischen dem Goethe-Institut, dem Österreichischen Sprachdiplom (ÖSD), und der Schweizer Erziehungsdirektorenkonferenz (EDK).<sup>55</sup> Es orientiert sich hauptsächlich an den Bedürfnissen von Lernenden in der Erwachsenenbildung im In- und Ausland. Neuere Entwicklungen stellen das gleiche Prüfungsformat auch jugendlichen Lernenden zur Verfügung. In der Regel brauchen Lernende zwischen 350 und 600 Unterrichtsstunden, um die notwendige Sprachbeherrschung für eine erfolgreiche Prüfung zum *Zertifikat Deutsch* zu erlangen. Das *Zertifikat Deutsch* orientiert sich an der Stufe B1 (*Lernen, lehren, beurteilen*) des Gemeinsamen europäischen Referenzrahmens für Sprachen, der international anerkannten Skala des Europarats. Die Prüfung dauert vier Stunden und wird in zwei Teilen durchgeführt, einem mündlichen und einem schriftlichen Teil. Im schriftlichen Teil werden das Leseverstehen und das Hörverstehen überprüft. Das Hörverstehen wird durch den Einsatz einer Hörkassette durchgeführt. Der Lerner antwortet mit „richtig“ oder „falsch“ auf dem Prüfungsblatt.

Im Zusammenhang mit Sprachtesten scheint es mir wichtig, kurz über den Verband „ALTE“ zu sprechen. Die Abkürzung „ALTE“ gilt für „Association for language Testers“ und bezieht sich auf den Verband von Sprachtestern in Europa. Der Verband ALTE zielt auch auf die

---

<sup>54</sup>TRIM et al., (2000, S.07)

<sup>55</sup>[http://fr.wikipedia.org/wiki/Zertifikat\\_Deutsch](http://fr.wikipedia.org/wiki/Zertifikat_Deutsch) besucht am 10/01/2009 um 18h 35

Festlegung gemeinsamer Standards für das Sprachtesten bzw. Testentwicklung- und Testdesign, sowie Testdurchführung.

### **1.3 Testgütekriterien**

Die Überprüfung der verschiedenen Faktoren, die die Performanz der Lerner in Sprachtests beeinflussen ist zentrale Bedingung für die Erstellung und Anwendung von Sprachtests. Ferner stellt die Überprüfung dieser Faktoren den Gegenstand der Sprachtesttheorie.

Beim Testen und Prüfen sind drei Gütekriterien von Belang und nämlich redet man von:

- a- Validität,
- b- Reliabilität, und
- c- Objektivität.

Es wird darüber viel geschrieben und gestritten, diese Testkriterien stehen im Mittelpunkt der Diskussion, und stellen die Achse worum es in der Testtheorie geht.

#### **a-Validität**

Ein Test ist valide wenn er genau das testet, was er testen muss. Die Testaufgaben müssen die Testziele verwirklichen, darum müssen sie entsprechend ausgewählt werden.<sup>56</sup> Dabei eignen sich bestimmte Testsorten für das Testen bestimmter Fertigkeiten.

Dies ist allerdings von der genaueren Bestimmung von Lernzielen nicht zu trennen. Testspezialisten leiten die Validität eines Tests von einem Vergleich zwischen Lernzielen und Testinhalten ab. (BOLTON/ALBERS, 1995, 23)<sup>57</sup>

Der Schwerpunkt der Validität besteht darin, die Gültigkeit und Zuverlässigkeit der aufgrund der Testperformanz erhobenen Beurteilung der Testergebnisse zu beweisen.<sup>58</sup> Problematisch bei diesen Entscheidungen, die aufgrund dieser Tests gemacht werden, ist ob diese Tests wirklich richtige und zuverlässige Informationen liefern d, h. ob der Test wirklich misst, was er zu messen vorgibt.

Faire Beurteilung IST daher nur durch eine valide Informationserhebung möglich. Zweck der Validierung im Sprachtesten ist die Zuverlässigkeit der an der Testperformanz basierten Beurteilungen zu versichern.<sup>59</sup> Die Testvalidierung schließt das Überprüfen der logischen Struktur des Tests ein, wie Testdesign und Testintentionen. Validierung ist also das

---

<sup>56</sup> ALBERS, H/ BOLTON, S. (1995): Fernstudieneinheit Testen und Prüfen in der Grundstufe, München: Goethe-Institut,

<sup>57</sup> Ebd., 1995 S. 23

<sup>58</sup> NAMARA, T. (2000, S.48)

<sup>59</sup> Ebd., S. 48

Überprüfen der empirischen Gültigkeit „ empirical evidence „ des Tests. Abschließend sei hier an BACHMAN(1990, 289) verwiesen, BACHMAN meint, das wichtigste Kriterium, das der Testersteller, bei der Entwicklung und Interpretation von Sprachtests berücksichtigt, sei die Validität.<sup>60</sup> Testtheoretiker haben die Testvalidität sehr intensiv erforscht, dabei werden zahlreiche Validitätstypen unterschieden:

### **a-Facevalidität**

Die Facevalidität ist das holistische<sup>61</sup> Urteil des Laien über einen gegebenen Test. Sie besagt auch, dass der Test für die Verwender ohne besondere Überprüfung oder genauere Validierung valide aussieht bzw. mit dem kommunikativen Sprachgebrauch übereinstimmt. (ALDERSON/WALL/ CAROLLINE, 1995, 167)<sup>62</sup>

### **b-Konstruktvalidität**

Die Interpretationen, die die Lehrkraft aufgrund der Lernerperformenz in einem Sprachtest macht sind nur dann gültig, wenn die Konstruktvalidität des Tests nachgewiesen worden ist. Die Konstruktvalidität bezieht sich auf die Übereinstimmung der Teskompetenz bzw. der Testergebnisse mit dem Bereich der sprachlichen Kompetenz, die der Test misst. In diesem Zusammenhang ist das Konstrukt die Definition oder Festlegung der sprachlichen Kompetenz, die der Test zu messen vorgibt.

„The term construct validity is therefore used to refer to the extent to which we can interpret a given test score as an indicator of the ability (ies), or construct(s), we want to measure.<sup>63</sup> Die Konstruktvalidität ist demzufolge die Begründung der Lernerergebnisse in einem Test im Hinblick auf die überprüfte Sprachkompetenz. Die Konstruktvalidität bezieht sich auch auf die angemessene Festlegung des Testkonstrukts in Verbindung mit dem jeweiligen Bereich des realen Sprachgebrauchs (TUL).<sup>64</sup> Bei der Überprüfung der Konstruktvalidität eines Tests wird der Übereinstimmungsgrad der Testaufgaben mit dem entsprechenden Bereich des Sprachgebrauchs überprüft, und danach die Übereinstimmung des Testkonstrukts mit der

---

<sup>60</sup> BACHMAN,L. (1990, S. 289)

<sup>61</sup> Eine holistische Beurteilung bezieht sich auf ein Urteil, das sich nicht auf wissenschaftliche Kriterien stützt, sondern allgemein und subjektiv erfolgt.

<sup>62</sup> ALDERSON/Wall/CAROLLINE.(1995): Language Test Construction and Evaluation, Cambridge Language Teaching Library, von Cambridge University Press.

<sup>63</sup> BACHMAN/PALMER.(1996, S. 21).



Zielkompetenz. Die Konstruktvalidität kann weiterhin durch die Errechnung der integrierenden Korrelation engl.: (internal correlation) überprüft bzw. durch den Vergleich der erbrachten Lernerergebnisse in den Testteile. Bei sehr niedrigen Korrelationsquoten ist zu fragen, ob die Testaufgaben valide sind oder nicht. Während die Validität das wichtigste Kriterium für die Testdurchführung ist, ist die Reliabilität die unabdingbare Bedingung für Validität<sup>65</sup> BACHMAN(1990).

Testergebnisse, die nicht reliabel sind, können nicht eine Basis für valide Interpretationen und Anwendung bieten. Das zweite Kriterium ist die Reliabilität, d. h. die Zuverlässigkeit der Leistungsmessung; eine Testwiederholung in den gleichen zeitlichen und örtlichen Verhältnissen muss die gleichen Ergebnisse erbringen.

### **Reliabilität**

Die Reliabilität ist ein Kriterium der Testergebnisse. Wenn wir die Reliabilität überprüfen, werden potenzielle Messfehler und ihren Einfluss auf die Testergebnisse überprüft. Bei der Validität werden Faktoren überprüft, die die Performanz in den Tests beeinflussen außer den sprachlichen Kompetenzen, die es im Test zu testen ist. Die Reliabilität betrifft demzufolge die Testergebnisse; ein perfekt reliabler Test sollte ohne Messfehler durchgeführt werden. ( American psychological Association 1985).<sup>66</sup>

Reliabilität besagt auch, dass ein Test die gleichen Ergebnisse ergibt, wenn er wiederholt wird. Die Lernerperformanz kann von äußeren Faktoren beeinträchtigt werden, wie Müdigkeit und Stress. Folglich können die Ergebnisse nicht perfekt « korrelieren ». Testergebnisse können z.B. auch nicht die gleichen sein, wenn die Prüfer verschiedene Punkteskalen (Ratings) für den gleichen Test anwenden. Die Ergebnisse werden in den beiden Fällen nicht reliabler sein. Reliabilität folglich bezieht sich auf die Zuverlässigkeit der Messung über verschiedene Zeiten, Testformen, und andere Charakteristika der Messungsmerkmale des Beurteilungskontexts.<sup>67</sup>

Die Identifizierung von Potenzialquellen der Messfehler impliziert die Beurteilung dieser Tests aufgrund Kenntnisse und Grundsätze der Fehlertheorien.

---

<sup>65</sup> BACHMAN, L. ( 1990, S. 289)

<sup>66</sup> American psychological Association. American Educational Research Association., et al. (1985) Standards for educational and psychological Association.

<sup>67</sup> BACHMAN, L. (1990, S.24)

Das Kriterium der Objektivität besagt, dass der gleichen Leistung die gleiche Punktzahl von Verschiedenen Prüfern gegeben wird. Dieses Kriterium hat einen kritischen Zusammenhang mit ethischen Fragen wie "Gerechtigkeit des Lehrers". (KRUMM/ CHRIST/BAUSCH, 2003, 365)<sup>68</sup> Bei geschlossenen Aufgaben stellt dies allerdings kein Problem, bei offenen Aufgaben schon. Die Diskussion über das Testen zieht zwangsläufig die Beurteilung mit sich.

Mit Prüfen wird die Ermittlung von Leistungen auf der Basis von offenen Aufgaben verstanden. Testen bezieht sich dagegen auf die Vermittlung von Leistungen auf der Basis von geschlossenen Aufgaben wie z. B. Multiple-Choice -Aufgaben. Die Überprüfung der Objektivität regt die ethische Problematik bezüglich Sprachtesten auf, obwohl die Ethik nicht nur bei diesem Kriterium auftritt. NAMARA weist darauf hin, dass die Befürworter des Testens als ethische Aktivität das Testen als wissenschaftliche und technische Aktivität ausschließen. Dies führt zu neuen Einsichten wie die konsequente Validität; die besagt, dass die Lehrkraft bei der Validierung des Tests die gewollten und ungewollten Berechnungen des Testerstellers bei der Einführung des Tests in Betracht nimmt.<sup>69</sup>

Zur Typologie von Testaufgaben kann gesagt werden, dass die Testaufgaben früher nach den vier Fertigkeiten eingeteilt wurden. Jede Fertigkeit wurde einer entsprechenden Testform zugeordnet<sup>70</sup>. Dazu enthalten viele Tests Aufgaben zu den verschiedenen sprachlichen Elementen Wortschatz und zur Grammatik. Heute zutage wird eine Typologie von Testaufgaben, die nach der Testform bzw. nach dem Grad ihrer Offenheit konzipiert, da dieses Kriterium für die Objektivität der Testergebnisse von Belang ist. Dies hat besonders Konsequenzen auf die Gütekriterien: Validität, Reliabilität und Objektivität.

Die Testtypologie ist dazu wichtig, da sie einen engen Zusammenhang mit der Feststellung der Prüfungsziele hat. Man unterscheidet allgemein drei Formen:

### **a-Die offenen Aufgaben**

Bei denen die Fragen „offen“ für die produktive und kreative Leistung des Lernenden sind, wie die Testaufgaben, die z.B. zu einer mündlichen Leistung in einem Prüfungsgespräch veranlassen. Eine Bildinterpretation oder ein Bildkommentar ist eine offene Aufgabe. In der Prüfung Grundstufe 1 des Goetheinstituts sollen die Schüler nach vorgegebenen Stichpunkten

---

<sup>68</sup> KRUMM/CHRIST/BAUSCH. (2003):Handbuch Fremdsprachenunterricht, Tübingen : A. Franke Verlag Tübingen und Basel.

<sup>69</sup> NAMARA, T. ( 2000 , S. 72)

einen Brief schreiben. Diese Aufgaben sind vor allem dazu aufgestellt, die produktive Leistung des Lernenden im Schreiben und Sprechen zu überprüfen. Dennoch bemängelt man in der offenen Aufgaben den bestehenden Nachteil subjektive Bewertungen zu geben, denn sie sind immer der Tatsache ausgesetzt, dass die Schüler zwar inhaltlich richtige Antworten geben, aber grammatisch-orthographisch falsche Strukturen formulieren. Problematisch ist daher ob der Prüfer die volle Punktzahl geben soll oder nicht.

### **b-Die halboffenen Aufgaben**

In einer halboffenen Aufgabe muss der Lernende eine eigen formulierte Antwort geben, in einem begrenzten Kontext. Beispielhaft sind die Ergänzungsaufgaben, in denen der Lerner Einzelsätze mit Lücken erstellt oder Lückentexte ergänzt. Im Cloze- Test z.B. geht es um eine Halboffene Aufgabe.

### **c-Die geschlossenen Aufgaben**

Bei diesen Aufgaben wird die richtige Antwort von dem Kandidaten lediglich erkannt. Ihm "dem Kandidaten" stehen Lösungsvorgaben zur Verfügung und braucht nicht die richtige Antwort selbst zu formulieren. In einer geschlossenen Aufgabe müssen die Lerner unter vorgegebenen Antworten "Distraktoren" die richtige Antwort auswählen.

Die Distraktoren müssen erstellt werden, so dass nur ein einziger Distraktor die richtige Antwort darstellt. Diesem Typ sind folgende Aufgaben untergeordnet:

### **d-Zuordnungsaufgaben**

Bei denen kann der Kandidat entweder passende Satzteile oder einzelne Wörter zuordnen z.B. Antwort-Auswahlaufgaben oder die „Multiple- Choice- Items“

### **e-Alternativaufgaben (True - False Items)**

Hier muss sich der Kandidat entscheiden, ob die vorgegebenen Antworten richtig oder falsch seien, d.h. ob sie inhaltlich in Bezug auf den Text richtig oder falsch sind.

### **f-Antwort-Auswahl –Aufgaben engl.(Multiple-Choice -Items)**

Bei diesem Aufgabentyp stehen dem Kandidaten mehrere Antworten (Destraktoren), wobei der Kandidat die richtige Antwort erkennen muss. Es wird hingewiesen, dass jede Aufgabe

mit einem einleitenden Satz eingeführt ist. Je mehr Distraktoren die Aufgabe enthält, desto geringer die Ratewahrscheinlichkeit beim Lösen.<sup>71</sup>

Als Beispiel einer geschlossenen Aufgabe sei ein Multiple-Choice-Aufgabe vorgegeben:

*Wie geht der Satz weiter?*

Weil wir kein Geld hatten, ....

a)-in Urlaub fahren konnten wir letztes Jahr nicht.

b)-konnten wir letztes Jahr nicht in Urlaub fahren

c)-letztes Jahr konnten wir nicht in Urlaub fahren.

d)-wir konnten letztes Jahr nicht in Urlaub fahren.<sup>72</sup>

Dazu gehören auch Ja/Nein, Richtig/Falsch-Aufgaben.

Diese Aufgabenform eignet sich am Besten zur Überprüfung des Hör- und Leseverstehens und erweist sich als eine durchaus objektive Testform. Es wird diesem Typ zwar bemängelt, dass die Lerner lediglich durch Raten die richtige Antwort finden können. Hierzu sagen ALDERSON und al., (1995, 53) :“True/False or Yes/No items are generally unsatisfactory, as there is a 50% possibility of getting any item right by chance alone „<sup>73</sup>Zuordnungsaufgaben sind auch den geschlossenen Aufgaben untergeordnet. Diese Aufgaben eignen sich zur Überprüfung des Wortschatzes. Bei solchen Aufgaben verläuft die Bewertung ganz maschinell.<sup>74</sup>Im Rahmen der Fremdsprachendidaktik haben viele Schulen die Testtypologie beeinflusst. Als Beispiel sei die Audiolinguale Schule erwähnt. Für die Erklärung LADOs Einsichten als Vertreter dieser Schule wurde hier Raum zugegeben.

Die Form der Aufgabe, sei sie halb, geschlossen oder offen, hat Konsequenzen auf die Fertigkeiten, die sie überprüfen kann. Sie entscheidet über die überprüfte Fertigkeit und beeinflusst die Gütekriterien des Tests bzw. die Auswertungsobjektivität und Reliabilität, d.h. die Messgenauigkeit und die Zuverlässigkeit des Messinstruments. Hierzu meint BOLTON, dass Offene, halboffene und geschlossene Aufgaben nicht nur nach den Fertigkeiten, die sie überprüfen unterteilt werden können, sondern auch hinsichtlich ihrer Auswertungsobjektivität und Reliabilität. Offene Aufgaben erhöhen den Ermessensspielraum

---

<sup>71</sup> BOLTON, S. (1985, S.13)

<sup>72</sup> BOLTON,S./ ALBERS. ( 1995, S.30)

<sup>73</sup> ALDERSON et al., ( 1995,S. 53)

<sup>74</sup> BACHMAN, L. (1990 ,S. 03)

bei der Bewertung der sprachlichen Leistung und haben somit eine geringere Auswertungsobjektivität.<sup>75</sup>

Neuere Einsichten in der Didaktik des Sprachunterrichts setzen den Kompetenzbegriff als Kernbegriff in der theoretischen Diskussion des FSUs und auch in der Testtheorie.

Ziel jedes Sprachenlernens ist der Erwerb von kommunikativen Fertigkeiten zur Bewältigung von realen Kommunikationssituationen. Der Sprachunterricht sollte grundlegend mit der Zielsetzung: "kommunikative Sprachkompetenz" gestaltet werden<sup>76</sup>

Zusammenhängend muss die Lehrkraft die Lerner einer perpetuellen Kontrolle aussetzen, um zu überprüfen ob die Lerninhalte assimiliert wurden oder nicht und das Sprachlernen progressiv und aufbauend zu gestalten

---

<sup>75</sup> BOLTON, S. (1985, S. 14)

<sup>76</sup> TRIM et al., (2000)

## **1.4 Theoretische Grundlagen zum Testen**

### **1.4.1 Sprachtesten und Spracherwerbforschung- und Lehren**

UPSHUR hat 1971 die intrinsische Beziehung zwischen der Leistungsbeurteilung und den Forschungen im Bereich des Spracherwerbs und der Unterrichtspraxis behandelt. Sprachtests sind für die Sprachkompetenzerhebung unerlässlich. Darüber hinaus liefern Sprachtests Informationen über die Effektivität des Lernens und der Lehrwerkeffektivität. BACHMAN/PALMER (1996, 2)

Die Forschungen im Bereich des Sprachtestens können nicht von den anderen didaktischen Bereichen isoliert durchgeführt werden. Sie schließen Forschungen und Entwicklungen in verschiedenen Bereichen ein, die den Spracherwerb und das Sprachlehren betreffen, sowie theoretische Rahmen für die Beschreibung der Sprachlichen Kompetenz. BACHMAN und PALMER betrachten die Testgültigkeit als das wichtigste Kriterium, das es in der Testerstellung zu berücksichtigen sei. Die Testgültigkeit wird nach den sechs Gütekriterien definiert und nämlich: Reliabilität, Validität, Authentizität, Interaktivität, Impact und Durchführbarkeit. Dazu ist die Bedeutung jedes Kriteriums relativ und unterscheidet sich je nach der Testsituation. Hierzu ist es wichtig alle Gütekriterien zu berücksichtigen, dass kein Gütekriterium ignoriert werden könnte.<sup>77</sup> In einem anderen Zusammenhang können neue Einsichten und Reformen in der Unterrichtspraxisforschung die Entwicklung von Sprachtests beeinflussen. Kommunikative Sprachtests sind von den neuen Entwicklungen in der kommunikativen Wende im Sprachunterricht durchaus geprägt.<sup>78</sup>

Entwicklungen im Bereich des Sprachtestens können andererseits Entwicklungen in den didaktischen Forschungen fördern und den Forschungen in den Spracherwerbstheorien Anstoß geben. "... advances in language testing do not take place in a vacuum ; they are stimulated by advances in our understanding of the processes of language acquisition and language teaching and developments in language testing can provide both practical tools and theoretical insights for further research and development in language acquisition and language teaching »<sup>79</sup> Abschließend kann dazu gesagt werden, dass Entwicklungen im Bereich des Sprachtestens es erfordern, Tests anzuwenden, die über das Testen der Sprachfertigkeiten hinaus, auch « authentisch » sind.

### **1.4.2 Sprachliche Fertigkeiten in der kompetenzorientierten Testtheorie**

---

<sup>77</sup> BACHMAN/PALMER. (1996, S. 38)

<sup>79</sup>BACHMAN, L. ( 1990, S. 03)

Traditionell wird die Sprachbeherrschung als die Beherrschung der vier Fertigkeiten aufgefasst: Hören, Lesen, Sprechen und Schreiben. Neue Einsichten in der Testforschung legen die funktionellen Beziehungen zwischen den sprachlichen Fertigkeiten fest und beschreiben dies innerhalb theoretischen Rahmen. Die sprachlichen Fertigkeiten werden nach dem Kanal, indem sie produziert sind aufgelistet bzw. audio- oder visuell und nach dem Mode d, h. als produktive oder rezeptive Fertigkeiten eingeteilt. Folglich werden Sprechen und Schreiben als produktive Fertigkeiten aufgefasst, während Lesen und Hören als rezeptive Fertigkeiten. Fraglich bleibt ob diese Fertigkeiten einfach nach dem Mode unterschieden werden sollten, und ob es sinnvoll ist, alle Sprachgebrauchsformen in rezeptivem und produktivem Mode einzuschränken.<sup>80</sup> Die bestehenden Zweifel über die Inadäquatheit dieses Ansatzes haben zu neuen Auffassungen über die sprachlichen Kompetenzen, die mit den kommunikativen Theorien binden, getrieben. BACHMAN und PALMER (1996, 75) begründen die Inadäquatheit dieses Ansatzes dadurch, dass der Sprachgebrauch unter divergierenden Formen vorkommt.

Eine „Face- To- Face“ Konversation ist mit dem Hören von Nachrichten im Radio nicht identisch; eine Email-Konversation hat viele Gemeinsamkeiten mit einem mündlichen Gespräch als mit dem Lesen einer Zeitschrift.<sup>81</sup> Ferner gehen BACHMAN und PALMER darauf ein, dass der Sprachgebrauch ein mit einem Kontext verbundener Prozess ist. Diese neue Konzeption hatte Konsequenzen auf die traditionell benannten Fertigkeiten, die zur Sprachkompetenz nicht mehr gehören, vielmehr sind Fertigkeiten die kontextualisierte Realisation der Fähigkeit, die Sprache in spezifischen sprachlichen Bereichen einzusetzen. „We would thus not consider language skills to be part of language ability at all, but to be the contextualized realization of language in the performance of specific language use tasks“<sup>82</sup> Es wird nicht mehr auf die Fertigkeit gestützt, sondern auf spezifische Aktivitäten Oder Aufgaben. Statt das Sprechen als abstrakte Fertigkeit zu beschreiben, wird es als eine spezifische Kombination der sprachlichen Fertigkeit und der Eigenschaften des Sprechens als Aufgabe „a Task“ definiert. Dieses Konzept hatte für die Theorie des Sprachtestens Konsequenzen in Bezug auf die Testmethode und Testfacetten. Beim Testen einer sprachlichen Kompetenz werden zuerst die Charakteristika der sprachlichen Kompetenz als „Handlung“ identifiziert. Die Rahmenbedingungen und der Input, die aufgeforderte Reaktion und der Zusammenhang zwischen Input und Reaktion werden berücksichtigt. Seitdem es sich im Sprachtesten um Entscheidungen über die Lerner handelt, muss die Testperformanz dem

---

<sup>80</sup> BACHMAN/ PALMER. (1996, S. 76)

<sup>81</sup>Ebd., S. 76

<sup>82</sup>Ebd., S. 76

Sprachgebrauch in realen sprachlichen Lebenssituationen entsprechen. Dafür haben BACHMAN und PALMER(1996) einen theoretischen Rahmen, in dem sie die Testperformanz von den Sprachbereichen nicht trennen; die geprüfte Person wird als ein Sprachverwender im Kontext des Sprachtests, der Sprachtest wird wiederum als ein spezifischer Kontext innerhalb des realen Sprachgebrauchs angesehen.

#### **1.4.2.1 Kommunikative Sprachkompetenz und Sprachtestforschung**

Die Forschung in der Sprachtesttheorie soll auf der Grundlage einer Sprachkompetenztheorie erfolgen.<sup>83</sup>

Die Sprachkompetenz stellt das Testobjekt dar. Laut BACHMAN müssen Sprachtests dem Testzweck entsprechend konzipiert werden. Darum muss die Testentwicklung auf einer klaren Definition des Testzwecks und der getesteten Kompetenz beruhen. Darauf aufbauend hat BACHMAN (1990) einen theoretischen Rahmen für das Sprachtesten auf der Basis einer klaren Festlegung der kommunikativen Kompetenzen aufgestellt. (KLIEME und al , 2003, 59) definieren die kommunikative Kompetenz und erläutern dabei das Zusammenspiel der kommunikativen individuellen Facetten wie folgt: „Die so verstandene Fremdsprachenkompetenz drückt sich darin aus, wie gut man kommunikative Situationen bewältigt (Handeln und Erfahrung), wie gut man Texte verfassen kann, aber auch in der Fähigkeit, grammatische Strukturen korrekt aufzubauen und bei Bedarf zu korrigieren (Fähigkeit und Wissen), oder in der Intention und Motivation, sich offen und akzeptierend mit anderen Kulturen auseinander zu setzen.“<sup>84</sup> Die individuelle Ausprägung einer Kompetenz wird nach WEINERT (2001) von verschiedenen Facetten bestimmt und nämlich von: Fähigkeit, Wissen, Verstehen, können, Handeln, Erfahrung und Motivation.<sup>85</sup>

In diesem Zusammenhang weisen kommunikative Sprachtests zwei wichtige Merkmale auf:

- Sie verwickeln den Lerner anhand des Performanztests in einem kommunikativen Akt,
- sie richten sich an kommunikative Situationen der realen Welt.<sup>86</sup>

Der theoretische Rahmen von BACHMAN (1990) teilt die kommunikative Sprachkompetenz in drei Bereiche:

- Die Sprachkompetenz.
- Die strategische Kompetenz.
- Die psychologische Kompetenz.

---

<sup>84</sup> KLIEME, E und al.(2003):Zur Entwicklung nationaler Bildungsstandards. Bonn: BMBF.

<sup>85</sup>WEINERT, F. ( 2002): Leistungsmessungen in Schulen, Weinheim.

<sup>86</sup>NAMARA, T. (2007)



Die Kommunikative Sprachkompetenz wird als der Gebrauch bereits erworbener Sprachkenntnisse in angemessenen sprachkommunikativen Situationen verstanden. CANDLIN (1986, S. 40) definiert die kommunikative Kompetenz als der Einsatz von organisiertem Wissen, um neue Probleme in der Kommunikation zu lösen, d. h. kreativ zu kommunizieren. "In sum ...a coming together of organized knowledge to solve new problems of communication that do not have ready-made and tailored solutions."<sup>87</sup> Der theoretische Rahmen von BACHMAN und PALMER (1996) war eindeutig von CANALE und SWAIN beeinflusst. Die kommunikative Sprachkompetenz lässt sich in drei Bereiche aufteilen: Sprachkompetenz, strategische Kompetenz und psycho- physiologische Mechanismen.

#### **1.4.2.2 BACHMANs Modell der Sprachkompetenz**

Im Rahmen seiner testtheoretischen Forschung hat BACHMAN (1990) ein Modell der kommunikativen Kompetenz erstellt und den Theoretischen Rahmen der Sprachtestforschung mit den kommunikativen Kompetenzen verknüpft. BACHMANs Modell der Sprachkompetenz umfasst zwei Hauptbegriffe:

**Das Sprachwissen** umfasst zwei bereite Kategorien: organisatorisches Wissen und pragmatisches Wissen.

Obwohl viele Sprachtests nur einen einzigen Aspekt der Sprachkompetenz hervorheben, sollten die Testaufgaben im Zusammenhang mit den Kompetenzen, die der Testteil misst bearbeitet werden.

**Das organisatorische Wissen** wird bei der Kontrolle der formellen Struktur der Sprache eingesetzt und um grammatisch richtige Äußerungen zu bilden und deren Organisation bei der Produktion von mündlichen und schriftlichen Texten. Organisatorisches Wissen umfasst wiederum Grammatikalisches Wissen, und sprachliche Elemente wie Wortschatz, Syntax, Phonologie und Graphologie.

**Das textuelle Wissen** wird bei dem Verständnis des Textes oder Inputs eingesetzt bzw. bei Abschnitten, die mehr als drei Sätze enthalten. Es werden zwei Arten des textuellen Wissens unterschieden:

---

<sup>87</sup> CANDLIN, C. N. (1986): Explaining communicative competence limits of testability, Stansfield.

**Das Kohäsionswissen** wird bei der Produktion und Verständnis explizit markierter Beziehungen zwischen den Sätzen in einem geschriebenen Text oder in einem Gespräch eingesetzt.

Wird von dem Sprachverwender bei der Produktion und organisatorischen Entwicklung geschriebener Texte und Konversationen eingesetzt.

**Das Pragmatische Wissen** der Sprache bezieht sich auf die Produktion von Sätzen oder Äußerungen im Zusammenhang mit den Intentionen der Kommunikationspartner.

Das pragmatische Wissen teilt sich wiederum in:

**Das funktionale Wissen** bezieht sich auf die Beziehungen zwischen Sprachproduktion und den Sprachverwender; die Äußerung „could you tell how to get to the post office?“ wird als Aufforderung den Weg in die Post zu zeigen verstanden und nicht als Ja/Nein Frage.<sup>88</sup> Eine Antwort mit nur „Ja“ oder „Nein“ ist in diesen kontextuellen Bedingungen unangemessen. Das funktionale Wissen umfasst vier Kategorien der sprachlichen Funktionen: manipulatives, instrumentales und imaginatives Wissen.

### 1.4.2.3 Zur Sprachtestforschung

Einer der Hauptanliegen der Sprachtestforschung ist die Überprüfung der Faktoren, die die Performanz in Sprachtests beeinflussen. Die Forschungen haben bewiesen, dass die Testmethode die Performanz in den Sprachtesten beeinflusst, sowohl bei einer Facette in dem Sprachtest oder wegen der Mitwirkung dieser Facetten oder Seiten.<sup>89</sup>

Viele Forscher haben diese Facetten im Cloze-Test erforscht und damit das theoretische Wissen über den Cloze-Test als Technik des Sprachtestens bereichert und erweitert<sup>90</sup>

---

<sup>88</sup> BACHMAN, L. (1990, S.69)

<sup>89</sup> BACHMAN, L. (1990, S. 69)

<sup>90</sup> Siehe. OLLER, 1975; ALDERSON, 1979b ; OHNMACHT / FLEMING, 1974; BACHMAN, 1982

Es wird festgestellt, dass die Validität der Testergebnisse und ihre Interpretation bzw. Beurteilung beeinträchtigt wird wenn die Testperformanz durch andere Faktoren als die zu überprüfte Kompetenz beeinflusst wird. Die Testmethode und die Interaktion zwischen dem am Test teilnehmende und den individuellen Charakteristika können die Ergebnisse des Tests beeinflussen. Die Erforschung dieser Faktoren bietet relevante Grundlage und Informationen, welche Testformen am besten mit den individuellen Charakteristika korrelieren, um sie im Testdesign zu berücksichtigen. Einige Schüler können zum Beispiel bessere Ergebnisse in einem Test erbringen in Form eines mündlichen Gesprächs als in einem Sprachlabor, Andere performen besser in einem Test in dem sie individuelle Sätze ergänzen sollen als in einem Cloze-Test<sup>91</sup> Diese Testformen stellen das „Wie“ des Testens dar.

Laut BACHMAN ( 1990) kann die Testperformanz sowohl durch die Testmethode, aber auch durch die Lernerkompetenz beeinflusst werden. Die Testperformanz kann auch von individuellen Faktoren beeinflusst werden, die mit der Sprachkompetenz keinen Bezug haben. Dies schließt die kognitiven und affektiven Eigenschaften ein, wie auch ihr allgemeines Weltwissen und andere Faktoren wie Geschlecht und Muttersprache oder erzieherische Hintergründe.<sup>92</sup> Individuen mit verschiedenen Charakter und Hintergründe können verschiedenartig mit Testtypen agieren. Auch Vorerfahrungen mit einer gegebenen Testform können die Testergebnisse beeinflussen. die Teilnehmer entwickeln Teststrategien, die die Testerfahrung erleichtern.

Es ist unverkennbar, dass die Testmethode Auswirkungen auf die Performanz der Lerner haben könnte; dies ist in der Testtheorie unter dem Terminus „Testeffekt“ bekannt.

Darüber hinaus sind einige Testmethoden zur Überprüfung bestimmter Kompetenzen und Fertigkeiten geeigneter; eine Multiple –Choice Aufgabe dient nicht dazu, die Aussprache der Lerner zu überprüfen.<sup>93</sup> Testforscher können aber nicht eine vorgegebene Testmethode zur Überprüfung einer einzigen Sprachkompetenz indizieren, obwohl viele von ihnen die Auswirkungen der Testmethode auf die Lernerkompetenz bewiesen haben.<sup>94</sup>

---

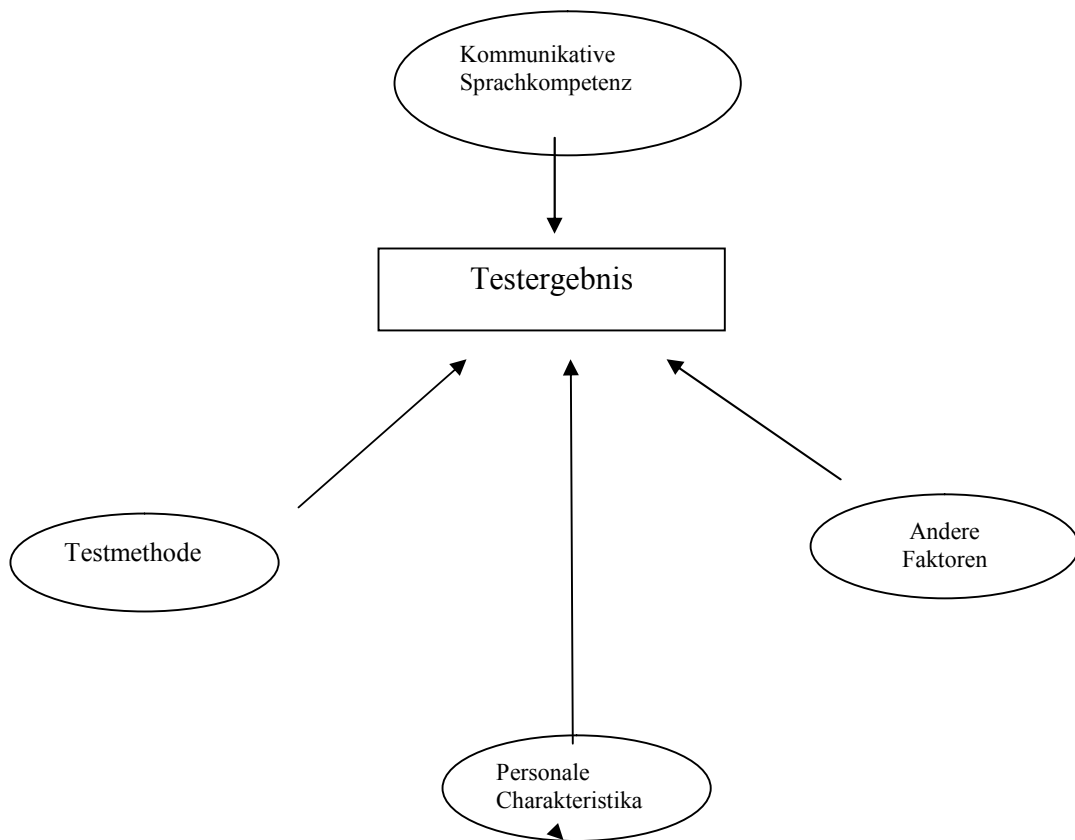
<sup>91</sup> BACHMAN, L (1990, S. 113-115)

<sup>92</sup> Ebd., S.111

<sup>93</sup> ALDERSON et al., (1995, S.46)

<sup>94</sup> BACHMAN/PALMER. (1996, S. 46)

Abb.1 Faktoren, die die Testperformanz beeinflussen nach Bachman



Quelle: BACHMAN, L (1990, 165). *Fundamental Considerations in Language Testing*. Oxford University Presse.

Es ist wichtig die Charakteristika der Testaufgabe zu bestimmen und dies aus verschiedenen Gründen: Die Angemessenheit der Testaufgabencharakteristika mit dem realen Sprachgebrauch in einer bestimmten sprachlichen Situation bestimmt den Grad der Testauthentizität. Der theoretische Rahmen von BACHMAN und PALMER (1996) zielte u.a. auf die Beschreibung der Beziehung zwischen Sprachverwendung im realen Leben und Testaufgaben. Im Nachfolgenden wird dieser theoretische Rahmen in Bezug auf diese Beziehung detailliert umgerissen. Laut CAROLL, J (1993)<sup>95</sup> ist eine Aufgabe jede Aktivität, in der sich eine Person, in einer angemessenen Situation einsetzt, um eine spezifische Klasse von Zwecken zu leisten. Er nennt zwei Bedingungen, die sowohl für die Sprachverwendung als auch für das Sprachtesten relevant sind: Der Lerner soll über die aufgeforderte Arbeit und deren Ergebnis bewusst sein. Der Lerner sollte über die Mess- und Beurteilungsverfahren seiner Performanz teilweise bewusst sein. Sprachliche Aufgaben sind allgemein:

- eng mit spezifischen Situationen verbunden.<sup>96</sup>
- zielorientiert.
- Schließen die objektive Teilnahme der Sprachverwender (engl.: language users) ein.

Die sprachliche Äußerung wird zu einem bestimmten Zweck geäußert (tache linguistique).

Nach der Sprachakttheorie ist eine Äußerung (language task) immer mit einem spezifischen Akt verbunden. Es muss nicht übersehen werden, dass die „sprachliche Aufgabe“ (engl. „Language task“) zielorientiert ist und mit einer sprachlichen Situation verbunden ist. Sprachverwendung, von Natur aus, hängt von bestimmten Situationen ab und wird von diesen Situationen hervorgerufen.

In dem Sprachgebrauchsbereich (LUD) wird primär zwischen zwei Bereichen unterschieden:

- den sogenannten „real-life“ Bereichen in denen die Sprache hauptsächlich für kommunikative Zwecke verwendet wird
- und einem zweiten Bereich in dem die Sprache für spezifische Zwecke, die das Lehren und Lernen der Sprache betreffen, verwendet wird.<sup>97</sup>

---

<sup>95</sup> CAROLL, J. (1993) Practical unification-based parsing of natural language, Technical Report No. 314, Computer Laboratory, University of Cambridge. PhD Thesis

<sup>96</sup> Vgl. BACHMAN / PALMER. (1996, S.76)

<sup>97</sup> Ebd., S. 44

Die Performanz in einem Sprachtest hängt von der Angemessenheit dieser Kompetenz zu bestimmten Sprachsituationen über den Test hinaus. Die Entsprechung der Testergebnisse mit dem Testkonstrukt wird anhand der Testkonstruktvalidität überprüft. Die Wichtigste Frage bei der Interpretation der Testergebnisse ist: Zu welchem Grad ist diese Interpretation richtig und Aussagen über die Kandidatenkompetenz dabei erlaubt? Und wie kann diese Interpretation begründet werden? Es ist notwendig das Konstrukt zu definieren. Der Test sollte so konstruiert werden, dass er die getestete Sprachkompetenz wirklich misst. Das Konstrukt ist laut BACHMAN und PALMER(1996) eine spezifische Beschreibung der Kompetenz, die die Basis für den Test und Ergebnisbeurteilung des jeweiligen Tests stellt.<sup>98</sup> Die Wichtigkeit der Konstruktvalidität lässt sich dadurch begründen, dass sie bei der Testergebnisinterpretation eine unabdingbare Bedingung ist. Es ist erwähnenswert, dass die Testvalidierung nie absolut richtig sein kann; ein Test kann nicht immer absolut valide sein. Da die Gütekriterien konfluieren, ist es wichtig für die Testersteller eine Gewichtung der Gütekriterien, durch die sorgfältige Entwicklung des Tests zu versichern. <sup>99</sup>

Eine Frage, die sich häufig Lehrer und Testentwickler (Testdesigner) stellen, betrifft die geeignetste Form eine bestimmte Fertigkeit zu testen. Dies stellt eine heikle Frage dar, da die Aufgabenform die Performanz an dem Sprachtest beeinflusst. Viele Forschungen haben die Auswirkung der Aufgabenform auf die Kandidatenperformanz bewiesen. BACHMAN/PALMER (1996) meinen, die Charakteristika der Testaufgaben( die Testrubrik, die Arbeitsanweisungen, Aufgabenform) beeinflussen die Testergebnisse, dass es virtuell kein Test gibt, der Daten lediglich über die zu testende Kompetenz erlaubt. Da wir die Konsequenzen der Testaufgabenform nicht ausschließen können, müssen sie (die Konsequenzen) verstanden und kontrolliert werden, dass die Testaufgabe die Gütekriterien erfüllt.<sup>100</sup> Die Testaufgabencharakteristika wurden auch in dem Modell von BACHMAN und PALMER (1996) skizziert, um eine optimale Angemessenheit zwischen der Zielsprache und der Testaufgabensprache zu garantieren. Darüber hinaus wollten sie Testformen, die die Parameter der Lernergruppe respektieren erstellen, damit sie am besten in dem Test leistet. Bisher sind die Testcharakteristika für Testentwickler die einzigen Faktoren, die unter Kontrolle sind.

---

<sup>98</sup> BACHMAN/PALMER. (1996, S. 21)

<sup>99</sup> Ebd., S. 22

<sup>100</sup> Ebd., S. 46

Bei der Durchführung eines Tests können andere Faktoren die Performanz beeinflussen u.a. individuelle Eigenschaften der Lerner und eventuelle mentale und psychische Veränderungen.

Die Charakteristika der Testaufgaben, sowie die Aufgabenform (offen, geschlossen, halb-offen) können die Performanz der Lerner beeinflussen, deshalb muss der Test sorgfältig erstellt werden, da die Entscheidungen, die aufgrund der Testergebnisse getroffen werden, Menschen direkt betreffen. Die Testqualität, die die Testersteller vor allem betreffen muss, bezieht sich auf die Gütekriterien: Reliabilität und Objektivität.

„Since the decisions we make will affect people, we must be concerned about the Quality – Reliability and validity of our test results. »<sup>101</sup>

In dem erzieherischen Bereich hängt die Evaluation mit Entscheidungen zusammen, die die Einstufung, Diagnose und den Lernfortschritt betreffen. Entscheidungen, die Sprachcurricula betreffen beziehen sich auf Merkmale wie die Angemessenheit, die Effizienz und Effektivität des Curriculums. Verschiedene Testarten können hinsichtlich fünf Facetten unterschieden werden:

- Einstufung-Diagnose und Lernfortschritt,
- des Zwecks, wozu sie erstellt sind (engl. achievement, profeciency, Aptitude),
- der Standards, die bei der Interpretation der Testergebnisse verwendet werden (Norm. Kriterium)
- und hinsichtlich der Benotungsverfahren ( scoring procedures).

Wie oben schon erwähnt wurde, ist Einer der Hauptanliegen der Sprachtestforschung die Überprüfung der Faktoren, die die Performanz in den Sprachtests beeinflussen. Die Forschungen haben bewiesen, dass die Testmethode die Performanz in den Sprachtests beeinflusst, sowohl in Sprachtests oder wegen der Interaktion zwischen diesen Facetten oder Seiten.<sup>102</sup> Viele Forscher haben diese Facetten im Cloze- Test erforscht, und haben damit das theoretische Wissen über den Cloze-Test als Technik des Sprachtestens bereichert<sup>103</sup>. Es wird auch festgestellt, dass die Testperformanz durch andere Faktoren als die überprüfte Kompetenz beeinflusst wird. Und, dass die gleichen Faktoren (Testform, Müdigkeit,...) die Validität der Testergebnisse und ihre Interpretation( Beurteilung) beeinträchtigen. Wie die

---

<sup>101</sup> BACHMAN, L. (1990)

<sup>102</sup> Ebd., S. 155

<sup>103</sup> Siehe: OLLER 199/ALDERSON1979 b/ OHNMACHT/FLEMMING1974 und BACHMAN 1982

Testmethode und die Interaktion zwischen den Probanden und den individuellen Charakteristika.

Die Erforschung dieser Faktoren bietet relevante Informationen, welche Testformen am besten mit den individuellen Charakteristika korrelieren, um diese im Testdesign zu berücksichtigen. Einige Schüler können bessere Ergebnisse in einem Test in Form eines mündlichen Gesprächs erreichen als in einem Sprachlabor.<sup>104</sup> Andere ziehen einen Test vor, in dem sie individuelle Sätze ergänzen sollen als ein Cloze-Test.

Die Beschreibung der Testfacetten stellt eine wichtige Basis für die Testerstellung, da sie die Testperformanz beeinflussen. BACHMAN meint, dass diese Facetten in der Testentwicklung und der Testdurchführung einen bedeutenden Impact haben.<sup>105</sup> CLARK (1972) verwendet den Terminus „Stimulus“ und bezeichnet damit alle angebotenen mündlichen und schriftlichen Materialien, mit denen der Lernende in der Testsituation behandeln muss, er unterscheidet die Stimulus- Modalitäten je nach ihrer Darbietungsform, sei sie schriftlich oder mündlich oder in Form eines Bildes. CLARK (1972) definiert die Antwort „Reaktion“ folgendermaßen: „any physical activity on the part of the student in reaction to the stimulus materials“<sup>106</sup> Er unterscheidet zwischen den freien Antworttypen, in denen der Lerner eine schriftliche oder mündliche Produktion leistet und berücksichtigt dabei ob diese Produktion in der Fremdsprache oder in der Muttersprache erfolgt.

#### **1.4.2.4 Zu den Testfacetten**

Es gibt sechs Kategorien von Testfacetten:

- a- Die Testumgebung.
- b- Die Testrubrik.
- c- Die Natur des Inputs.
- d- Die aufgeforderte Antwortform.
- e- Die Beziehung zwischen Input und Reaktion im Test.<sup>107</sup>
- f- Die Testumgebung:

Die Testumgebung umfasst: die Vertrautheit des Ortes, des Personals, die Testzeit oder die zur Verfügung stehende Zeit. Die Testteilnehmer leisten besser mit vertrautem Material wie Papier und Kugelschreiber als mit unvertrautem Material wie Computers. In mündlichen

---

<sup>104</sup> BACHMAN, L. (1990, S. 113)

<sup>105</sup> Ebd., S. 115

<sup>106</sup> CLARK, D. (1972, S. 27): Foreign Language Testing: Theory and Practice. Philadelphia, Pa: Center for Curriculum Development, Inc.



Tests können die Testteilnehmer verschiedene Ergebnisse haben, wenn sie von einem unbekanntem Prüfer geprüft werden.<sup>108</sup> Die Testzeit beeinflusst die Performanz in den Sprachtests wie z.B. dass die Teilnehmer besser oder minder in Tests arbeiten, die sie am frühen Morgen haben. Die Testperformanz kann wegen physikalischen Bedingungen wie Lärm oder Humidität im Testraum und seiner Umgebung beeinflusst bzw. beeinträchtigt werden.

Die Teststruktur umfasst Faktoren wie die Testorganisation, die Zeitallokation und die Arbeitsanweisungen. *Die Testorganisation* bezieht sich auf die Testteile (Testaufgaben) (engl.: „Parts“, oder Items“). Die Verteilung dieser Testaufgaben kann die Performanz in den Sprachtests beeinflussen.

**Die Aufgabenorganisation** (engl.:“ Saliency of parts“): Einige Tests bestehen aus einer bestimmten Zahl von Subtests, seitdem der Test eine Gruppe von Tests oder Testteile ist. Diese Testteile werden durch Sätze gesondert, die den Testzweck und Inhalt erklären oder die getestete Kompetenz angeben. Diese Sätze sind die Arbeitsanweisungen. In der ersten Aufgabe ist der Kandidat z.B. aufgefordert ein grammatisches Problem zu lösen, während er in der Aufgabe 2 einen korrekten Aufsatz zu einem gegebenen Thema schreiben sollte. In anderen Tests gibt der Aufgabentitel schon die Testinformationen an wie „Leseverstehen“, „Wortbildung“ usw.

„The parts of the ILR oral interview, for example warm-level check, probe, and wind-down are not explicitly marked, and are not intended to be noticed at all by the test taker.“<sup>109</sup> Die Sequenzen, in denen die Teilaufgaben präsentiert sind, spiegeln in mancher Hinsicht die Intentionen der Testersteller wieder und schließen die Kontrollabsicht des Testerstellers auf die Lernerreaktion ein. Die Testprobanden reagieren auf diese Aufgabenverteilung verschieden. Einige gehen auf die Testaufgaben direkt ein, ohne den Schwierigkeitsgrad der Aufgaben zu berücksichtigen, Andere setzen eine Strategie ein, d. h. sie gehen auf die Testaufgaben vom Einfachen zum Schweren und bedienen sich der ihnen stehende Zeit strategisch.

Bei der Messung der Lernerleistungen in den Sprachtests entscheiden die Testaufgaben weniger oder mehr im Testergebnis mit. Die Aufgaben in denen die Testkandidaten korrelierende Noten bekommen, entscheiden schwach im Testergebnis mit. Testaufgaben, die

---

<sup>108</sup> BACHMAN, L. (1990, S. 118)

<sup>109</sup> Ebd., S.180

mit den anderen Aufgaben innerhalb des Tests wenig korrelieren, entscheiden also stärker im Testergebnis mit.

Die Zeit, die den Kandidaten zur Verfügung steht, kann die Testergebnisse beeinflussen, wenn die Testzeit z. B. nicht zur Lösung der Testaufgaben genügt<sup>110</sup>. Testforscher legen fest, dass die Testzeit angemessen sein soll; sie soll weder zu lang noch für die Lösung der jeweiligen Aufgaben ungenügend sein. Es wird aber allgemein festgestellt, dass bezüglich der Testzeit, ist der Test reliabler wenn die zur Erstellung eingesetzte Zeit genügend ist.<sup>111</sup> Testaufgaben, die objektiv getestet werden können, sind potenziell mehr reliabel als Testaufgaben, die z.B. subjektive Beurteilung benötigen.

Die Arbeitsanweisungen spielen eine bedeutende Rolle in der Optimierung der Kandidatenperformanz. Die Testkandidatenperformanz hängt vor allem von der Klarheit, auf der die Arbeitsanweisungen erstellt werden. Die Sprache in der die Arbeitsanweisungen präsentiert werden, kann die Muttersprache oder die Zielsprache sein. Die Arbeitsanweisungen werden auch in den beiden Sprachen (Muttersprache und Zielsprache) präsentiert. Sie können visuell oder auditiv dargeboten werden.

Die Arbeitsanweisungen allgemein bestimmen die Testverfahren, d. h. die Antwortform, ob z.B. die Schüler an dem Testblatt antworten sollen oder an einem separaten Blatt und wie sie die Alternativen wählen und wie sie die Antworten auf dem Antwortblatt schreiben, und ob sie nur bei dem Antworten auf die Aufgaben die Lücken füllen müssen. Die Arbeitsanweisungen definieren die verschiedenen Testteile oder Testparts. Diese Spezifizierung schließt die Antwortform ein (Selektion oder Konstruktion), d. h. die Kandidaten aus vorgegebenen Distraktoren die richtige Antwort wählen oder die Antwort selbst formulieren sollen. Die Testperformanz ist also zu erhöhen durch den Einsatz einfacher Testaufgaben, mit klaren Arbeitsanweisungen, die bestimmen, wozu der Lerner aufgefordert ist.

Der Typ der aufgeförderten Antwort bezieht sich auf die Form in der der Kandidat seine Leistung aufbringt; er kann die richtige Antwort wählen oder die falsche Alternative identifizieren. Die zweite Form ist die „constructed-responses“<sup>112</sup>, in der der Testkandidat seine Antwort selbst formulieren muss, d. h. sie bezieht auf die Produktion einer sprachlichen Stichprobe als Reaktion auf einen Input, d. h. die sprachliche Reaktion auf ein Inputmaterial.

---

<sup>110</sup> BACHMAN, L. (1990, S. 123)

<sup>111</sup> UPSCHUR, J / GENESE, G. ( 1996, S.199)

<sup>112</sup> POPHAM, W. (1978): Criterion- Referenced Measurement. Prentice Hall.

Viele Forschungen vertreten die Hypothese, Multiple-Choice Aufgaben seien einfacher als selbst formulierte Antworten.<sup>113</sup>

#### **1.4.2.5 Zu den Arbeitsanweisungen**

Die Arbeitsanweisung ist ein Teil von dem Test, den die Lernenden verstehen müssen. Die Hauptrolle der Arbeitsanweisung ist folglich das Lernerverständnis des Testinhalts zu sichern. Die Arbeitsanweisungen sollten gegeben werden, so dass sie von den Lernenden verstanden werden. Während des Tests müssen die Arbeitsanweisungen so gegeben werden, dass der Testkandidat sie vollständig aufnimmt. Dies schließt auch die kontextuellen „materiellen“ Bedingungen bei der Durchführung des Tests ein.

Die Performanz, die die Prüflinge in einem Test aufweisen, ist weitgehend von der Teststruktur und Testform abhängig. Es ist wichtig, dass die Lernenden, die an der Prüfung teilnehmen, klar verstehen, was sie im Kontext dieser Prüfung leisten sollten bzw. wozu sie aufgefordert sind und auch wie werden ihre Antworten evaluiert werden. Die Testarbeitsanweisungen sind besonders wichtig, da die Testkandidaten nur durch sie verstehen wozu sie aufgefordert sind und wie sie die Aufgaben lösen werden. Die Arbeitsanweisung soll vier Sätze enthalten:

- einen ersten Satz, wodurch die Intention und Zielstellung des Tests klar werden (Testzweck).
- einen zweiten Satz, in dem der Lernende über die sprachlichen Kompetenzen, die der Test misst erfährt.
- Erwähnung der Aufgabenprozeduren und auch der Beurteilungsmethode „scoring methode“

<sup>114</sup>

Zusammenfassend, dienen Testarbeitsanweisungen laut BACHMAN und PALMER (1996) einem wichtigen Ziel: der Motivierung der Lernenden ihr Bestes zu tun. Vielmehr dienen effektiv erstellte Arbeitsanweisungen dazu, die Lernenden zu versichern, dass der Test relevant, angemessen, und objektiv ist. Diese Gliederteile sind nicht unbedingt auf einem Schluck den Lernenden anzugeben. Allgemein werden Arbeitsanweisungen vor der Testdurchführung eingegeben und werden auch mündlich während die Lernenden den Test

---

<sup>113</sup> SAMSON, D. M. M. (1983): *Rash and Reading* in J. Van Weeren (Ed): *practice and problems in language testing*. Arnhem: CITO.

<sup>114</sup> BACHMAN/ PALMER. (1996, S.182)

machen von den Testadministratoren gelesen. Zusätzlich können spezifische Arbeitsanweisungen bei jeder Testaufgabe gegeben werden.<sup>115</sup>

Die Arbeitsanweisungen müssen keine zusätzliche Quelle für Schwierigkeiten in dem Test stellen, sondern den Lernern beim Verstehen des Inputs helfen. Ziel des Tests ist also die Überprüfung einer bestimmten sprachlichen Kompetenz und nicht wie gut die Lerner die Arbeitsanweisungen verstehen.

„We want to discover what students have learned, not how well they can understand test instructions or how well they can guess what they are supposed to do.“<sup>116</sup> Es ist besser diese Arbeitsanweisungen in der Muttersprache zu schreiben, wenn die Lernenden in Bezug auf die Muttersprache eine homogene Lerngruppe darstellen. Wenn die Lernenden einen heterogenen sprachlichen Hintergrund aufweisen, werden die Arbeitsanweisungen in diesem Fall in der Zielsprache erstellt. (UPSHUR, 201)<sup>117</sup>

UPSHUR und GENESE haben die Tests die in Klassenzimmer durchgeführt werden beschrieben. In einem rein erzieherischen Kontext sollen die Arbeitsanweisungen angeben:

- die verfügbare Zeit zur Lösung der Aufgaben .
- Zahl der Aufgaben „Items“.
- Spezielle Testbedingungen, die normalerweise in der Klasse erlaubt sind wie z.B das Zurückgreifen zu Wörterbüchern oder Büchern usw. oder ob der Lernende mit den anderen Mitschülern im Kontakt kommen kann.

-Erwähnung der Inputs, mit denen der Lernende handeln wird, z.B. Texte, Bilder, Gespräche und die erwartete Antwortform wie z.B. Gespräch, Schreibantwort. UPSCHUR weist darauf hin, dass diese Arbeitsanweisungen, wenn die Lernenden den Test schon durchgeführt haben fakultativ werden.

Nach ihm kann ein Beispiel (Muster) die Funktion der Arbeitsanweisungen erfüllen. Die Arbeitsanweisungen sind nicht in jedem Test unbedingt einzusetzen; jedem Test sind bestimmte Arbeitsanweisungen angemessen.<sup>118</sup> UPSHURs Einsichten bezüglich Arbeitsanweisungen in erzieherischen Institutionen können in den folgenden Punkten zusammengefasst werden:

- Die Arbeitsanweisungen können in der Muttersprache gegeben werden, oder in der Zielsprache.
- Es ist besser die Arbeitsanweisung in der Muttersprache anzugeben.
- Die Arbeitsanweisungen müssen nicht in der Zielsprache gegeben werden.
- Ein Muster kann in manchen Aufgaben die Rolle der Arbeitsanweisungen erfüllen.

---

<sup>115</sup>Ebd., S. 182

<sup>116</sup> UPSHUR/GENESE. (1995, S. 201)

<sup>117</sup> Ebd., S. 201

<sup>118</sup> Ebd., S. 203

Die Arbeitsanweisungen werden nicht zwangsweise in geschriebener oder mündlicher Form eingegeben, sondern in einer dem Test angemessensten Form dargeboten. Arbeitsanweisungen zum Erfüllen eines Lückentextes werden folglich in geschriebener Form dargeboten. Tests in denen der Input in akustischer Form dargeboten worden ist, bieten auch Arbeitsanweisungen in derselben Form.

Die Mustereingabe dient dazu die Testaufgaben zu vereinfachen, insbesondere wenn die Testform für die Lernenden unbekannt ist. In manchen Fällen sind die Arbeitsanweisungen für die Lernenden unklar, auch wenn sie detailliert erstellt sind. Die Musterwahl sollte daher nicht willkürlich erfolgen “providing good examples can be as difficult as writing good test tasks“<sup>119</sup>

Sprachtests können mehr als einen Aspekt der Sprachkompetenz messen und darum bestehen sie aus mehr als einer Aufgabe. Es ist sehr üblich am Anfang des Tests generelle Arbeitsanweisungen zu geben und jeder Aufgabe angemessene Arbeitsanweisungen einzuordnen. Die Arbeitsanweisung in einem Sprachtest muss folgende Punkte enthalten:

- Die überprüften Sprachkompetenzen.
- Testteile und ihre relative Wichtigkeit.
- Die Beurteilungsmethode „the scoring method“<sup>120</sup>

Wie bei der Festlegung des Testzwecks helfen die Arbeitsanweisungen, die den Testinhalt bzw. das Testkonstrukt definieren, die Lerner zu motivieren ihr Bestes im Test zu tun. Folglich hilft die Bestimmung der Testinhalte in den Arbeitsanweisungen den Validitätsgrad des Tests zu erhöhen.<sup>121</sup> Diese sollten aber in einer dem Lerner weitaus verständlichen und einfachen Form dargeboten werden.

Durch den Testzweck erfahren wir welche Entscheidungen durch den Test getroffen werden werden. Es muss eine Übereinstimmung zwischen den spezifischen Arbeitsanweisungen jedes Testteils und dem Zweck, wofür der Test erarbeitet wird, bestehen. Wenn dies nicht der Fall ist kann der Test missgebraucht werden<sup>122</sup> Die Arbeitsanweisungen sollen auch eingeben: wie die Testergebnisse mitentscheiden und wie werden die Leistungen benotet. Hierzu wird auf BACHMAN/PALMER (1996) verwiesen: “we believe that if test takers understand that there is a fair and legitimate use for a given test, they will be more likely to take the test seriously and to attempt to do their best „<sup>123</sup>

---

<sup>119</sup> BACHMAN/PALMER. (1996, S. 184)

<sup>120</sup> Ebd., S.184

<sup>121</sup> Ebd., S. 186

<sup>122</sup> Ebd., S. 184

<sup>123</sup> BACHMAN/PALMER. (1996, S. 85)

### **1.4.3 Vorgehensweisen bei der Erstellung standardisierter Tests**

Die Testerstellung bezieht sich auf alle Verfahren, die die Konzeption und Vorbereitung des Tests umfassen. Die Zeit und Mühe, die bei der Erstellung des Tests verwendet werden, hängen von den Zwecken ab, wozu der Test erstellt wird.

Es werden drei Etappen bei der Erstellung eines standardisierten Tests unterschieden:

- das Design
- das Operationalisieren
- die Administration oder Durchführung.

#### **Das Design**

In der Designetappe wird die Testperformanz in Bezug auf den realen Sprachgebrauchsbereich festgelegt. In dieser Etappe wird ein Design-Statement entwickelt, das folgende Aktivitäten umfasst:

- a-Beschreibung des Testzwecks.
- b-Beschreibung des realen Sprachgebrauchsbereichs.
- c-Beschreibung der Zielgruppe.
- d-Definition des Testkonstrukts.
- e-Eine Checkliste zur Testbeurteilung im Hinblick auf die Testgütekriterien.

Der Zweck der Prüfungsanleitung besteht darin, dass das Design-Statement alle notwendigen Prinzipien für die Erstellung der Testaufgaben, der Prüfungsanleitung und des Tests zur Verfügung stellt. Entsprechend umfasst die Designetappe die folgenden sechs Aktivitäten:

#### **Beschreibung des Testzwecks**

Die Beschreibung des Testzwecks schließt die klare Festlegung der spezifischen Rückschlüsse und Entscheidungen über die getesteten sprachlichen Kompetenzen ein, die aufgrund der Testergebnisse getroffen werden. Die Beschreibung des Testzwecks hilft bei der Einschätzung des potenziellen Impacts der Prüfung.

**Die Beschreibung der entsprechenden Aufgaben im Bereich des realen Sprachgebrauchs** Die Beschreibung der Aufgaben im realen Sprachgebrauch stellt die Basis für die entsprechenden Testaufgaben dar. Diese Beschreibung hilft der Lehrkraft in der Einschätzung der potenziellen Authentizität und Interaktivität des Tests.

#### **Beschreibung der Sprachverwender Charakteristika und der Testgruppe**

In dieser Etappe wird die Testpopulation beschrieben, also die Leute, für die der Test erstellt wird. Diese Etappe liefert eine weitere Basis zur Einschätzung des Testimpacts.

### **Die Beschreibung des Testkonstrukts**

Bei dieser Aktivität beschreibt die Lehrkraft die genaue Natur der sprachlichen Kompetenz, die der Test misst; es geht um die Beschreibung des Testkonstrukts. Diese Beschreibung ermöglicht die Einschätzung der Validität des jeweiligen Tests<sup>124</sup>

### **Die Entwicklung eines Beurteilungsrasters zur Überprüfung der Testgültigkeit**

Die Entwicklung eines Plans zur Beurteilung der Testgültigkeit garantiert einen minimalen Grad der Gültigkeit, Die Testgültigkeit wird durch die Gewichtung der Testgütekriterien realisiert. Der Plan enthält auch einen Beurteilungsraster zur Überprüfung der inneren Validität und Gültigkeit der Testaufgaben. Zum Plan gehört auch die Feedbacksammlung, die sich auf die quantitativen Informationen wie die Testergebnisse und ihre Analyse sowie die Posttestreports stützt.

### **Beschreibung der Ressourcen und ihr Management**

Die Beschreibung der Ressourcen bezieht sich auf das Klarmachen der vorhandenen Ressourcen wie Menschen, Material und Zeit, die zur Testerstellung erforderlich sind. Ein Plan zum Management und Steuerung dieser Ressourcen soll entwickelt werden, um die Einschätzung der potenziellen Praktikabilität des Tests zu ermöglichen.

### **Das Operationalisieren**

In einem Sprachtest umfasst die Etappe des Operationalisierens die Entwicklung einer Prüfungsanleitung und die Festlegung der Testaufgaben. Das Operationalisieren bezieht sich auch auf das Schreiben einer aktuellen Testaufgabe, auf die Erstellung der Arbeitsanweisungen und Spezifizierung einer Beurteilungsmethode für den jeweiligen Test. Die Erstellung der Testaufgaben erfolgt mit der Beschreibung des Sprachgebrauchsbereichs, der schon im Design-Statement festgelegt wurde. Die im Design-Statement festgelegten Sprachgebrauchsbereiche können in der Entwicklungsphase modifiziert um den Gültigkeitskriterien zu entsprechen. Die Charakteristika der festgelegten Sprachgebrauchsbereiche liefern die Basis für die Erstellung der Testaufgaben. Die Prüfungsanleitung enthält Prinzipien und Anweisungen zur allgemeinen Strukturierung und

---

<sup>124</sup> Ebd., S. 89

Organisierung des Tests. In diesem Zusammenhang unterscheidet sich eine Prüfungsanleitung von einem Design-Statement dadurch, dass eine Prüfungsanleitung detaillierter ist, während ein Design-Statement die allgemeinen Parameter für die Testerstellung bzw. den Testzweck, den Sprachgebrauchsbereich, die Lerngruppe usw. festlegt. Andererseits beschreibt eine Prüfungsanleitung vielmehr die innere Struktur des Tests, wie die Arbeitsanweisungen und die Testaufgaben und stellt sie in Verbindung mit den Prinzipien in dem Design-Statement auf. Eine andere wichtige Aktivität in der Entwicklungsphase ist das Schreiben der Arbeitsanweisungen. Effektive Arbeitsanweisungen beschreiben expliziterweise die Teststruktur und die Natur der sprachlichen Aufgaben, und wie die Lerner aufgefordert sind, auf diese Aufgaben zu reagieren. Die Spezifizierung der Beurteilungsmethode umfasst zwei Etappen:

- die Festlegung der Kriterien anhand deren die Lernerantworten beurteilt werden.
- die Festlegung der Prozeduren, denen die Lehrkraft folgt um ins Testergebnis zu gelangen. (Richtigkeitskriterien, Punkteverteilung, Notenskala)

### **Administration (Durchführung)**

Die Testdurchführung ist eine Testphase, die die Einreichung des Tests zu einer Lerngruppe, die Sammlung der Informationen und die Analyse der gesammelten Informationen umfasst. Die gesammelten Daten können sowohl für die Messung der Testgültigkeit als auch für die Beurteilung der Testergebnisse zwecks Entscheidungen zu treffen, die die Lerner direkt betreffen, von Belang sein. PALMER/ BACHMAN (1996) haben die Durchführungsphase in zwei Etappen unterteilt: die Try- Out Etappe, in der die Testersteller hauptsächlich Informationen über den Test sammeln. Die gesammelten Informationen dienen zur Festlegung der Testgültigkeit oder zur Optimierung der Testverfahren. Wenn die gesammelten Daten von der Testdurchführung nur zur Erhebung der Testgültigkeit dienen, wird es in diesem Fall um ein „Pretesting“ oder Testerprobung gehen.

### **Operationalisierter Testgebrauch**

Andererseits bezieht sich der operationalisierte Testgebrauch auf die Durchführung des Tests zwecks Informationen über die Lernalterssprachkompetenz zu erheben, um einen spezifischen Testzweck zu realisieren. In dieser Etappe können die gesammelten Informationen zur Überprüfung der Testgültigkeit dienen. Die Testdurchführung bezieht sich u.a. auch auf die Vorbereitung der Testumgebung und des Testmaterials, auf die Ausbildung der Testadministratoren, die Einreichung des Tests zur Lerngruppe, die Sammlung des Feedbacks



und die Analyse der Testergebnisse. Bei der Analyse der Testergebnisse sind drei Operationen von Belang, es handelt sich um:

### **Itemanalyse**

Die Einschätzung der Reliabilität der Testergebnisse anhand von statistischen Verfahren, wie die Errechnung des Einfachheitsgrads und die Durchführung eines Korrelationstests. Eine weitere letzte Operation, die für die Testdurchführung wichtig ist, ist die Archivierung der gesammelten Daten sowie der Lernerkopien zwecks der Bildung einer Prüfungsbank, um die Erstellung neuer Tests in der Zukunft zu erleichtern.

### **Post Test Reports**

Tests sind wichtig für diejenigen, die an denen teilnehmen und auch für die Lehrkräfte, die die Ergebnisse interpretieren. Darum sollten sie in ihren Entscheidungen am genauesten umgehen. Es wird aber viele Zeit und Mühe verwendet bei der Erstellung von Sprachtests und ihrer Durchführung. Daher ist es wichtig für die Testersteller und die Institutionen diese Tests und Testprozeduren zu speichern bzw. eine Prüfungsbank zu bilden. Von dem Testvorgehen, bis zum Testergebnis und Feed-Back zu speichern, und zur entsprechenden Institution weiterzuleiten.<sup>125</sup>

### **Testanalyse**

Bei der klassischen Analyse der Testaufgaben sind zwei Kriterien zu berücksichtigen: Der Einfachheitsgrad und der Diskriminierungsindex. Erwähnenswert bleibt, dass die kompetenzorientierte Testtheorie diese Analyseverfahren anerkannt und auch adoptiert hat. Dies lässt sich dadurch erklären, dass die Analyse, die die traditionelle Testtheorie führt, sich an logischen und statistischen Verfahren orientiert. Die logische und statistische Analyse (Einfachheitsgrad/ Diskriminierungsindex) erfolgt unabhängig von jeder Theorie.

### **A-Der Einfachheitsgrad**

Der Einfachheitsgrad einer Aufgabe, (engl.: "Facility value") bezieht sich auf den Anteil der Lerner, die sie richtig gelöst haben.

Wenn 50 Lerner von 100 die Aufgabe richtig gelöst haben, dann ist der F.V von 50%. Wenn der Einfachheitsgrad einer Aufgabe von 2% ist, ist die Aufgabe dann sehr schwer, eine

---

<sup>125</sup> ALDERSON et al.,(1995, S. 197)

Testaufgabe deren Einfachheitsgrad von 95% ist, ist folglich sehr einfach. Sprachtestexperten mahnen Testaufgaben von einem Einfachheitsgrad zwischen 40 bis 60%<sup>126</sup>. Eine Testaufgabe von einem sehr hohen oder sehr niedrigen Einfachheitsgrad ist wenig informativ, da sie die wirklichen sprachlichen Kompetenzen der Lerner nicht erheben kann. Eine Testaufgabe, deren Einfachheitsgrad 0% ist, wird uns keine Information liefern außer, dass der Test sehr schwer ist.<sup>127</sup>

### **B-Der Diskriminierungsindex**

Der Diskriminierungsindex bezieht sich auf die Diskriminierungsfunktion des Tests. Dies besagt, dass in einer validen Testaufgabe, schneiden die hochleistenden Lerner besser ab, als die schwachleistenden. Das Errechnen des Diskriminierungsindex wird durch das Ranking der Lerner in zwei verschiedenen Gruppen je nach ihren erbrachten Noten im ganzen Test, und diese in zwei Gruppen einteilen. In der Gruppe „A“ werden die hochleistenden Lerner aufgelistet, während die schwachleistenden in der Gruppe „B“ aufgelistet werden.

Das Errechnen des Diskriminierungsindex erfolgt durch den Vergleich der Lernerergebnisse in den beiden Listen zu vergleichen. Nehmen wir an, dass 70% der Lerner in der Gruppe „A“ die Aufgabe richtig gelöst haben, während 20% der Lerner in der Gruppe „B“ die selbe Aufgabe richtig gelöst haben, wird in diesem Fall der Diskriminierungsindex wie folgt errechnet:  $D.I = 70 - 20 = 50$ . Der Diskriminierungsindex der Aufgabe ist 50%. Der höchste Diskriminierungsindex kann 100% erreichen, wenn alle Lerner der Gruppe „A“ die Aufgabe richtig gelöst haben, und Keiner in der Gruppe „B“ die gleiche Aufgabe gelöst hat.

### **C-Korrelation**

Die Korrelation ist die Konvergenz der Ergebnisse in zwei Tests. Die Schülerleistungen korrelieren in zwei Tests, wenn die Leistungen des einzelnen Lerners im ersten und im zweiten Test nahebei liegen. Ein Beispiel für die Korrelation gibt ALDERSON<sup>128</sup> Sein Schema erläutert den Fall einer perfekten Korrelation wenn die Lerner genau die gleichen Ergebnisse im ersten wie im zweiten Test bekommen. Also sie sind in den beiden Tests gleich eingeordnet. Die perfekte Korrelation der Testergebnisse ist rundweg unrealistisch. Die Punkte in der Graphik entsprechen der Schülerleistungen im Test 1 und 2.

---

<sup>126</sup> <http://www.btinternet.com/~ted.power/esl0724.html>. besucht am 12. 01. 2010

<sup>127</sup> ALDERSON et al., (1995, S.76)

<sup>128</sup> ALDERSON et al., ( 1996, S. 77)

### 1.3.4.1 Testimpact auf die Testkandidaten

Tests können die Testteilnehmer beeinflussen durch:

1. die Erfahrungen der Lerner mit dem Test und durch ihre Vorbereitung für den jeweiligen Test,
2. durch das Feed-back , das sie über ihre Testperformanz erhalten
3. und durch die Entscheidungen, die getroffen werden aufgrund der Testergebnisse.

Standardisierte Tests wie das Abitur oder anerkannte Zertifikate fordern eine relativ lange Vorbereitung. Manche Lerner müssen wochenlange Vorbereitungen durchmachen oder sich einem mehrjährigen Kurs unterziehen um der Prüfung ausgesetzt zu sein. Für das Eignen des Zertifikats Deutsch z.B nimmt der Lerner einen relativ langen Weg bis zum Zertifikat. Die Erfahrung des Tests kann auch das strategische Wissen der Lerner verbessern, wenn der Test weitere Informationen über das Thema liefert. Die Testdurchführung selbst hilft dem Teilnehmer dabei, seine eigene Leistung durch den Prozess des Wiederrufens schon gelernter Inhalte zu erkennen und sich seiner eigenen Stärke und Schwäche bewusst zu sein, und dies ohne die Rolle des Feedbacks in Betracht zu nehmen. Die Probandenperzeption des Sprachgebrauchs in dem entsprechenden sprachlichen Bereich (TLU) kann von dem Testinhalt orientiert werden. BACHMAN/PALMER (1996) verteidigen den Standpunkt, dass der Lerner auch sein Sprachvermögen verbessern kann, durch seine Testeinnahme, da er durch den Test auch eine sprachliche Erfahrung erlebt oder durch das Feedback, das er bekommt. Er kann auch durch den Input andere Informationen wahrnehmen, sowohl grammatisches Wissen als auch Fachwissen, wie auch beim Feedback. Die Testform oder die Testaufgaben können auch das strategische Verhalten des Lerners während des Tests verbessern. Ferner meinen BACHMAN/PALMER (1996), dass die Lerner bei der Entwicklung des Tests miteinbezogen werden müssen. Der Testersteller soll auch die Lernerperzeption des Tests erheben.<sup>129</sup>

Das Feedback kommt in den meisten Fällen in Form einer Ziffernote vor, die das Ergebnis der Leistung darstellt. Es ist daher wichtig, dass wir das Testergebnis für den Lerner und eventuell für die Eltern sinnvoll und klar präsentieren. Dies schließt die richtige Interpretation des Testergebnisses und die klare Erstellung der verbalen Beurteilungen bzw. der formativen Leistungsbeurteilung ein. Sie helfen den Eltern auch dabei, die Testnote zu

---

<sup>129</sup>BACHMAN/PALMER. (1996, S.32)

interpretieren. Entscheidungen, die aufgrund dieser Ergebnisse getroffen werden, können wichtige Konsequenzen auf die Testkandidaten haben. Sei es auf den Vorzug in eine höhere Klasse oder die Aufnahme in einem Arbeitsplatz. In Algerien entscheiden die Ergebnisse im Fach Deutsch über die Aneignung des Abiturs, wie auch über die Zulassung zu einem Deutschstudium an der Universität mit.

### **Testtyps in Bezug auf das ermittelte Feedback**

#### **Non-reziprokale Tests**

Wie beim Lesen, gibt es in einem nicht reziproken Sprachgebrauch weder Feedback noch Interaktion zwischen den Lernern. Lesen ist ein typisches Beispiel, da die Reaktion des Lesers auf den Input die Form des Inputs nicht verändern wird.<sup>130</sup>

#### **Adaptive Tests**

In denen die Antworten in einer gegebenen Aufgabe von der richtigen Antwort in einer Ersteren abhängig ist. In adaptiven Tests erhalten Testkandidaten kein Feedback über ihre Performanz in den einzelnen Teilaufgaben. Die Verwendung einer Checkliste für Aufgabencharakteristika bei der Erstellung angemessener Testaufgaben und bei der Modifizierung schon gegebener Testaufgaben, um den Gültigkeitsgrad dieser Tests zu erhöhen, ist wichtig. Ziel jedes Sprachtestens ist Daten über die Sprachliche Kompetenz der Lerner zu sammeln, zwecks Entscheidungen zu treffen, die diese Sprachlerner betreffen. Daher war es wichtig zu beweisen, ob die Testperformanz dem Sprachgebrauch entspricht. Allgemein ist der Sprachgebrauch als die Schöpfung oder Interpretation von ermittelten Inhalten von einem Individuum in einem Gespräch aufgefasst, oder als die dynamische und interaktive Negotiation von intendierten Bedeutungen oder Inhalten zwischen zwei oder mehreren Individuen in einer bestimmten Situation.<sup>131</sup> BACHMAN/PALMER (1996) meinen, dass nicht nur Testaufgabencharakteristika die Performanz an einem Sprachtest beeinflussen. Individuelle Charakteristika wie Sprachwissen, persönliche Eigenschaften, Wissen über den Testinhalt können die strategische Kompetenz sowohl in alltäglichem Sprachgebrauch als auch in der Testperformanz beeinflussen.

---

<sup>131</sup> BACHMAN/PALMER. (1996, S. 63)

## **Individuelle Charakteristika**

Der reale Sprachgebrauch, sowie die Performanz in Sprachtests kann von personalen Eigenschaften beeinflusst werden, wie Müdigkeit, die weitgehend unvorhersehbar ist. Es gibt andere persönliche Eigenschaften, die bei der Verarbeitung von Sprachtests zu berücksichtigen sind. Individuelle Charakteristika wie: Alter , Geschlecht und Muttersprache, das Wissen über den Inhalt des Tests und die Sprachkompetenz der Testteilnehmer<sup>132</sup>

### **1.4.3.2 Affektive Schemata und Testperformanz**

Die individuellen Eigenschaften sind nicht ein Teil der Sprachkompetenz, aber könnten sie beeinflussen. Viele Autoren haben diese Eigenschaften studiert u.a. BROUN, 1994; COHEN, 1994 und eine Liste verschiedener Persönlichkeitseigenschaften erstellt. Diese Eigenschaften sind u.a. Alter, Fremdsprachliches Können, soziopsychologische Faktoren, Persönlichkeit, Sprachgebrauchsstrategien, ethnologische Faktoren und multilinguistisches Können.<sup>133</sup> Sprachtestersteller sollen folglich bei der Sprachtesterarbeitung eine Liste von Merkmalen berücksichtigen, wie Alter, Muttersprache, Niveau und Typ der allgemeinen Ausbildung, und eventuelle Vorerfahrungen mit dem jeweiligen Test. Einige Testaufgaben, die kulturelle Inhalte enthalten könnten für die Lerner, die über das jeweilige Wissen verfügen, einfacher sein.

Affektive Schemata oder die emotionale Korrelation von Themen und Objekten des Weltwissens können auch die Testperformanz beeinflussen.

Hierzu schreiben BACHMAN/PALMER (1996):” In a language test, test takers affective schemata may influence the ways in which they process and attempt to complete the test tasks. “ Die Berücksichtigung der affektiven Schemata impliziert die Vermeidung viel emotionaler Themawahl oder schockierender Themen bei der Themawahl wie z.B. zu aktuelle Katastrophen, da die Lernerperformanz durch die allzu emotionale Ausschöpfung des Textes beeinträchtigt oder behindert werden könnte. Andererseits könnten emotionale Reaktionen nach BACHMANs Meinung den Sprachgebrauch erleichtern<sup>134</sup>. Testaufgaben und Testdesign müssen den Testkandidaten Sicherheitsgefühlen garantieren. Da setzt sich das Prinzip durch, dass die Testform den Persönlichkeitsmerkmalen der Kandidaten entsprechen soll. Zusammenfassend bestimmen Affektive Schemata in Verbindung mit den Testcharakteristika die affektiven Reaktionen der Lerner auf die Testaufgaben.

---

<sup>132</sup> Ebd., S. 64

<sup>133</sup> COHEN. (1994, S. 74) zitiert in BACHMAN/PALMER (1996, S. 64)

<sup>134</sup> BACHMAN/PALMER. (1996, S. 66)

### 1.4.3.3 Die Beurteilungsmethode

Die Beurteilungsmethode gehört zum Prozess des Operationalisierens eines Tests. Sie kann die Testaufgaben oder die Lernerergebnisse in dem Test beeinflussen. Es ist in der Testtheorie bekannt, dass die Beurteilung geschlossener Testaufgaben einfacher ist als die Beurteilung der offenen Aufgaben. Bei der Beurteilung der Lernerarbeiten hat die Lehrkraft zwei Prozeduren zu folgen: Die Lehrkraft kann holistische oder analytische Beurteilungsmethode anwenden. Bei einer holistischen Beurteilungsmethode wird eine allgemeine Punktezahl der Lernerperformanz zugewiesen. Diese Operation beruht auf dem allgemeinen Lehrereindruck über die Lernerperformanz.<sup>135</sup> „Holistic scores represent teacher’s overall impressions and judgments“<sup>136</sup>

Diese Beurteilungsmethode ist, wenn es um die Lenkung und Optimierung des Lernens geht, keine angemessene Alternative. Bei der analytischen Beurteilungsmethode hingegen wird je einer gelösten Aufgabe bzw. Lernerperformanz eine Punktezahl zugewiesen. So werden die sprachlichen Elemente und Kompetenzen separat beurteilt. Meistens werden die Punktzahlen hinzugefügt um eine holistische Punktezahl einzugeben. Die Lehrkraft kann auch die Punktzahlen separat lassen, Diese werden „Performanzprofite“ genannt.<sup>137</sup>

Die Beurteilungsskala (engl.:“the scoring categories“), die ein analytisches System aufweist, sollte bildungsbezogene Ziele und Pläne widerspiegeln und diesen Bildungsprinzipien Folge leisten.

Die Punkteverteilung und die Beurteilungsskala müssen den Lernern in den allgemeinen Arbeitsanweisungen angegeben werden, damit der Lerner über die Kriterien, die bei der Beurteilung benutzt werden, erfährt. Die Bewertungskriterien werden in den Arbeitsanweisungen angegeben<sup>138</sup>Die Interpretation der Testergebnisse lässt sich in zwei Komponenten aufteilen: Inferenzen über die Kompetenzen, die getestet werden, und Inferenzen über den entsprechenden Bereich des realen Sprachgebrauchs.

### Zusammenfassung

In dem theoretischen Teil der vorliegenden Arbeit wurden Stellenwert und Funktionen der Leistungsmessung- und beurteilung in der schulischen Umgebung, sowie die Reformen und Versuche zur Abschaffung der Noten aus der Schule aufgelegt.

---

<sup>135</sup> UPSCHUR/GENESE. ( 1996, S. 206)

<sup>136</sup> Ebd., S. 206

<sup>137</sup> Ebd., S. 116

<sup>138</sup> BACHMAN/PALMER. ( 1996, S. 189)

Hinsichtlich der Funktionen der *L. B* gilt sie nach wie vor als ein notwendiges Unterrichtsvehikel und Motivationsfaktor. Außerdem fungiert sie als Rückmeldung für Lehrer und Eltern.

Auf der Macro- Ebene wird die Testpraxis eingesetzt, um schulentwicklungsbezogene Zwecke zu überprüfen, wie bei der Überprüfung der Lehrplaneffektivität und der Inszenierung neuer Schulreformen.

Die *L. B* kann nur dann ihre Funktionen erfüllen, wenn die Bewertungsprozeduren zuverlässig sind. Die Testergebnisse aus ungültigen Testverfahren sind vielmehr die Resultate der Messfehler und weniger des getesteten Konstrukts.

Die Problematik der Messfehler ist von ethischen Fragen nicht zu trennen, da die *L. B* für die Selektionsfunktion der Schule als Grundlage gilt. Die ethische Problematik wird auch bei offenen Testformen erregt, die dem Lehrer einen größeren Ermessensspielraum erlauben.

Die Testfehler können durch den Einsatz gültiger Testformen reduziert werden. Dies erfolgt durch eine gezielte Gewichtung der Testgütekriterien Validität, Reliabilität und Objektivität.

Im diesem Kontext werden Abweichungen von den Testgütekriterien hauptsächlich in den Testfacetten diagnostiziert. Dafür müssen die Testfacetten wie Testrubrik und Testaufgabenform klar beschrieben und sorgfältig erstellt werden.

Bevor ein Test zum Einsatz fertig gestellt wird, müssen statistische Prozeduren zur Beurteilung der Testgültigkeit durchgeführt werden, u. a. eine Testerprobung, die Feststellung des Testschwierigkeitsgrades und dessen Diskriminierungsgrad.

Die Testform -und methode sind für Testdesigner, die einzigen Merkmale, die unter Kontrolle gesetzt werden können, um die Lernerperformanz der Lerner zu fördern.

## **2. Empirische Untersuchungen und Befunde**

### **2.1 Curriculumanalyse**

Der Lehrplan legt die Kompetenzen, die im Lehrwerk zu vermitteln sind, fest und stellt die Selbstbeurteilung innerhalb der methodischen Kompetenzen, die der Lerner sich eignen sollte. (les competenes d'ordre methodologique)<sup>139</sup>. Es wird anhand der durchgeführten Analyse beobachtet, dass weder das Curriculum noch das Lehrwerk die Richtigkeitskriterien, die als allgemeine gültige Bewertungsreferenzen dienen können, festgelegt haben. In Bezug auf das Kriterium, ob es im Lehrplan Hinweise auf die Leistungsbeurteilung vorzufinden sind, wird festgestellt, dass der geltende Lehrplan für das algerische Gymnasium auf die Leistungsbeurteilung und zwar auf die formative und summative Leistungsbeurteilung hinweist. Die entsprechenden Hinweise wurden theoretisch und oberflächlich gestaltet, was der Lehrkraft in der Praxis keine große Hilfe bieten könnte. Überdies weist das Curriculum auf die Notwendigkeit der Rückmeldung hin.<sup>140</sup> Es wird dem algerischen geltenden Curriculum in diesem Zusammenhang entgegengehalten, dass es die Besonderheiten des algerischen Deutschunterrichts und des algerischen Prüfungssystems nicht detailliert aufstellt. Anhand der Curriculumanalyse wurde auch festgestellt, dass die praktischen Vorgehen der Leistungsbeurteilung nicht festgelegt sind und, dass die L.B mit den Kompetenzen, die die Lerner erwerben sollten nicht gebunden wird. Das Curriculum enthält keine Punkteskalen und keine Hinweise auf den Koeffizienten des Faches Deutsch. Hierzu sollte zwischen einer Punkteverteilung und einer Punkteskala unterschieden werden. Die anderen Evaluationsformen werden kaum erwähnt. Die Curriculumanalyse hat ergeben, dass das algerische geltende Curriculum für das Gymnasium Defizite im Hinblick auf die Leistungsbeurteilung aufweist.

---

<sup>139</sup> S. das algerische geltende Curriculum für das Gymnasium (2006, S. 12)

<sup>140</sup> Das algerische geltende Curriculum für das Gymnasium. (2006, S. 12)



Die Analyse der Testgültigkeit hängt von den Gütekriterien, die er aufweist. Die Gütekriterien betreffen dieselben Testteile. Wenn ein Test die Reliabilitätskriterien erfüllt, erfüllt er auch die Konstruktvaliditätskriterien in ansteigendem Maß. Darum rekurren bei jeder Analyse dieselben Maßstäbe. Es wird bei der Analyse der Arbeitsanweisungen auf die Kriterien von BACHMAN und PALMER(1996) gestützt, die die logische Gültigkeit des Tests betreffen. Diese Kriterien wurden gewählt, da sie in enger Verbindung mit den kommunikativen Kompetenzen und dem realen Sprachgebrauch erstellt wurden, was für die allgemeinen Ziele des algerischen DaF-Unterrichts angemessen ist. Darüber hinaus stellt BACHMANs theoretischer Rahmen die Referenz für das Sprachtesten<sup>141</sup>. Die Analyse zielt auf die Prüfungsanleitung Eng. "Blue-Print", die vom algerischen Ministerium vorgelegt wurde. Prüfungsanleitungen sind sehr wichtig, da sie eine allgemeine Referenz für die Lehrer bei der Erstellung der Hauptschulprüfungen und gegebenenfalls auch der Lernkontrollen sind. In der nachfolgenden analytischen Arbeit wird hauptsächlich mit der Überprüfung des Gültigkeitsgrads des oben erwähnten Testmodells beschäftigt. Es wird bei der Auswahl der Testanleitung davon ausgegangen, dass die Lehrer dem Testmodell, das von der Testanleitung vorgeschlagen wird, bei der Erstellung der Hauptschulprüfungen nachkommen. Darum sind die Prüfungen, die in allen algerischen Gymnasien eingesetzt werden, in einer weitgehend einheitlichen Form dargeboten. Dieses Modell wird ausgewählt, da es praktisch unmöglich ist, alle Einzeltests zu überprüfen. Eine ökonomische Methode den Umstand des Prüfens und Bewertens im algerischen DaF-Unterricht unter der Lupe zu setzen ist, das vom algerischen Erziehungsministerium vorgelegten Testmodell zu überprüfen. Weitere Gründe für die Auswahl der Hauptschulprüfung für algerische DaF-Gymnasiasten sind ihre entscheidende Rolle als Vorbereitungstest für das Abitur und allgemeine Gültigkeit unter Lehrern und Schülern bzw. ihre allgemein verbreitete "Facevalidität".

Das erste Kriterium betrifft die Form der Prüfung und ihre Teile. Bezüglich dieses Kriteriums habe ich während meiner Hospitationen in verschiedenen Gymnasien in Setif und Bordj Bou Arreridj während des Schuljahres 2009/2010 beobachtet, dass die in algerischen Gymnasien eingesetzte Hauptschulprüfung nur eine einzige Prüfung darstellt. Die Prüfung misst die schriftlichen Kompetenzen der Lerner und dauert 3 Stunden. Die Lerner werden einem Mündlichen Teil nicht unterzogen. Es muss erwähnt werden, dass die Lerner in identischen Bedingungen die Prüfung machen. Der Test wird gewöhnlich in vertrauten Klassenzimmern durchgeführt, in denen jeder Schüler in einem individuellen Tisch sitzt. Im

---

<sup>141</sup> UPSHUR/GENESE ( 1996)

Administratorführer wird ein Sitzplan vorgegeben, der imperativ respektiert werden muss. Jeder Lerner soll an der rechten Seite des Tisches sitzen, so dass die Interaktion zwischen den Lernern während der Prüfung vermieden wird. Die Testadministratoren sind meist Gymnasiallehrer. Die Testbedingungen sind dem Lerner von weitgehender Vertrautheit. Während der Prüfungshospitationen im experimentellen Abitur und in Hauptschulprüfungen wurde festgestellt, dass die Prüfungsbedingungen mit denen des Abiturs (Testmethode, Testumgebung, Testpersonal) weitaus übereinstimmen.

Was das zweite Kriterium anbelangt, ob die Merkmale der Testrubrik von einem Test zu dem anderen nicht variieren, dass die Lernerperformanz in dem Test nicht dadurch beeinflusst werden könnte, erfüllt die Prüfung hinsichtlich der Zeitzuweisung dieses Kriterium, da den Lernern 3 Stunden zur Verfügung stehen. Wenn die Prüfung zwei Themen vorlegt, wird eine Halbstunde hinzugefügt, um den Lernern genug Zeit für die Themawahl zur Verfügung zu stellen. Im Fall einer Doppelthematik wird das zweite Thema nach derselben Methode wie das Erste erstellt. Es wird in dieser Hinsicht festgestellt, dass dieser Punkt in der Prüfungsanleitung nicht erwähnt wurde. Es wird in der Testrubrik auch anhand der Arbeitsanweisung hingewiesen, dass die Prüfung zwei Themen vorschlägt und, dass der Lerner nur auf eine einzige „Prüfungsvariable“ antworten muss. Die Analyse der Prüfungsanleitung hat ergeben, dass sie keine detaillierte Anweisungen bezüglich der Erstellung der beiden Prüfungsvariablen bietet, was zu ernster Beeinträchtigung der Reliabilität, sowie der Konstruktvalidität der Prüfung führt.

Die Beeinträchtigung der Konstruktvalidität könnte vorkommen, wenn die Lehrkraft die überprüften Kompetenzen nicht in den beiden Prüfungen angemessen proportioniert, so dass die Prüfungswahl zwischen den Lernern diskriminiert. In der ersten Testaufgabe könnte dieses Kriterium beeinträchtigt werden, wenn die Lehrkraft zwei Texte von sehr verschiedenem Schwierigkeitsgrad einsetzt, so dass ein Text für den Schüler einfacher scheint als der Andere.

Dies kommt klar zum Ausdruck, insbesondere in dem zweiten Subtest „Grammatik“. Die Lehrkraft kann in der ersten Prüfungsvariable das Passiv überprüfen und in der Zweiten das Präteritum. Der Einsatz von zwei Prüfungsalternativen wird im Grunde um die Lerner eine zweite Chance zu widmen festgelegt, und auch für psycho-pädagogische Zwecke, wie die Stimulierung der Motivation bei den Lernern. Dies führt aber zu Schwierigkeiten bei der Festlegung des Testkonstrukts wegen der Prüfungsform; bei der Überprüfung des Passives im

Subtest Grammatik müssen die beiden Prüfungsalternativen die gleichen grammatischen Strukturen testen und auch von analogem Schwierigkeitsgrad sein.

Diese wichtigen Bedingungen für die Erstellung der Prüfungsvariablen können nicht respektiert werden, wenn darauf in der Prüfungsanleitung nicht hingewiesen wird. Bezüglich der Testergebnisse, enthält die Prüfungsanleitung eine Punktwertverteilung, die zu allen Arbeiten einheitlich einzusetzen ist. Es wird jeder gelösten Aufgabe ein Punktwert zugewiesen. Die Punktwertverteilung ist mit einer detaillierten Punkteskala nicht begleitet. Bezüglich der Teststruktur teilt sich die Prüfung in drei Hauptteile, je nach der Kompetenz, die der Subtest misst.

Der erste Teil überprüft das Leseverständnis; er enthält drei Aufgaben, die auf das Leseverstehen als rezeptive Sprachkompetenz zielen. Die Erste Aufgabe ist eine Richtig-Falsch Aufgabe, die zweite Aufgabe umfasst vier Fragen zum Textverständnis. Die Lerner sollen auf die vier Fragen durch die Entnahme der Informationen aus dem Text antworten. Viertens liegt eine Kombinationsaufgabe zum Testen des Wortschatzes vor. Der zweite Teil innerhalb des ersten Teils betrifft die Sprachkompetenz und teilt sich in drei Aufgaben:

- Eine Wortschatzaufgabe, in der der Lerner die Synonyme und Gegenteile aus dem Text herausfinden sollte.
- Eine Wortbildungsaufgabe .
- Eine Übersetzungsaufgabe; zum Übersetzen ins Arabische wird ein Satz aus dem Text entnommen.

Der zweite Hauptteil bezieht sich auf die Grammatik und teilt sich wiederum in vier Aufgaben:

-Zwei Aufgaben bzw. „Punkte“ zum Testen der Tempora.

Die anderen drei Punkte werden dem Testen der anderen grammatischen Strukturen, die schon während des Schulkurses behandelt wurden, zugewiesen. Die Vierte Aufgabe testet die Deklination in Form eines kleinen Cloze-Tests. Es wurde bei der Analyse der Prüfungsanleitung festgestellt, dass das von der Prüfungsanleitung aufgestellte Testmodells keine Temporaaufgaben enthält, obwohl die Testanleitung es imperativ festlegt. Der Lehrer muss bei jeder Testdurchführung eine Temporaufgabe einsetzen. Der dritte Teil der Prüfung betrifft die Schreibkompetenz; die Aufgabe bietet dem Lerner zwei Themen zur Wahl: die erste Aufgabe ist eine offene Aufgabe zu einem der neun Themen, die im Lernprogramm

behandelt wurden. Die Zeitanweisung des Tests macht 3 Stunden aus, der Lerner kann potenziell jeden Teil in einer Stunde lösen.

In Bezug auf die Beurteilungsmethode wurde festgestellt, dass die Prüfung eine Punkteverteilung von 0 bis 20 Punkte vorlegt, die im algerischen gymnasialen Prüfungssystem üblich ist. Die erbrachte Note wird in 3 vervielfältigt. Dies bedeutet, dass der Mitwirkungsgrad im Vergleich zu den anderen Evaluationsprozeduren, zu Lernkontrollen und dauernder Evaluation im Klassenzimmer hoch ist. Der Mitwirkungsgrad hat einen Bezug mit dem Impact der Prüfung auf die Lerner; wenn der Mitwirkungsgrad im Zusammenhang mit den anderen Tests hoch ist, wird auch der Impact der Prüfung auf die Lerner hoch.

In diesem Zusammenhang wird die Punkteverteilung mit einer detaillierten Punkteskala nicht begleitet, was zu ernststen Benotungsproblemen führt.

Und folglich, wie im theoretischen Teil ersichtlich wurde, zur Beeinträchtigung der Testreliabilität. In der Lehrerumfrage, die im Rahmen dieser Arbeit durchgeführt wurde, haben 100% der Lehrer angegeben, dass sie spezielle Probleme bei der Benotung treffen. Die Benotungsprobleme, die aus dem Nicht- Vorhandensein eines detaillierten Beurteilungsrasters hervortreten, sind u.a. die Benotungsschwierigkeit des Passivs. Wenn der Lerner das Hilfsverb und die Satzstruktur richtig schreibt, aber Fehler bei dem Partizip 2 macht, weißt der Lehrer nicht, welchen Punktwert er zuweisen soll.

Hierunter werden die Prüfungsaufgaben im Hinblick auf die Punktwertverteilung analysiert werden. Was die erste Aufgabe anbetrifft, handelt es um eine“ Richtig-Falsch „ Aufgabe. Das Benotungsproblem ist, dass die Lerner allein durch Raten die richtige Antwort herausfinden können. Dies führt zur Beeinträchtigung der Reliabilität der Testaufgabe und der Zuverlässigkeit der Messverfahren. Um den Ratings-Grad zu entkräften, wird von den Lernern verlangt, eine Begründung aus dem Text herauszufinden; dies hilft dabei den Validitätsgrad der Aufgabe zu erhöhen.

Dieser Aufgabe wird ein Punktwert zugewiesen. Bei jeder richtigen Antwort wird ein Punktwert von 0,25 zugewiesen. Es wird bei der Kopienanalyse festgestellt, dass manche Lerner zwar richtige Antworten durch die Eingabe des „richtig oder falsch“ geben, aber falsch bei der Begründung liegen. Problematisch ist auch, wenn der Lerner die Begründung richtig findet und die Richtigkeit bzw. die Unstimmigkeit des Satzes nicht richtig löst. Der Lehrer kann den Punktwert 0,25 nicht teilen. Dies führt zu Divergenzen in den Punkteskalen, die jeder Lehrer auf seine Art verwendet. Die Lehrkraft könnte denselben Lerner verschiedenwertig evaluieren und folglich unobjektiv beurteilen.

Weitere Benotungsprobleme treten bei der Bewertung der Aufgabe „Fragen zum Text“. In dieser Aufgabe wird das Leseverstehen überprüft und der Aufgabe werden 4 Punktwerte zugewiesen. Analog wie in der ersten Aufgabe treten die Benotungsprobleme hauptsächlich wegen Das- Nicht- Vorhandensein einer detaillierten Punkteskala. Dies stellt zwar ein weltweit verbreitetes Problem, diesem Aufgabentyp eigen. Die Lehrer im algerischen Gymnasium sind über den Punktwert, den sie zuweisen sollen nicht einig. Dies führt zu uneinheitlicher Benotung und folglich zu fehlender Reliabilität bzw. Objektivität<sup>142</sup>. In diesem Fall riskiert der Lehrer etwas anders als die Zielkompetenz zu testen. In der Teilaufgabe, die die Übersetzungsfähigkeit überprüft, tritt dasselbe Problem auf. Bezüglich des dritten Kriteriums, inwieweit variieren die Charakteristika des Testinputs von einem Test zu einem Anderen, und in den Testaufgaben, dass sie die Lernerperformanz in dem Test beeinflussen, wird Folgendes festgestellt: Das Problem des sprachlichen Inputs wird insbesondere in mündlichen Tests erregt. In der formellen Hauptschulprüfung des 3. Studienjahrs im algerischen gymnasialen DaF-Unterricht handelt es sich um einen in geschriebener Form dargebotenen Text. In der Prüfungsanleitung ist festgelegt, dass die Thematik des Inputs aus den neun Themen, die vom Curriculum bestimmt wurden ausgesucht werden muss. In der Prüfungsanleitung steht fest, dass die Thematik der Testaufgaben und der Teilaufgaben deren des Textes entsprechen muss. „le vocabulaire utilisé dans les exercices de grammaire doit avoir un lien avec la thématique du text“<sup>143</sup>

„Die in den Testaufgaben eingesetzte Vokabel sollte mit der Textthematik binden.“

Eventuelle Abweichungen können wegen Defizite in der Prüfungsanleitung, und auch wegen unklarer Konstruktdefinition hervortreten, da die Lehrkraft bei der Erstellung der beiden Prüfungsvariablen die getesteten Kompetenzen nicht angemessen proportioniert. In der im Blue-Print vorgegebenen Checkliste zur Beurteilung der Lernerantworten auf die Prüfung, steht fest, dass die Themen, die in der Schreibfähigkeit behandelt werden, einen Bezug mit der Textthematik haben müssen.

Die Analyse hat ergeben, dass der sprachliche Input der Thematik angemessen ist und in seiner geschriebenen Form einheitlich und kohärent dargeboten wird. Die Lehrkraft hat in der ausgewählten Prüfung ein Thema gewählt, das dem Themenkatalog, der im Lehrplan festgelegt wurde nicht widerspricht. Das Thema betrifft den Einfluss der Massenmedien auf den Konsum und vertritt auch eine problematische Frage bezüglich Massenmedien. Was die

---

<sup>142</sup> S. Theoretischer Teil der vorliegenden Arbeit, S. 25

<sup>143</sup> ONEC. (2010 ):Prüfungsanleitung, S. 05

innere Ebene der Prüfung anbetrifft, erfüllt die Prüfung mit Vorbehalt dieses Kriterium, da die Prüfungsanleitung keine Hinweise darauf enthält, dass die Lehrkraft auch in den anderen Subtests „ Wortbildung und Grammatik“ in dem Kontext bleibt, d. h. Beispiele geben, die der Textthematik entsprechen.

Darüber hinaus, wurde nicht in der Prüfungsanleitung darauf hingewiesen, dass die Lehrkraft so genau wie möglich ähnliche Strukturen in den zwei Testvariablen geben muss, d. h. die zwei Tests müssen fast einen gleichen Schwierigkeitsgrad besitzen.

Was das vierte Kriterium anbelangt, ob die Beziehungen zwischen Input und der aufgeförderten Reaktion einheitlich sind, so dass sie die Lernerperformanz nicht beeinflussen, hat die Analyse ergeben, dass die Hauptschulprüfung ein nicht reziprokaler Test ist. Die Lerner erhalten Feedback über die Relevanz ihrer Antworten erst wenn die Lehrkraft die Kopien korrigiert hat.

Die Beziehungen zwischen Input und der aufgeförderten Antwortform sind einheitlich, so dass die Lernerperformanz dadurch nicht beeinträchtigt werden könnte.

Bezüglich des fünften Kriteriums zur Analyse der Testreliabilität, das die Charakteristika der Beziehungen zwischen Input und der erwarteten Reaktion betrifft, wurde festgestellt, dass die Prüfung einen einzigen Teil vorlegt. Die Lerner werden nicht einem anderen Test ausgesetzt im Rahmen derselben Prüfung, was positiv zu bewerten ist. Dies trägt dazu bei, die Prüfung lernerfreundlich zu machen. Es wurde auch bei der Analyse festgestellt, dass die allzu Konzentrierung auf die Überprüfung des sprachlichen Regelwissens im zweiten Subtest lediglich einem Hang zur Methodenvielfalt dient und den in der Prüfungsanleitung tatsächlichen Testzielsetzungen widerspricht.

Die algerischen curricularen Richtlinien richten sich an der kommunikativen Methode, die in der Grammatik nur ein Mittel zum Zweck sieht. Davon leitet sich das Prinzip ab, dass das Testen der Grammatik nicht explizit erfolgen muss, sondern im Kontinuum der Lerner Sprachkompetenz als integrierende Kompetenz. Mit anderen Worten: die Respektierung der grammatischen Regeln muss in dem dritten Subtest „ Schreibkompetenz“ überprüft werden.

### **2.1.1 Ergebnisse und Auswertung**

Nachdem das Prüfungsmodell des algerischen Erziehungsministeriums im Blick auf die Erfüllung des Kriteriums der Reliabilität untersucht wurde, wird nun auf die Auswertung eingegangen. Die Auswertung will feststellen, ob die Reliabilitätskriterien bei der Erstellung

dieser Prüfung berücksichtigt wurden. Was das erste Kriterium anbelangt, ob die Charakteristika der Testkonstrukts einheitlich sind, d. h. ob die Zeitzuweisung und der Prüfungsraum, sowie die Charakteristika des Prüfungspersonals berücksichtigt wurden oder nicht, hat die Analyse ergeben, dass die materialen Bedingungen weitgehend erfüllt sind. Bezüglich des zweiten Kriteriums, stellt die Zeitzuweisung 3 Stunden dar. Die äußeren Rahmenbedingungen sind weitgehend erfüllt und mit denen des Abiturs weitaus kongruent. Die Merkmale der Testrubrik sind nur teilweise erfüllt. Dies wird im Laufe dieser Forschung erklärt und nachgewiesen. Vielmehr stellt die Überprüfung der Testrubrik eine Kernaufgabe der vorliegenden Forschung.

Die Arbeitsanweisungen sind in einer geschriebenen Form dargeboten, so dass sie die Performanz der Lerner nicht verschiedenweise beeinflussen. Die Prüfungsanleitung weist ein Defizit im Fall einer Doppelthematik auf. Die Charakteristika der Prüfungsvariablen variieren bei der Testerstellung, so dass sie die Lernerperformanz beeinflussen könnten, da die Prüfungsanleitung keine genauen Anweisungen bietet. Dieses Defizit in der Prüfungsanleitung führt zu ernster Beeinträchtigung der Konstruktvalidität. Was die Testergebnisse anbelangt, hat die Analyse ergeben, dass die Prüfung das Kriterium der Reliabilität nicht erfüllt; das Nicht-Vorhandensein einer detaillierten Punkteskala führt zur divergierenden Benotung und folglich zur mangelnden Objektivität. Bezüglich der Beurteilungsmethode, erfüllt die Prüfung das Kriterium des Impacts indem das Testergebnis in 3 vervielfältigt wird. Was die Punkteverteilung anbelangt, weist die Beurteilungsmethode gravierende Defizite auf, da die Prüfungsanleitung keine detaillierte Punkteskala vorlegt. Was das dritte Kriterium anbelangt erfüllt die Prüfung dieses Kriterium nur teilweise. Im Hinblick auf die Form ist die Prüfung in geschriebener Form dargeboten und somit ist das Kriterium der Form erfüllt.

Auf der inneren Ebene weist die Prüfung Defizite in der Prüfungsanleitung, da die Prüfung zwei Themen vorlegt, könnte der Input in den zwei Prüfungsvariablen variieren und zur Beeinträchtigung der Objektivität führen.

Tabelle1. Ergebnisse der Konstruktvalidität in der Hauptschulprüfung des Gymnasialen DaF-Unterrichts.

	erfüllt	teilweise erfüllt	nicht erfüllt
Inwiefern sind die Charakteristika der äußeren	×		

Rahmenbedingungen einheitlich.			
Inwieweit sind die Charakteristika der Testrubrik einheitlich.		×	
Inwieweit sind die Charakteristika des Testinputs einheitlich. (Eindeutigkeit des Testinputs)		×	
Die Charakteristika der aufgeführten Reaktion.		×	
Die Charakteristika der Beziehungen zwischen Input und erwartete Reaktion.		×	
Die Charakteristika der Beziehungen zwischen Input und erwarteter Reaktion bei jeder Testdurchführung und in jedem Testteil.		×	

*Quelle: Eigene Darstellung*

Nachdem die Reliabilität des Tests nachgeprüft wurde, wird nun die Analyse auf die Überprüfung der Konstruktvalidität fokussieren, d. h. es wird überprüft, zu welchem Grad stimmen die Fertigkeiten und Kompetenzen im Testziel mit dem Testkonstrukt überein, so dass die Inferenzen über die Lerner in Bezug auf die getesteten Kompetenzen, die aufgrund dieser Tests erhoben werden richtig sind. Die Analyse will die Konstruktvalidität der Hauptschulprüfung im algerischen gymnasialen DaF-Unterricht überprüfen. Die Kriterien anhand deren ich die Prüfung analysieren werde, fokussieren auf Aspekte des Testdesigns, die die Performanz der Lerner beeinträchtigen könnten. Diese teilen sich wiederum in Kriterien, die sich auf die Beziehung zwischen Konstruktbeschreibung und Testaufgaben beziehen und Andere auf eventuelle Überschreitungen in den Aufgabencharakteristika. In Bezug auf die klare Beschreibung des Testkonstrukts wurde festgestellt, dass die Prüfung klare und genaue Beschreibung der Kompetenzen, die der Test misst darbietet, hierzu handelt es um die Überprüfung des Leseverstehens und der Schreibkompetenz des Lerners.

„L’épreuve d’allemand vise à évaluer chez le candidat la compréhension d’un texte, la maîtrise de la langue et sa compétence à l’expression écrite.“<sup>144</sup> Über den Text hinaus, beinhaltet die Prüfung drei Hauptteile, die den drei sprachlichen Kompetenzen entsprechen. Es werden:

- Der Input in Form eines Textes unterschieden.
- Textverständnis: Diese Aufgabe zielt auf die Überprüfung des Leseverstehens und beinhaltet drei Aufgaben; eine Richtig/Falsch- Aufgabe, Fragen zum Text, eine Kombinationsaufgabe.

---

<sup>144</sup> S. Prüfungsanleitung. (2010, S. 04)



Der zweite Teil überprüft die Sprachkompetenz, es wird auf die lexikalischen als auch auf die grammatischen Kenntnisse der Lerner betont.

Der Lexik-Teil schließt drei Aufgaben ein:

- Eine Aufgabe, in der die Lerner Synonyme aus dem Text herausfinden sollen.
- Einen kleinen Cloze-Test zum Ausschuchen der Gegenteile vorgegebener Wörter aus dem Text.
- Eine Wortbildungsaufgabe.
- Einen Satz aus dem Text zum Übersetzen.

In dem zweiten Hauptteil der Prüfung wird die sprachanalytische Kompetenz der Lerner überprüft. Der dritte Prüfungsteil überprüft die Kompetenz der Lerner einen Aufsatz zu einem der Zielthemen zu schreiben. Es werden dem Lerner zwei Themen, die der Thematik des Textes entsprechen vorgeschlagen. Es wird in Bezug auf die klare Beschreibung des Konstrukts festgestellt, dass die Prüfungsanleitung klare und detaillierte Anweisungen über die Kompetenzen, die die Prüfung testet, liefert. Vielmehr werden bei jeder Aufgabe bestimmte Kompetenzen geprüft. Die Analyse der Prüfung hinsichtlich dieses Kriteriums hat ergeben, dass sie dieses Kriterium weitgehend erfüllt, so dass die Beurteilung der Testergebnisse aufgrund der Kompetenzen, die im Test getestet werden sollen, von der Lehrkraft gemacht werden kann. Es wurde auch hervorgehoben, dass die Lerner die Begriffe „Textverständnis“ und „Schreibkompetenz“ nicht verstehen, was die Interpretation ihrer Ergebnisse in Bezug auf die Kompetenzen, die sie erwerben sollten unmöglich macht. Nur die Lehrkraft kann das Beurteilungsverfahren durchführen. Bezüglich des siebten Kriteriums zur Evaluation der Konstruktvalidität der Hauptschulprüfung hat die Analyse ergeben, dass der erste Teil der Prüfung, der das Leseverstehen testet, hinsichtlich der getesteten Kompetenz das Kriterium mit Vorbehalt erfüllt.

Die drei Aufgaben der Prüfung „Textverständnis, Sprachkompetenz und Schreibkompetenz“ entsprechen der Form nach und hinsichtlich des Testzwecks dem oben genannten Kriterium dadurch, dass sie die drei Kompetenzen, die die Prüfung testet, darstellen. Die erste Testaufgabe wird zum Überprüfen des detaillierten Textverständnisses eingesetzt und entspricht folglich dem Testkonstrukt bzw. Textverständnis. Die dritte Aufgabe stellt eine Kombinationsaufgabe dar. Der Lerner sollte die vorgegebenen Elemente „Wörter“, die semantisch miteinander hängen kombinieren. Dieser Aufgabentyp eignet sich nicht in erster Stelle zur Überprüfung des Leseverstehens. Der Lerner wird beim Lösen dieser Aufgabe nicht seine Kompetenz den vorgegebenen Text zu verstehen nachweisen, sondern ein Beleg

für seine Wortschatzkenntnisse stellen. Ferner können die Lerner die Aufgabe auch ohne Rückgriff auf den Text lösen.

Wie bei dem folgenden Beispiel ersichtlich wird:

### **Kombinieren Sie!**

a-Schreiben	1- Kolonialismus
b-Unabhängigkeit	2-Filme
c-Regisseur	3- Schriftsteller

Beim Überprüfen des Leseverstehens wird der Text als Ganze erzielt. Auch wenn die ausgewählten Wörter aus dem Text entnommen werden, dient diese Übung zur Überprüfung des Leseverstehens nicht. Dieser Tatbestand ist einem Defizit in der Prüfungsanleitung zurückzuführen. Es wird nicht erwähnt, dass die Lehrer bei der Auswahl der Elemente sich auf dem Text beschränken müssen. Hier erlaube ich mir kurz BACHMAN und PALMER zu zitieren: „asking the test taker to describe a picture orally would not adequately reflect a construct broadly defined as the ability to participate in conversations”<sup>145</sup>

Bezüglich der dritten Aufgabe, die das Leseverstehen betrifft, soll der Lerner einen Satz aus dem Text erklären. Diese Frage weist einen hohen Schwierigkeitsgrad auf. (siehe, Lernerkopienanalyse) 2,5% der Lerner antworten richtig auf die Frage.

7,5 der Lerner lösen die Aufgabe nicht.

90% der Lerner geben falsche Antworten an.

Die Arbeitsanweisungen zu dieser Aufgabe sind mehrdeutig. Mit der Arbeitsanweisung: “Erklären Sie!“ könnte der Lerner den Satz mit anderen sprachlichen Mitteln ausdrücken. In diesem Fall riskiert der Prüfer bei den Lernenden auch die Vokabel und nicht das Leseverstehen zu überprüfen. Bei der durchgeführten Kopienanalyse haben einige Übersetzungen vorgefunden, einige Lerner haben den Satz direkt ins Arabische übersetzt. Von der 4. Arbeitsanweisung könnten andere Aufforderungen als die Erklärung des Satzes verstanden werden, wie beispielsweise die Übersetzung. Diese Aufgabe misst in diesem Fall die Kompetenz der Lerner die Arbeitsanweisungen zu verstehen, oder den Satz zu übersetzen und nicht das Leseverstehen, was die Benotungsverfahren erschwert. Nach der der Prüfungsanleitung beigelegten Punkteverteilung wird im Fall der richtigen Antwort dieser

---

<sup>145</sup> BACHMAN/PALMER( 1996)

Frage einen Punktwert von 0,75 zugewiesen. Problematisch bleibt, dass die Punkteskala die Richtigkeitskriterien nicht eingibt. In der letzten Aufgabe soll der Lerner einen Titel für den Text eingeben. Dadurch kann eine Idee über das Lerner- Textverständnis erhoben werden, indem der Titel eine allgemeine Idee über das Textthema liefert.

Der dritte Aufgabenteil überprüft die Lernerkompetenz einen Paragraphen über eine bestimmte Thematik zu schreiben; es werden zwei Themen zur Wahl vorgeschlagen; eine offene Aufgabe in der sich der Lerner frei ausdrückt und die Zweite ist eine halboffene Aufgabe, in der die Lerner anhand vorgegebener Stichpunkte einen Paragraphen schreiben.

Diese Aufgabenformen bzw. die Offene und Halboffene eignen sich am besten zur Überprüfung der Schreibkompetenz. Folglich reflektiert die Aufgabenform des dritten Teils die Konstruktbeschreibung der Prüfung und erfüllt somit das dritte Kriterium, das sich auf die Entsprechung der Testaufgaben mit dem Testzweck bezieht. Bezüglich des vierten Kriteriums, ob die Benotungsverfahren die Konstruktbeschreibung reflektieren, stützen sich die Lehrer bei der Beurteilung der Lernerergebnisse auf die vorgegebene Punkteverteilung. Kriterium ist erfüllt.

Was das fünfte Kriterium anbelangt, ob die Testergebnisse den Lehrkräften dabei helfen, die erwünschten Interpretationen über die Lenersprachkompetenz zu erheben, wurde anhand der obigen Analyse ersichtlich, dass die Prüfung die Kompetenz der Lerner einen geschriebenen Text zu verstehen, und die Lernerkenntnisse der sprachlichen Elemente Wortbildung, Grammatik, Wortschatz und die Lernerkompetenz kleine Sätze zu übersetzen überprüft.

Dieser zweite Hauptteil „Sprachkompetenz“ der Prüfung beschränkt sich auf die Überprüfung der Sprachanalytischen Kompetenz der Lerner und testet nicht die Sprachkompetenz der Lerner als die Kompetenz sich der Sprache in realen Kommunikationssituationen zu bedienen. Darüber hinaus erhält der Lerner im Fall einer richtigen Antwort auf den zweiten Teil 8 Punktwerte; in der Sprachanalyse allein. Die von den Lernern erbrachten Ergebnisse in dem Sub-Test „Sprachkompetenz „ sagen nur wenig über die Lenersprachbeherrschung, da die Kenntnis der sprachlichen Elemente kein reliabler Indikator der Sprachbeherrschung ist.

Die Ergebnisse dieser Aufgabe können nicht die erwünschten Informationen über die Lenersprachkompetenz vermitteln. Was die Schreibkompetenz anbelangt, werden theoretisch die erlernten Strukturen und sprachlichen Elemente, sowie das landeskundliche Wissen in Schreiben eines Paragraphen in der deutschen Sprache eingesetzt. Der Lerner erhält im Fall einer optimalen Leistung den Punktwert „5“.

Diese Zuweisung erfolgt wie folgt: Für die richtige Planung des Aufsatzes bzw. seiner Einteilung in einer: Einleitung, Körper, und Schlussfolgerung, erhält der Lerner einen Punktwert (01). Es sei hier darauf hingewiesen, dass die Punkteskala es nicht erklärt, ob mit dem "Plan" nur die Form gemeint worden ist oder auch die inhaltliche Graduierung. In diesem Fall kann das erbrachte Ergebnis nicht viel über die Lernerkompetenz methodisch zu schreiben bzw. ihre Methodenkompetenz indizieren. Dem Stil, Rechtschreibung und Wortschatz gibt der Test einen Punktwert von 3 Einheiten an, d.h. drei- Punkte werden zugewiesen. Es wird zwar in der Prüfungsanleitung nicht detailliert, was mit Stil gemeint worden ist. Wenn der Lerner beim Schreiben die grammatischen Regeln respektiert, erhält er Dafür einen Punkt.

In der Teilaufgabe „Sprachkompetenz“ erhält der Lerner 3 Punkte für die Wortschatzkenntnisse und einen zusätzlichen Punkt im Subtest „Schreibkompetenz“. Dies widerspricht dem didaktischen Prinzip, dass der Lerner nur einmal in einem sprachlichen Bereich getestet werden muss. Was die grammatischen Kenntnisse anbelangt, wird der Lerner zweimal in dem Subtest Grammatik, und in dem dritten Subtest Schreibkompetenz getestet. Der Lerner erhält 5 Punkte im Subtest Grammatik und einen Punktwert in der Testaufgabe „Schreibkompetenz“ für die richtige grammatische Gestaltung des Aufsatzes. Insgesamt bekommt der Lerner 11 Punkte /20 für die sprachlichen Mittel allein. Es wird da riskiert den Lerner für die gleichen Fehler mehrmals zu benachteiligen. Nach dieser Analyse hat sich herausgestellt, dass die Evaluation der Testgültigkeit anhand der Testergebnisse eine Aufgabe, die jeder Lehrer durchführen kann. Die sprachlichen Mittel werden zwei Mal überprüft, was die Validität der Aufgabe beeinträchtigt. Problematisch ist, dass die Prüfungsanleitung nicht darauf hinweist, wie der Prüfer sich im Fall eines rekurrierenden Fehlers verhält. Die Punkteverteilung der Teilaufgabe „Schreibkompetenz“ hat das Kriterium erfüllt, da die Schreibfähigkeit auch die sprachlichen Elemente Grammatik, Rechtschreibung und Stil einschließt. Was die eventuellen Quellen der Unobjektivität in den Teilaufgaben anbelangt, wird die Hauptschulprüfung im algerischen DaF-Unterricht im Hinblick auf die Testcharakteristika, die die Lerner dazu bringen, verschiedenerweise zu leisten bzw. die Lernerperformanz in dem Test beeinflussen analysiert. Nach dieser Analyse wurde klar, dass die Testergebnisse potenzielle Messfehler aufweisen. Das Kriterium ist nur teilweise und nämlich im ersten Sub-Test erfüllt. Bezüglich des ersten Kriteriums, ob die Testzüge der äußerlichen Bedingungen des Tests für alle Lerner identisch sind, dass sie die Lernerleistungen in dem Test nicht beeinflussen, hat die Analyse ergeben, dass die Tests in Klassenzimmern statt finden, die einen vertrauten Raum für die Lerner darstellen. Die

Testadministratoren sind in den meisten Fällen Gymnasiallehrer. Nach meinen empirischen Beobachtungen bzw. Hospitationen in algerischen Gymnasien in Setif und Bordj bou Arreridj habe ich festgestellt, dass die physikalischen und äußerlichen Bedingungen weitgehend kongruent sind, und keine Differenzen in der Lernerperformanz verursachen.

### **2.1.1 Analyse der Prüfungsanleitung**

Die Analyse der Prüfungsanleitung hat ergeben, dass sie keine Hinweise auf die Berücksichtigung der äußerlichen Bedingungen des Tests enthält. Bezüglich des zweiten Kriteriums, das den Einfluss der Testrubrikzüge auf die Lernerperformanz betrifft, so dass die Lerner verschieden in dem Test leisten, wird die Prüfung in schriftlicher Form dargeboten. Was die dritte Aufgabe der Teilaufgabe anbetrifft, hat sich herausgestellt, dass die Arbeitsanweisung zu dieser Aufgabe unklar erstellt wurde, dass einige Lerner davon das Übersetzen verstehen. Die Lernerkopienanalyse hat ergeben, dass einige Lerner den Satz übersetzt haben anstatt ihn mit anderen lexikalischen Mitteln zu erklären.

Was die einheitliche Form der Prüfung anbelangt, hat es sich herausgestellt, dass die Prüfung in weitgehend einheitlicher Form dargeboten wird. Dasselbe gilt für die 13. und 14. Kriterien. Bezüglich des 15. Kriteriums erzielt die Prüfung die Lerner des 3. Studienjahres d, h. für eine Lerngruppe, die einer zweijährigen Ausbildung in deutscher Sprache ausgesetzt wurden. Die betreffende Lerngruppe stellt eine in hohem Maß homogene Gruppe dar, die die gleichen Chancen gehabt haben, so dass die aufgeforderte Reaktion keinen Lerner auf den anderen benachteiligt. Somit wird das 15. Kriterium, das die Charakteristika der Beziehungen zwischen Input und aufgeforderter Reaktion betrifft weitgehend erfüllt.

In Bezug auf die Beschreibung des Testkonstrukts, widerspricht nicht die Prüfung den Zielstellungen in der Prüfungsanleitung zum Leseverstehen, zu den sprachlichen Elementen und den grammatischen Strukturen sowie zur Schreibfähigkeit. Die Validierung der Prüfung hat ergeben, dass die erste Teilaufgabe das Kriterium der Konstruktvalidität mit Vorbehalt erfüllt, da die Richtig/Falsch Aufgabe zur Überprüfung des Leseverstehens wenig zuverlässig ist<sup>146</sup>. Es hat sich herausgestellt, dass die Teilaufgaben den in der Prüfungsanleitung festgelegten Testzielen entsprechen. Was die Kombinationsaufgabe in der ersten Aufgabe

---

<sup>146</sup> Siehe: Theoretischer Teil der vorliegenden Arbeit. S: 24

anbetrifft, eignet sie sich nicht zur Überprüfung des Leseverstehens. Die Lerner können diese Aufgabe lösen ohne auf den Text zurückzugreifen.

Die Analyse der vierten Aufgabe im Subtest Leseverstehen hat ergeben, dass die Frage:“Erklären Sie diesen Satz“ den behandelten Inhalten im Lehrbuch nicht entspricht und weist somit in Bezug auf das Kompetenzkonstrukt eine völlige Divergenz auf. Es hat sich auch herausgestellt, dass die Aufgabe 4 in der Teilaufgabe Leseverstehen einen hohen Schwierigkeitsgrad aufweist. Die offene Form der Aufgabe lässt die genaue Festlegung der Richtigkeitskriterien nicht möglich, was die Lehrkraft dazu führt, holistische Beurteilungen zu geben. Die Analyse der 5. Aufgabe hat ergeben, dass sie in Bezug auf die Konstruktvalidität die Kriterien erfüllt. Der zweite Prüfungsteil entspricht der Kompetenzbeschreibung im Curriculum nicht, da sich der Deutschunterricht im algerischen Gymnasium dem kommunikativen Ansatz verpflichtet. Das Testen des sprachlichen Regelwissens des Lerner ist in diesem Fall überflüssig und benachteiligt den Lerner, der im dritten Subtest auch gefordert ist, grammatisch richtige Sätze zu schreiben. Die Analyse hat bezüglich der dritten Teilaufgabe ergeben, dass die offene Form der Aufgabe zur Überprüfung des Kompetenzkonstrukts „Schreibkompetenz“ angemessen ist, und demzufolge valide ist.

Bezüglich der Inferenzen, die aufgrund der Testergebnisse über die Lernerkompetenzen gemacht werden können, hat die Analyse ergeben, dass der zweite Subtest “Sprachkompetenz“ nur wenige Daten über die sprachliche Kompetenz der Lerner erhebt. Die Analyse der Punkteverteilung in Bezug auf das Testkonstrukt hat ergeben, dass die grammatische sowie die Wortschatzfähigkeit der Lerner wiederholt überprüft werden, was eine Überschreitung der Konstruktvaliditätsprinzipien erkennen lässt.

**Tabelle 2** Ergebnisse der Evaluation der Reliabilität von Hauptschulprüfung.

	erfüllt	nicht erfüllt
Das Testkonstrukt ist klar definiert.	×	
Das Testkonstrukt entspricht dem Testzweck.		×
Die Testaufgaben spiegeln die Konstruktbeschreibung des Tests wieder.	×	
Inwiefern reflektieren die Beurteilungsverfahren die Konstruktbeschreibung.		×
Die Testergebnisse erlauben der Lehrkraft die genauesten Interpretationen über die Lersprachkompetenz zu bekommen.		×
Die Wesenszüge der Testrubrik deskriminieren nicht zwischen den Lernern.	×	

*Quelle: Eigene Darstellung*

Nachdem die Validität des Tests bzw. die Übereinstimmung der Testinhalte mit den Testzielen überprüft wurde, wird nun die Authentizität der Prüfung überprüft d. h., die Übereinstimmung der Testinhalte mit den Charakteristika der Sprachzielaufgaben (TLU) oder Handlungen, die wie im theoretischen Teil schon ersichtlich wurde, vom Design- Statement festgelegt werden müssen. Im Folgenden wird die Prüfung für algerische DaF-Gymnasiasten hinsichtlich ihrer Authentizität überprüft werden. Die Authentizität ist für die Testvalidierung unerlässlich. Da die Testergebnisse bei der Selektion entscheidend sind, sollten auch die Lernerperformanz in dem Test dem Zielsprachgebrauch, sowie dem instruktionalen Bereich angemessen sein.

Das erste Kriterium betrifft die klare Beschreibung der Handlungen in realen Sprachgebrauchsbereichen und der Informationen über die äußerlichen Bedingungen, den Input und die aufgeförderte Reaktion sowie, den Zusammenhang zwischen Input und Reaktion in dem Blue-Print(Prüfungsanleitung).

Bevor die Analyse hinsichtlich dieses Kriteriums angefangen wird, soll darauf hingewiesen werden, dass die Authentizität eines Tests in Bezug auf das Design-Statement, worüber jeder Test verfügen muss, gemessen wird. Es wird zuerst festgestellt, dass das Design-Statement in dem die nötigen Informationen für die Analyse vorzufinden sind, nicht vorhanden ist. Als Alternative wird auf das Curriculum zurückgegriffen, in dem die Sprachgebrauchsbereiche sowie der Lehrbereich explizit festgelegt wurden. Bezüglich des ersten Kriteriums steht im Curriculum Folgendes:

„Le programme d’Allemand doit contribuer à construire des compétences à la communication. De ce fait, il met l’accent sur une perspective de l’emploi de la langue en situation. Les apprenants a l’issue de l’enseignement secondaire, auront : Acquis des bases fondamentales de la langue, utilisé la langue dans les situations d’interlocution pour différents buts en prenant en compte les contraintes de la vie sociale.

Les compétences à installer en troisième année secondaire :

Acquérir des entretiens sur l’ensemble des thèmes traités.

Acquérir des connaissances diverses spécifiques aux pays germanophones.<sup>147</sup>

---

<sup>147</sup> Erziehungsministerium (2006) Algerisches geltendes Curriculum für Deutsch im Gymnasium.

Die Analyse hat ergeben, dass bezüglich des Input-Settings, findet der Zielsprachgebrauch (TLU) im deutschen Sprachraum oder im algerischen Sprachraum statt. Die Sprachgebraucher sind deutsche Muttersprachler oder deutsche Touristen, sowie algerische Germanisten. Die Themen, die die Lerner behandeln, sind weltweit bekannte Probleme wie Umweltverschmutzung und Computerrevolution. In der ersten Lektion: „Die deutschsprachigen Länder und Algerien heute“, wird ein interkultureller Ansatz versucht. Es wird im Curriculum zwar betont, dass der Lerner landeskundliche Informationen über die deutschsprachigen Länder und interkulturelle Konzepte aneignen sollte. In dem vorgegebenen Text der Prüfung muss es um eine aktuelle Thematik der 9 vorgeschlagenen Themen im Curriculum handeln, wie u.a.: Die deutschsprachigen Länder und Algerien heute, Der technische Fortschritt, Die Umweltprobleme.

Diese Themen stellen die Schwerpunktthemen, die die Jugend in dieser Epoche interessieren. Die aufgeforderte Reaktion ist, dass der Lerner in der Lage wird, Informationen über diese Themen zu erwerben, einen Standpunkt zu diesen Themen zu entwickeln und Ereignisse in Bezug auf erlebte Erfahrungen zu erzählen. Die Beschreibung des Sprachgebrauchsbereiches ist in den Lerninhalten des Curriculums vorhanden, wenn auch nicht mit der entsprechenden Terminologie, die von BACHMAN und PALMER(1996) eingebracht wurde. Bezüglich des 17. Kriteriums, ob die Eigenschaften der Testaufgaben denen des TLUs entsprechen, wurde festgestellt, dass die Prüfung den mündlichen Gebrauch der Sprache ausschließt, und widerspricht somit der Beschreibung des TLUs im Curriculum indem der Lerner die mündlichen Sprachkompetenzen erwerben soll, und einen universitären Studium in dem Bereich aufgreifen kann. Es wird in der Prüfungsanleitung festgelegt, dass die Lerner auf die Inhalte, d. h. den instruktionalen Bereich der Sprache stützen und sich darauf einschränken. Dies impliziert die Entsprechung der Testaufgaben mit den Lernaufgaben.

Es hat sich herausgestellt, dass die Konzentrierung auf die Sprachanalyse das Prinzip der kommunikativen Orientierung des Deutschunterrichts überschreitet; der Grammatik werden 5 Punkte zugewiesen, was die Erhebung der genauen Inferenzen über die kommunikativen Kompetenzen der Lerner erschwert oder verhindert.

Darüber hinaus werden die Lerner auch in den anderen Testteilen aufgefordert grammatisch richtige Sätze zu schreiben und werden dazu benotet.

Die Prüfung widerspricht den Prüfungszielen nicht, aber der Konzeption des Curriculums über die Kompetenzen, die die Lerner erwerben sollten, da das Testziel ohne Berücksichtigung der festgelegten Prinzipien im Curriculum verwirklicht wurde.



Die Analyse zur Überprüfung der Textauthentizität der Hauptschulprüfung für algerische DaF-Gymnasiasten hat ergeben, dass der Bereich des Zielsprachgebrauchs, sowie der situative Sprachgebrauch im Curriculum festgelegt sind. Es wurde auch anhand der durchgeführten Analyse festgestellt, dass die Testzwecke bzw. das Verstehen eines Textes in der Zielsprache Deutsch, die Sprachbeherrschung die Schreibkompetenz in der Prüfungsanleitung festgelegt sind. Es hat sich herausgestellt, dass die Prüfung im Hinblick auf den Testzweck des zweiten Testteils, der die grammatische Kompetenz testet, den Zielsetzungen des Curriculums widerspricht, da die Förderung der Grammatik nicht im Vordergrund des Curriculums steht. Die Bereiche der Zielsprache sind zwar beschrieben, wenn auch nicht mit der entsprechenden Terminologie.

Im Curriculum werden die Bereiche der Zielsprache wie folgt festgelegt:

Bezüglich der ersten Lektion:

- Die Lerner sollen das entsprechende Wissen erwerben, anhand dessen sie die deutschsprachigen Länder identifizieren oder erkennen können.
- Auf der Ebene der Phonetik müssen sie die Ländernamen richtig aussprechen.
- Der Erwerb von Vergleichsfunktionen.
- Der Lerner soll über das nötige Wissen verfügen um Konversationen zu führen.

In der zweiten Lektion, soll der Lerner über das notwendige Wissen verfügen, um eine bekannte Persönlichkeit zu identifizieren. Die entsprechenden Bibliographien werden aus der deutschen Landeskunde und auch aus der algerischen Geschichte gewählt. Es werden u.a. Bach, Albert Einstein, Abdelhamid Ben Badis gewählt.<sup>148</sup>

Es wird dem Test die allzu Konzentrierung auf die Sprachanalyse entgegengehalten, was die kommunikative Orientierung des algerischen gymnasialen Deutschunterrichts widerspricht und folglich die Authentizität der Prüfung verringert. Es wurde anhand der durchgeführten Analyse der Prüfungsanleitung festgestellt, dass die Prüfung die Testzwecke, ausgenommen die allzu Konzentrierung auf die Grammatik im zweiten Subtest nicht widerspricht und folglich in Bezug auf die Prüfungszwecke mit Vorbehalt authentisch ist.

In Bezug auf die curriculare Konzeption hat die Analyse ergeben, dass die Prüfung hinsichtlich des Authentizitätskriteriums gravierende Defizite aufweist, da die Prüfung ohne treue Berücksichtigung der im Curriculum festgelegten Prinzipien erstellt wurde.

---

<sup>148</sup> Weitere Felder des Zielsprachgebrauches sind in dem geltenden Curriculum für das algerische Gymnasium (2006, 8) vorzufinden.

**Tabelle 3** Ergebnisse der Testanalyse im Hinblick auf die Authentizitätskriterien.

	teilweise erfüllt	erfüllt	nicht erfüllt
Die Beschreibung der Aufgaben des TLUs beinhaltet Informationen über Input und die erwartete Antwort sowie die Beziehung zwischen Input und der erwarteten Reaktion.	×		
Inwiefern die Charakteristika der Testaufgabe denen des TLUs entsprechen.	×		

*Quelle: Eigene Darstellung.*

Die Interaktivität in einem Sprachtest ist der Grad zu welchem die sprachlichen Charakteristika der Lerner in der Testerstellung mitberücksichtigt sind.

Die wichtigsten individuellen Charakteristika anhand deren die Analyse durchgeführt wird sind:

-Die sprachliche Kompetenz des Lerners, d.h. das Sprachwissen, die strategische Kompetenz, Fachwissen und affektive Schemata. Die Divergenzen in der Lernerperformanz werden grundsätzlich zu diesen Differenzen zurückgeführt. Ein interaktiver Test zum Beispiel ist ein Test, der die sprachlichen Kenntnisse und strategischen Kompetenzen des Lerners einruft.

Im vorherigen Kapitel wurde die Authentizität des Tests überprüft, d, h. den Übereinstimmungsgrad zwischen Testaufgaben und dem instruktionalen Bereich. Im Folgenden wird das Kriterium der Interaktivität überprüft bzw. den Übereinstimmungsgrad der Testaufgabencharakteristika mit den persönlichen Eigenschaften der Probanden. Bezüglich des 18.Kriteriums, inwiefern stellen die Testaufgaben die entsprechenden Kompetenzbereiche des Fachwissens dar, und inwiefern stimmen die Kompetenz und der Lehrbereich mit den Eigenschaften der Lerngruppe überein.

Was die Kompetenzen, die die Lerner in der gymnasialen Ausbildung erwerben sollten, anbetrifft, sollen die Lerner allgemeine Kompetenzen aneignen wie die Vertrautheit mit den Fremdsprachen. Sie sollen auch in der Lage sein, einen Input in der Zielsprache zu dekodieren, mit anderen Kulturen der Welt vertraut zu sein, mündliche und schriftliche Gespräche in der Zielsprache zu führen, eine kritische Haltung gegenüber dem riesigen

Schwung der Informationen einzunehmen. In der ersten Aufgabe sollen die Lerner Fragen zu einem Text antworten. Dies impliziert die Einsetzung der vorher erworbenen Kenntnisse und Kompetenzen u.a. Wortschatz, Wortbildung Syntax und Grammatik. Der Lerner muss beim Lesen des Textes, sowie bei der Antwort auf die Aufgaben seine Kenntnisse der sprachlichen Elemente nachweisen und affektive Schemata einsetzen, die für das Leseverstehen erforderlich sind. Es sei hier auch hingewiesen, dass das Lesen des Textes zur Lösung der Aufgabe gehört. Die Analyse hat ergeben, dass der Tatbestand, dass der Prüfungsaufbau auf einem Text in der Zielsprache beruht, hohe Ansprüche an die Lerner erhebt und somit hohe Interaktivität realisiert. In der Richtig- Falsch- Aufgabe setzt der Lerner seine Erinnerungskompetenz ein bzw. sein Kurzzeitgedächtnis; er soll auf den Text zurückgreifen um von der Richtigkeit seiner Antwort sicher zu sein. Hier sollte er auch sein Leseverständnis einsetzen, was auch das Wiederrufen seiner Vorkenntnisse und Fertigkeiten benötigt. Je umfangreicher die Fülle der Informationen und Kenntnisse, die der Lerner über den Text besitzt, desto höher wird die Interaktion zwischen den Lernereigenschaften und dem Text.

Ein hoher Interaktionsgrad hängt zusammen mit dem Richtigkeitsgrad. In diesem Zusammenhang ist eine Aufgabe potenziell valide, wenn sie einen hohen Interaktivitätsgrad mit den Lerner- eigenschaften realisiert. Mit anderen Wörtern, erheben valide Aufgaben hohe Ansprüche an die Lernereigenschaften.

Der zweite Teil überprüft die Sprachkompetenz der Lerner. Hinsichtlich des 15. Kriteriums erweist er relativ einen wenigen Interaktivitätsgrad. Bei der Synonymfindung setzt der Lerner seine Leseverständniskompetenz ein. Hierzu soll der Lerner ein Synonym finden, das in einem Satz eingebettet wird. Das entsprechende Wort soll im Text gefunden werden. Diese Methode hilft dem Lerner dabei den genauen Sinn des Wortes zu erkennen, und dies durch das Einsetzen seiner erworbenen Kenntnisse. Dasselbe gilt für die zweite Aufgabe des Wortschatzes, in der der Lerner Gegenteile in dem Text finden soll. Der Grammatiktestteil veranlasst den Lerner dazu sein strategisches Wissen einzusetzen, indem er auf ähnliche Strukturen in dem Text zurückgreift oder erlernte Strukturen wiederruft. Die Antwort auf eine Aufgabe wie der K 2 Vergangenheit erfordert die Einsetzung der Lernerkenntnisse des Konjunktivs 2 aber auch seine Kenntnisse des Perfekts. Das Lösen dieser Aufgabe erfolgt grundlegend auf Erinnerungsprozesse und weist einen niedrigen Interaktivitätsgrad auf. Es hat sich anhand der Analyse herausgestellt, dass der Grammatiktestteil an Authentizität mangelt, da die kommunikative Sprachkonzeption davon weit ist, die Sprache in einen Erinnerungsprozess zu reduzieren. Der Grammatiktestteil ist sprachanalytisch orientiert und besagt somit wenig über die Lernerkompetenz in der Zielsprache zu kommunizieren. Der

Dritte Subtest will die Schreibkompetenz der Lerner, einen Paragraphen zu schreiben, beurteilen. Hierzu setzt der Lerner komplexe mentale Prozesse wie das Langzeitgedächtnis ein.<sup>149</sup>

Das Wesen des dritten Teils als offene Aufgabe impliziert die hohe Einstufung der personellen Eigenschaften, da das Schreiben ein weitgehend komplexer Vorgang ist. Über die Einsetzung des Langzeitgedächtnisses hinaus, gliedert sich der Prozess des Schreibens in drei Phasen oder Prozesse: Das Planen, das Formulieren, und das Überarbeiten.

Die Planungsphase schließt das Generieren, die Zielsetzung, das Strukturieren ein. Der Lerner greift auf sein allgemeines Weltwissen sowie sein Fachwissen zurück. Die strategische Kompetenz könnte auch eingesetzt werden wie die Entnahme von Strukturen aus dem Text und in dem Revidierungsprozess. Das Schreiben des Aufsatzes ist eine integrierte Aufgabe, in der verschiedene Prozesse durchgeführt werden. Der Lerner soll über eine Thematik oder über eine problematische Situation schreiben. Nennenswert sind Themen wie: die Werbung und die Massenmedien oder Jugend und Drogen. Ohne den gezielten Einsatz der entsprechenden Kompetenzen, kommt die erwartete Leistung nicht zustande. Somit realisiert dieser Subtest einen hohen Interaktivitätsgrad. Nach der Benotung ermittelt das Ergebnis ein reliables Bild der Lernerschreibkompetenz.

Die Analyse hat ergeben, dass der dritte Subtest einen hohen Interaktivitätsgrad aufweist, und dies im Licht des 19. Kriteriums, das sich auf die Angemessenheit der Aufgabenzüge mit den persönlichen Eigenschaften der Lernergruppe bezieht. Die Entsprechung der Testaufgabencharakteristika mit den Lernereigenschaften sollte im Testdesign berücksichtigt werden. Die Testaufgabenzüge können den Lernereigenschaften entsprechen durch die genaue Analyse der Lernergruppe hinsichtlich Alter, muttersprachliche Hintergründe und schulische Laufbahn.

Die Zielgruppe stellt junge Lerner von 17- 19 Jahre alt, die Deutsch ab dem 2- Studienjahr gelernt haben. Die Lerner verfügen über einheitliche sprachliche Erfahrungen. Die Muttersprache ist in den meisten Fällen Arabisch oder Berberisch. Die Orientierung der Testaufgabencharakteristika an den Lernereigenschaften trägt dazu bei, den Interaktivitätsgrad zu erhöhen und somit die Konstruktvalidität des Tests zu verstärken. Testaufgaben, die mit den Lernereigenschaften übereinstimmen, erwecken die Motivation der Lerner.

---

<sup>149</sup> HAYES, J/ FLOWER, L. (1980) : Identifying the Organisation of Writig Processes. In: GREGG, W. / STEINBERG, R, Cognitive processes in Writing . Hillsdale: Lawrence Erlbaum Associates, 1980, S: 3-4-5.

Wissenschaftliche oder allzu technische Themen würden die Lerner der Fremdsprachen nicht interessieren.

Der Themenkatalog des 3. Studienjahres umfasst folgende Themen:

- Die deutschsprachigen Länder und Algerien heute.
- Künstlerleben.
- Der technische Fortschritt.
- Die Umweltprobleme .
- Die Massenmedien.
- Die Kommunikationsmittel.
- Die Jugendprobleme
- Die Arbeitswelt.

Es wurde bei der Analyse der Prüfungsanleitung festgestellt, dass sie keine Hinweise über die Themenwahl eingibt. Die Prüfungsanleitung weist nicht drauf hin, dass schockierende Themen vermieden werden müssen. Die ausgewählten Themen sind potenzielle Motivationsvehikel. Wir haben auch festgestellt, dass die Themen, die die deutschen Jungen vorziehen, algerische Lerner nicht zwangsläufig interessieren. Es wurde festgestellt, dass die Lerner Themen wie Jugendprobleme, und Kommunikationsmitteln höchstmotivierend finden, während sie für Umweltprobleme kaum Interesse zeigen. Für meine empirische Untersuchung des Testimpacts wird auf die von YETTOU( 2001) durchgeführte Befragung algerischer Lehrer gestützt. Die Lehrer wurden u.a. über die Themen, die das Interesse ihrer Schüler am meisten erwecken gefragt. Die Ergebnisse der Befragung sind folgende Quoten: Das Interesse der Schüler an bestimmten Themen des Deutschunterrichts wie Jugendprobleme ist groß.

60% Jugendprobleme. Umweltprobleme 10% . YETTOU weist auch darauf hin, dass das Interesse an Kunst und Kulturleben bei allen Schülern sehr gering ist. Zusammenfassend kann gesagt werden, dass die Schüler sich für Themen interessieren, die sie unmittelbar ansprechen.<sup>150</sup>

---

<sup>150</sup> YETTOU, B ( 2001) Zur Entwicklung eines Lehrerhandbuchs für algerische Deutschlehrer, Diss., Universität Es- Sania, Oran.

Von der Tatsache ausgegangen, dass die jungen Lerner für neue Themen sich interessieren, Veränderungen erleben, die modernsten Kommunikationsmitteln gern benutzen, und von den Medien stark beeinflusst sind, erweisen sich Themen wie Handy und Internet als motivierend. Der Themenkatalog „Die deutschsprachigen Länder und Algerien heute“ entspricht dem Prinzip der Interkulturalität und erfüllt den Hang der jungen Lerner über andere Kulturen zu wissen. Dazu stellt die deutsche Sprache das Ziel der fremdsprachlichen Ausbildung. Bezüglich der Themenauswahl entsprechen die ausgewählten Themen den Lernereigenschaften als Jugendliche des 21. Jahrhunderts. Dies erhöht den Interaktivitätsgrad der Prüfung.

Bezüglich der Testform sind die Aufgaben kurz und im Ganzen vom Einfachen zum Schweren graduiert, wenn die dritte Aufgabe im Subtest „Leseverständnis“ nicht in Betracht genommen wird. Es wurde festgestellt, dass diese Aufgabe einen hohen Schwierigkeitsgrad aufweist; Ihre Position in dem ersten Subtest kann als motivationshemmend betrachtet werden. Was die Eigenschaften der Lerner nicht respektiert, da sie als Junge Menschen dazu neigen, schnell demotiviert zu werden.

In dem dritten Subtest „Schreibfähigkeit“ schreiben die Lerner einen Paragraphen. Diese Aufgabe berücksichtigt die Lernercharakteristika in Verbindung mit den ausgewählten Themen, die oben besprochen wurden und in Bezug auf die Antwortform bzw. die aufgeforderte Reaktion.

Die Charakteristika, die potenziell die Lernerperformanz in dem Test beeinflussen könnten sind sehr umfangreich. Es werden Alter, Geschlecht, Muttersprache, Niveau und Typ der Allgemeinbildung unterschieden. Diese Eigenschaften sollten bei der Testerstellung berücksichtigt werden, da die Testersteller in enger Vertrautheit mit dem Lerner aufgrund des täglichen Kontakts in der Klasse stehen.

In der Prüfungsanleitung wurde festgestellt, dass die Lehrer unklare Wörter aus dem Text auf Arabisch oder Französisch erklären sollten. Es ist hier erwähnenswert, dass Arabisch in vielen Fällen nicht die einzige Muttersprache sein könnte. Die Prüfung berücksichtigt die Zielsetzungen des Curriculums, die die Offenheit auf der Welt betonen und berücksichtigt damit die Charakteristika der Lerner. In der Prüfungsanleitung steht fest, dass die Lehrkraft Themen wählen muss, die in der Klasse behandelt wurden. Es wurde darauf nicht hingewiesen, dass die allzu aktuellen Themen, sowie schockierende Themen, wie neue Naturkatastrophen vermieden werden müssen. Die Prüfungsanleitung weist darauf hin, dass die Prüfung von mittlerem Schwierigkeitsgrad sein muss, bzw. von dem mittleren Lerner gelöst werden kann.

Bezüglich des 21. Kriteriums, ob die aufgeforderte Reaktion in der Testaufgabe umfangreiche Bereiche der Lernerkompetenzen einruft, wurde Folgendes festgestellt: In der ersten Aufgabe setzt der Lerner sowohl sein deklaratives als auch sein prozedurales Sprachwissen ein. Die zweite Aufgabe benötigt die Einsetzung des Lernalphabetischen Regelwissens sowie sein landeskundliches Wissen. Was das prozedurale Wissen anbelangt, kommen Sprachrezeptionsverfahren zum Einsatz. Sprachproduktionsverfahren kommen auch hierbei zustande. Der Lerner versucht in der Leseverständnisaufgaben grammtisch und inhaltlich richtige Sätze zu produzieren, hier setzen die Lerner sowohl deklaratives Wissen als auch sprachliches Regelwissen bzw. die Kenntnis des Wortschatzes und der Assoziogramme ein. Bei der prozeduralen Bearbeitung des Inputs kommt das Leseverstehen ohne den Einsatz des deklarativen- und prozeduralen Wissens nicht zustande.

Im zweiten Prüfungsteil ist der Lerner aufgefordert, sein sprachliches Regelwissen bzw. die im Rahmen des Programms erlernten grammatischen Strukturen zu beweisen. In diesem Fall ist der Lerner nicht dazu aufgefordert sein prozedurales Wissen einzusetzen, da die Sätze schon in Form einer Ergänzungsaufgabe vorgegeben werden, und deshalb weniger interaktiv sind, wie im folgenden Beispiel ersichtlich wird:

Unsere Eltern waren jung. Sie machten Fehler.

Der Lerner ist aufgefordert die vorgegebenen Sätze mit einer Konjunktion zu verbinden.

Darüber hinaus sind diese Aufgaben für die Lerner schwer zu lösen, wegen Interferenzen aus dem Französischen oder Arabischen, was die Motivation hemmt.

Die Lernerkopienanalyse hat diesbezüglich ergeben, dass nur 33, 33% der Lerner in diesem Subtest gut abschneiden, bzw. den Durchschnitt haben. Dieser Subtest scheitert daran, einen großen Teil der Lerner dazu zu bringen, in der Aufgabe zu leisten, was den Interaktivitätsgrad abnimmt. Vielmehr schreiten viele Lerner die Aufgaben zur Grammatik über, ohne auf die Fragen zu antworten. Im dritten Teil der Prüfung wird der Lerner veranlasst, einen Paragraphen zu einem bestimmten Thema zu schreiben. Wie vorab erläutert wurde, misst diese Aufgabe die Schreibkompetenz der Lerner. Das Aufgabenumfeld umfasst acht Themen, die die Lerner in der Klasse im Laufe des Schuljahres behandelt haben. Die Form der Aufgabe als offene Aufgabe veranlasst zur Einsetzung verschiedener Kompetenzen und mentaler Wissensstrukturen. Spezielle Aspekte bei der Analyse dieser Aufgabe ergeben sich aus der offenen Form des Inputs und der erwarteten Reaktion. Das Schreiben ist ein langsamer Prozess, das den Einsatz von Wortschatzkenntnissen der Lerner sowie sein sprachliches Regelwissen impliziert. Prozedurale Sprachaktivitäten kommen auch zum

Einsatz wie das Leseverstehen beim Dekodieren des Inputs bzw. der Festlegung des Aufgabenumfelds.

Strategisches Wissen wird eingesetzt bei dem Nachweis der erlernten Strukturen. Der Lerner ist davon bewusst, dass er dem Prüfer nachweisen soll, dass er den erlernten Stoff assimiliert hat. Die Lehrkraft muss dem Schüler während des Unterrichts erklären, dass er beim Schreiben, über die klare und richtige Schreibform hinaus, auch die erlernten Strukturen einsetzen muss. Deshalb benutzt der Lerner die verschiedenen grammatischen Strukturen u.a. die mehrteiligen Konjunktionen, das Passiv, der Konjunktiv 2. usw. Der Einsatz des strategischen Wissens findet besonders bei der Revidierungsphase statt.

Strategisches Wissen wird auch eingesetzt, wenn der Lerner seine Ideen richtig organisiert, bzw. den Aufsatz in einer Einleitung, einem Körper und Schlussfolgerung einteilt.

Die Analyse macht deutlich, dass der dritte Subtest der Prüfung das Kriterium weitgehend erfüllt und somit einen hohen Interaktivitätsgrad realisiert.

Bezüglich des 23. Kriteriums, das die Interdependenz der Testaufgaben betrifft, hat es sich anhand der Analyse herausgestellt, dass die Testaufgaben in Bezug auf die Thematik interdependent sein können. Die Thematik des Inputs und der Testaufgaben, die Auswahl der Distraktoren und die Thematik des dritten Subtests „ Schreibkompetenz“ muss einheitlich konzipiert werden.

Die erste Aufgabe „ Richtig / Falsch“ ist mit der zweiten Aufgabe des Subtests „Textverständnis“ interdependent, d. h. der Lerner, der richtig in der ersten Aufgabe bei der Ankreuzung und auch bei der Begründung antwortet, auch die Fragen zum Text richtig lösen kann. Hierzu riskiert der Prüfer das Textverständnis wiederholt zu testen. Die Interdependenz der zwei Aufgaben drückt auch eine freundliche Haltung des Testdesigners aus. Der schwachleistende Lerner kann in den beiden Aufgaben gut abschneiden. Darüber hinaus ist in der Prüfungsanleitung explizit ausgedrückt, dass die Lehrkraft bei der Erstellung der Testteile der Textthematik entsprechende Themen wählen muss.

Die Interdependenz der Testaufgaben kann auch in der Beurteilungsmethode vorkommen. In den siebziger und achtziger Jahren verwendeten die amerikanischen Testdesigner in vielen Bereichen eine Beurteilungsmethode, die danach unter der Bezeichnung „ amerikanische Methode „ bekannt wurde. Die Messverfahren beruhen auf einer Interdependenz der Testaufgaben, so dass die richtige Antwort auf eine Aufgabe von der Ersteren abhängt. Somit verliert der Lerner bei jeder falschen Antwort Punkte nicht nur in der entsprechenden Aufgabe, sondern auch in der Nachkommenden.



In der Hauptschulprüfung des algerischen Daf-Unterrichts kommt eine solche scharfe Interdependenz nicht vor. Es kann zwar gesagt werden, dass die Interdependenz der Testaufgaben im Interesse der Lerner ist.

In den ersten und zweiten Aufgaben sind die Prüfungsziele identisch: die Überprüfung des Leseverstehens. Hierbei setzt der Lerner metakognitive Strategien wie die Projektion oder die Schattierung der Informationen der ersten Aufgabe auf die Zweite. Wenn die erste Aufgabe „eine Richtig-Falsch Aufgabe“ richtig gelöst wird, kann der Lerner prinzipiell auch die zweite Aufgabe „Fragen zum Text“ richtig lösen. Ist die erste Aufgabe gezielt und den Gütekriterien angemessen erstellt, wird der Lerner dabei Hypothesen über den Text bilden um sein Leseverständnis zu optimieren. In der Aufgabe „Fragen zum Text“ schneidet nur ein kleiner Anteil der Lerner ab in der Teilaufgabe: „erklären Sie diesen Satz“.

In der ersten Aufgabe „Richtig/Falsch“ haben nur 35,71% der Lerner den vollen Punktwert erhalten trotz der auffallenden Einfachheit dieser Aufgabe.

Die Testaufgaben allgemein erweisen sich zu einem kleinen Grad interdependent. In diesem Fall soll der Lerner jede Aufgabe unabhängig von der anderen lösen. Der Lerner kann auch einige Aufgaben unabhängig vom Text lösen, wie bei der Kombiniertenaufgabe des Textverstehens und den Aufgaben im Subtest „Grammatik“. Dabei sind vielmehr psychologische Strategien einzusetzen, damit eine Fehlleistung in einer Aufgabe nicht die Performanz in einer Anderen beeinflusst. Der Grammatik-Sub-Test veranlasst zur richtigen Lösung der grammatischen Strukturen. Es wurde in der Prüfungsanleitung nicht darauf hingewiesen, dass der Lehrer bei der Auswahl der betroffenen grammatischen Strukturen diejenigen vermeiden sollte, die analoge Strukturen aufweisen, oder die Eingabe eines Satzes im Passiv, der das Perfekt enthält und den Lerner nachher auch im Perfekt zu überprüfen oder einen anderen Satz zu geben, dessen Lösung die Einsetzung des Perfekts erfordert, wie das Konjunktiv 2 Vergangenheit. Im Fall eines solchen Umgehens mit der Strukturenwahl wird der Lerner benachteiligt.

Was die dritte Aufgabe anbetrifft und im Licht des 24. Kriteriums, ob die Testaufgaben dem Lerner Anstöße zum Einsatz seiner strategischen Kompetenzen bietet, wird Folgendes festgestellt:

-Die Zeitanweisung zum Einsatz des strategischen Wissens ist weitgehend genügend.

-Die Lerner sind aufgefordert, ihre sprachlichen Kompetenzen sowie ihr thematisches Wissen einzusetzen. Zielstellungen werden besonders in dem dritten Teil bei der Themawahl eingesetzt. Der Lerner soll das angemessene Thema für seine sprachlichen Kompetenzen und thematischen Kenntnisse auswählen.

Die Analyse hat ergeben, dass Anstöße zur Einsetzung des strategischen Wissens zwar dargeboten sind, aber weitgehend auf dem dritten Teil der Prüfung beschränkt.

Bezüglich des 25. Kriteriums ob die Testaufgaben affektive Reaktionen der Lerner hervorrufen, wird festgestellt, dass die Prüfungsanleitung keine Anweisungen diesbezüglich angibt. Die Lehrkraft muss ein der acht behandelten Themen wählen. Es wird nicht erwähnt, dass sehr emotionale Themen vermieden werden müssen. In diesem Fall könnte die Lehrkraft ein sehr aktuelles Thema über Umweltkatastrophen, wie Erdbeben oder Tsunami, oder Weltende wegen einer Umweltkatastrophe darbieten. Einige Lerner können die emotional erschöpften Themen nicht ertragen. Dasselbe gilt für das Thema „Technischer Fortschritt“, wo verbreitete Krankheiten zitiert werden können; was einigen Lernern betreffen könnte, wenn ein Familienmitglied davon neulich gestorben ist.

Was die Arbeitsanweisungen anbelangt, könnte der hohe Schwierigkeitsgrad der Arbeitsanweisungen dazu führen, die Motivation der Lerner zu hemmen, da sie die Aufgabe nicht verstehen und folglich wird es keine Interaktion statt finden.

Bezüglich der Graduierung vom Einfachen zum Schweren, wurde festgestellt, dass die erste Aufgabe von dem didaktischen Prinzip abweicht, da die vierte Frage einen hohen Schwierigkeitsgrad aufweist<sup>151</sup> und wurde trotzdem am Anfang gestellt.

Bezüglich des zweiten Testteils, wird festgestellt, dass obwohl die Prüfungsanleitung Hinweise darauf macht, dass die Aufgaben vom Einfachen zum Schweren graduieren müssen, wurde übersehen, dass auch die Teilaufgaben in den einzelnen Testaufgaben das Kriterium der Graduierung berücksichtigen müssen.

Die logische Analyse der Interaktivität hat ergeben, dass die erste Aufgabe „Fragen zum Test“ einen hohen Interaktivitätsgrad mit den Lernern realisiert, weil der Lerner beim Lösen dieser Aufgabe seine Kenntnis der sprachlichen Elemente und Assoziogramme, die für das Leseverstehen notwendig sind einruft. Die Analyse hat auch hinsichtlich des Zusammenhangs zwischen Validität und Interaktivität ergeben, dass valide Aufgaben höhere Ansprüche an die Lernereigenschaften erheben.

Bezüglich der vierten Aufgabe des Subtests Textverständnis weist die Frage: „Erklären Sie den folgenden Satz!“ einen hohen Schwierigkeitsgrad auf; nur 2,5% antworten richtig auf die Frage.

---

<sup>151</sup> Siehe: Lernerkopienanalyse der vorliegenden Arbeit. S: 94

7,5 der Lerner antworten auf die Frage nicht.

90% der Lerner geben falsche Antworten an.

. Die Analyse des Testteils „Sprachkompetenz“ hat ergeben, dass die vorgegebenen Aufgaben, obwohl sie den curricularen Prinzipien widersprechen, Ansprüche an die Lerner erfordern, da die Lerner auch passives Wissen, wie das Wiederrufen sprachlicher Strukturen und Regeln einsetzen müssen.

**Tabelle 4** Ergebnisse der Testanalyse im Hinblick auf die Interaktivitätskriterien.

	Weitgehend erfüllt	Erfüllt	Nicht erfüllt
Inwieweit ruft der Test der angemessene Bereich des thematischen Wissens ein.	×		
Die Charakteristika der Lerner(Probanden) sind im Designstatement mitberücksichtigt.		×	
Die Wesenszüge der Testaufgaben sind den Lernern angemessen in Bezug auf die im Design-Statement spezifizierten Charakteristika.	×		
Die aufgeförderten Verarbeitungen in den Testaufgaben setzen umfassendere Bereiche der sprachlichen Lernerkompetenz voraus.	×		
Inwieweit sind die Testaufgaben interdependent.	×		
Der Test gibt dem Lerner Gelegenheit sein strategisches Wissen einzusetzen.		×	

*Quelle: Eigene Darstellung*

Impact bezieht sich auf den Einfluss, den eine Prüfung auf eine Gesellschaft hat bzw. ihr Einfluss im Zusammenhang mit dem deduktiven System. Die Prüfung kann eine Gesellschaft betreffen auf zwei Ebenen:

- Auf der Micro- Ebene.
- Auf der Macro-Ebene.

Die Analyse wird zunächst auf die Micro-Ebene fokussieren bzw. auf den Lerner, Objekt des Testverfahrens.

## **A-Einfluss der Prüfung auf die Lerner**

Die Kriterien anhand deren die Analyse durchgeführt werden wird, befassen sich mit den drei Aspekten des Testverfahrens bzw. der Durchführung des Tests an sich (the experience of taking the test).<sup>152</sup>

Die Lerner sind zuerst von dem Test beeinflusst, bevor sie den Test machen, bei der Vorbereitung für den Test, und bei der Erstellung des Tests, wenn die personellen Eigenschaften der Lerner berücksichtigt werden.

Prüfungen können auf die Lerner einen Einfluss haben und dies durch die Verbesserung der Kontrollen und durch das erhaltene Feedback. Die Entscheidungen, die aufgrund der Testergebnisse getroffen werden, könnten die Lerner in manchen Fällen beeinflussen und über ihre berufliche oder schulische Laufbahn entscheiden.

Bezüglich des 26. Kriteriums: inwieweit beeinflusst die Erfahrung des Tests die Eigenschaften der Lerner, die mit dem jeweiligen Gebrauch der Sprache betroffen sind, wurde festgestellt, dass die Beschränkung auf die im Programm behandelten Themen in der Prüfungsanleitung fest verankert wurde. In diesem Fall kann der Lerner allein durch das Lesen des Textes neue Informationen erwerben oder vorheriges Wissen versichern. Im Curriculum steht fest, dass der Lerner einen offenen Standpunkt zu den anderen Kulturen entwickeln sollte, was in den Prüfungen nicht vernachlässigt werden muss.

Es wird bei der Analyse der Prüfungsanleitung festgestellt, dass die Aufmerksamkeit der Lehrer darauf nicht gezogen wurde. Es werden auch keine Hinweise über die Qualität der Sprache bzw. den Einsatz von Texten, die negative Werte enthalten, wie z.B. Texte, die rassistische Ideen und/oder Vorurteile reflektieren bzw. vertreten. Es wird festgestellt, dass die Prüfungsanleitung die Notwendigkeit, dass die Prüfung fehlerfrei erstellt werden muss, betont. Als Beispiel wird ein Abschnitt aus der Prüfungsanleitung zitiert: „Pour le choix de tout support il faudra veiller à ce que la langue et le style soient aisément accessibles et parfaitement corrects. »<sup>153</sup>

Der Text muss sowohl auf der Ebene der Struktur als auch auf der Ebene des Wortschatzes reichhaltig sein. Der Reichtum des Inputs wird dem Lehrer bei der Erstellung der Prüfung helfen, den entsprechenden Input aus dem Text zu entnehmen um die Testaufgaben zu erstellen. Der Reichtum des Inputs trägt dazu bei, die Prüfung authentischer und folglich valider zu erstellen. Bezüglich des Tests kann der Lerner bei der Einsetzung seiner funktionalen sprachlichen Aktivitäten bzw. des Leseverstehens neue Informationen

---

<sup>152</sup> BACHMAN, L( 1990,S. 146)

<sup>153</sup> Siehe. Prüfungsanleitung ( 2010, 08)

entnehmen. Die Synonymfindungsaufgabe wird konzipiert, dass der Lerner die Synonyme leicht heraus schließen kann, was pädagogisch positiv zu bewerten ist. Die Aufgabenform hilft dem Lerner eine bessere Performanz zu erbringen, seitdem das unterstrichene Wort in einem Satz, bzw. in seinem Kontext eingebettet wird, wie in der folgenden Aufgabe der Fall ist:

**Ersetzen Sie das unterstrichene Wort durch sein Synonym im Text!**

a-Die Experten warnen vor übermäßigem TV- Konsum.

b-Das Ziel der Werbung ist meistens ökonomisch.

Im dritten Prüfungsteilteil kann die Erfahrung des Tests den Sprachgebrauch der Lerner beeinflussen, da der Lerner einen Aufsatz schreiben sollte. Was das Feedback anbetrifft, wird es im algerischen Prüfungssystem fest verankert. 100% der befragten Lehrer haben angegeben, dass sie vor der Noteneingabe eine Verbesserungsstunde veranstalten. Der Impact des Feedbacks kann die Lernereinstellungen zum Gebrauch der Sprache beeinflussen, wenn die Testinhalte den behandelten Lerninhalten in der Klasse entsprechen. In diesem Fall wird der Sprachgebrauch des Lerners durch zwei Prozesse beeinflusst:

- Die Verstärkung des Sprachwissens durch psychologische Prozesse.
- Die Verbesserung der Fehler und das Erlernen neuer Strukturen.

Im Zusammenhang mit dem Impact der Prüfung durch das Feedback, enthält Die Prüfungsanleitung ein Verbesserungsblatt, das die Richtigkeitskriterien der Antworten festlegt. Darüber hinaus verfügt die Prüfungsanleitung auch über eine Punkteverteilung.

Die Analyse hat ergeben, dass die Erfahrung des Tests durch die Eingabe der Richtigkeitskriterien und die Punkteverteilung, sowie durch die im Curriculum verankerten Prinzipien des Feedbacks prinzipiell die Lerner positiv beeinflusst. Somit wird das Kriterium, das den positiven Einfluss der Testdurchführung auf den Lernaltersprachgebrauch betrifft, mit Vorbehalt erfüllt. Denn das Feedback erfüllt seine Funktion, nur wenn das Testkonstrukt mit den Lerninhalten übereinstimmt. Darüber hinaus, wird festgestellt, dass das Feedback lediglich durch eine gezielte Aktion die Lerner dazu verhilft die unerfüllten Kriterien und die Prüfungslücken zu verbessern, wie etwa bei der 4. Frage des Leseverständnisses, wo die Lerner solche Aufgaben während des Kurses nicht genug behandelt haben, könnte die Verbesserungsstunde für den Lehrer eine Gelegenheit den Lernern die Richtigkeitskriterien der Aufgabe zu erklären und auch die Arbeitsanweisungen, weil die Lerner der Prüfung mehrmals ausgesetzt werden.

Auf der Macro- Ebene ist die Hauptschulprüfung im algerischen gymnasialen Unterricht von großer Bedeutung. Die Prüfung wird als Vorbereitungstest für das Abitur in allen Daf-Klassen in algerischem Gymnasium eingesetzt.

Bezüglich des 27. Kriteriums, ob die Lerner bei der Erstellung der Hauptschulprüfung gefragt wurden, welche Aufgabentypen sie für angemessen halten, und in denen sie bessere Leistungen haben, stehen in der Prüfungsanleitung, sowie im Curriculum keine Hinweise darauf, dass die Lerner bei der Erstellung dieser Prüfung mitbeteiligt werden sollen. Auch keine Anweisungen darauf, dass die Lehrkraft in den Lernkontrollen die Lerner fragt, welchen Aufgabentyp sie vorziehen, sind vorgefunden. Obwohl es bekannt ist, dass die Aufgabenvielfalt die Performanz der Lerner verbessert. Darüber hinaus müssen die Lehrer wohl oder übel dem vom Ministerium vorgeschlagenen Modell Folge leisten. Es ist in der Prüfungsanleitung auch darauf nicht hingewiesen worden, ob die Testdesigner bei der Erstellung dieser Prüfung sich auf akademische Forschungen gestützt haben. Was das 28. Kriterium anbetrifft, ob das Feedback relevant, vollständig, und bedeutungsvoll für die Lerner ist, wurde festgestellt, dass wegen des Defizits in der Prüfungsanleitung an einer detaillierten Punkteskala, wird dem Lerner kein „qualitatives“ Feedback dargeboten. Obwohl die Lehrer vor der Noteneingabe eine Verbesserungsstunde organisieren, können die erbrachten Noten den eigentlichen Kompetenzstand der Lerner in Bezug auf die Zielstellungen im Curriculum nicht genau reflektieren.

Es wird lediglich ein Verbesserungsmodell dargeboten, sowie eine Punkteverteilung, d. h. Punktwerte werden der Lernerleistung zugewiesen, die an sich relevant und bedeutungsvoll sind. Die Note 10/20 bedeutet, dass der Lerner mittelmäßig ist.

Diese Beurteilung ist zu bezweifeln, wenn der Schwierigkeitsgrad der Prüfung hoch ist.

Die Analyse der Prüfungsanleitung hat ergeben, dass sie keine Hinweise auf qualitatives Feedback enthält. Es gibt keine Punkteskala, die die Punktwerte interpretiert und beurteilt.

Im Nachfolgenden wird auf die Notenverarbeitung im algerischen DaF-Prüfungssystem an algerischem Gymnasium eingegangen. Die Note in der Hauptschulprüfung wird in drei vervielfältigt, die Note der dauernden Evaluation, franz.: (Evaluation continue) wird aus der erbrachten Note in den Lernaktivitäten in der Klasse bzw. das Schülerverhalten während des Unterrichts und die Note der Klassenkontrollen gerechnet, das Ganze wird in 2 vervielfältigt. Die gesamte Lerner- note d, h.. der Durchschnitt im Fach Deutsch wird von der

Prüfungsnote  $\times 3$  + und der Note der dauernden Evaluation in 2 vervielfältigt gerechnet. Das Ergebnis wird über fünf geteilt.

Nach diesen Verfahren wird der Durchschnitt im Fach Deutsch gerechnet. Es soll darauf hingewiesen, dass das Ergebnis in 4 vervielfältigt wird. Je höher der Koeffizient ist, desto entschiedener wirkt das Ergebnis im entsprechenden Fach bei der Errechnung des allgemeinen Durchschnitts mit. Die Bedeutung des Faches hängt von dem Wert des Koeffizienten ab; je größer der Beurteilungsfaktor ist, desto höher wird der Mitwirkungsgrad. Bezüglich des 20. Kriteriums, das sich auf die Eindeutigkeit der Selektionsprozeduren für alle Lerner bezieht, wurde die Entscheidungsfunktion im Blue-Print nicht erwähnt. Die Beurteilungsprozeduren werden zwischen Lehrern übertragen und/ oder von der Verwaltung bekommen und stellen für viele Lehrer insbesondere diejenigen, die den Lehrerberuf zum ersten Mal ergreifen, Schwierigkeiten dar.

Bezüglich der 30. Frage zur Evaluation des Impacts „inwiefern ist das Testergebnis relevant für die Entscheidungen, die getroffen werden“, hat sich herausgestellt, dass das Ergebnis der Prüfung in 3 vervielfältigt, während das Ergebnis der Kontrolle und der dauernden Evaluation „evaluation continue „durch 2, das Ergebnis aus der Kontrollenote und der dauernden Evaluation wird in 2 vervielfältigt, das Ganze wird über 5 geteilt.

Der Durchschnitt =  $\text{Hauptschulprüfung} \times 3 + \frac{(\text{ständige Evaluation} + \text{Kontrolle1} + \text{Kontrolle2})}{5} \times 2$

5

Das\_Ergebnis wird durch 5 geteilt.

Der Mitwirkungsgrad der Prüfung in den Entscheidungen über den Lerner ist hoch. Das 31. Kriterium ist erfüllt, seitdem die Prozeduren für das Errechnen des Testergebnisses im Fach Deutsch dieselben sind wie in den anderen Sprachen, die im algerischen FSU unterrichtet werden bzw. Englisch, Französisch und Arabisch. Was die Ausnahme stellt, ist der Koeffizient des Faches Deutsch, der nur 4 Stufen stellt, während er in den anderen Sprachen bzw. im Französischen, Arabischen und Englischen bis 5 Stufen steigt. Wenn der Koeffizient im Fach Deutsch im Vergleich mit den anderen Fächern niedrig ist, entscheidet das Fach relativ schwach in der Errechnung des Lerner gesamt Durchschnitts mit. Fraglich ist ob dieser relativ niedrige Koeffizient der Tatsache, dass die Lerner Deutsch nur für 2 Schuljahre gelernt haben oder anderen bzw. kulturell- historischen und/oder politischen Ursachen zurückzuführen ist.

Fraglich ist auch, ob die Wiedereinschätzung des Koeffizienten im Fach Deutsch für das Fach und über die Lerner positiv auswirken wird.

## **B- Impact auf die Lehrer**

Das 33. Kriterium bezieht sich auf den Grad zu welchem die sprachlichen Kompetenzen, die der Test zu messen vorgibt mit denen im Lehrmaterial einbezogenen Kompetenzen übereinstimmen. Die Prüfung erfüllt dieses Kriterium, indem die Lehrer die Testinhalte ausschließlich aus dem im Curriculum bzw. im Lehrbuch festgelegten Inhalte entnehmen müssen. (Guide d'examineur, 2010)

Was das 34. Kriterium anbelangt, das die Übereinstimmung der Testaufgabencharakteristika mit der Natur der Lehr- und Lernaktivitäten betrifft, hat Folgendes herausgestellt: Um dieses Kriterium zu analysieren sollte zum Lehrwerk zurückgegriffen werden. Es wird im Lehrbuch die Lektion 3 analysiert, um zu überprüfen, ob die Aufgabenmerkmale bzw. die Arbeitsanweisungen, Testrubrik, Input und die erwartete Reaktion mit den Lehr- und Lernaktivitäten übereinstimmen. Diesbezüglich wird Folgendes erhoben:

### **Die Arbeitsanweisungen**

Was die erste Aufgabe des Leseverständnisses anbetrifft, kommen im Lehrwerk die Arbeitsanweisungen wie folgt:

**-Stimmt das oder stimmt das nicht?**

**-Kreuzen Sie an!**<sup>154</sup>

Diese Arbeitsanweisungen der Prüfungsaufgaben weisen in Bezug auf das Lehrbuch eine Kongruenz hinsichtlich der Form auf.

Es wird festgestellt, dass der Aufgabeninhalt zum Üben des Leseverstehens im algerischen geltenden Lehrwerk zum Üben des Leseverstehens und zur Überprüfung des Leseverstehens mit den Arbeitsanweisungen kongruiert, indem die Testersteller eine „richtig-falsch“ Aufgabe eingesetzt haben.

Bezüglich der Fragen zum Text wird dieselbe Form eingesetzt, aber meist mit der Arbeitsanweisung **“Antworten Sie!”**. In der Prüfung wird die Teilaufgabe **„Fragen zum Text“** mit der Arbeitsanweisung **„Fragen zum Text“** eingeführt.

Es könnte zu Missverständnissen führen, sowie zur Hemmung der Motivation oder zu negativen Prüfungsverhalten, wie das Überschreiten der Arbeitsanweisungen. In diesem Fall antworten viele Lerner lediglich nach der Form der Aufgabe, und dies könnte zu Fehlern führen.

---

<sup>154</sup> El Ahcen, et al (2010, 141) :Vorwärts mit Deutsch , 3 AS, Erziehungsministerium: Algier



Dies ist empirisch klar geworden, da einer der befragten Lerner in dem Fragebogen geschrieben hat, dass er die Arbeitsanweisungen nicht liest, sondern direkt antwortet, je nach der Aufgabenform ohne die Arbeitsanweisung zu lesen. Darüber hinaus, drückt die Arbeitsanweisung: „**Fragen zum Text**“ aus, dass der Lerner, der während des Unterrichts im Lehrwerk lediglich antworten muss, in der Prüfung die Antwort in dem Text herausfinden soll.

Bezüglich der Wortschatzarbeit im Unterricht werden genau die gleichen Formen vorgegeben, wie sie in der dritten Lektion vorkommen: „**Ersetzen Sie die hier unterstrichenen Wörter durch ein Synonym aus dem Text oder ein gleichbedeutendes Wort aus dem Text.**“<sup>155</sup>

Hinsichtlich der Wortbildungsarbeit im Unterricht, wird meist zur Bildung des Kompositums das Suffix oder Präfix eingegeben. Hingegen werden sie in der Prüfung nicht eingegeben, genau wie in Kombinerungsaufgaben, in denen die Arbeitsanweisungen im Lehrwerk wie folgt vorkommen: „kombinieren Sie! Oder was passt zusammen? (Vorwärts mit Deutsch, 94) Dasselbe gilt für die grammatischen Strukturen wie „Dass-Sätze“, die in der Lektion 3; Seite.60 vorkommen.

Die Sätze kommen wie folgt vor:

„Wussten Sie schon, dass.....?“

Der Lerner soll dabei die fehlende Struktur ergänzen. Hier handelt es um einen „Dass-Satz“, während in der Prüfung dem Lerner zwei Sätze vorliegen. Er muss die Sätze verbinden.

Das Üben und Testen der Modalverben und die eingesetzten Arbeitsanweisungen im Lehrwerk und in der Hauptschulprüfung stimmen völlig überein. Die Arbeitsanweisungen zum Üben der Modalverben kommen im Lehrbuch wie folgt vor: „**Ersetzen Sie die unterstrichenen Wörter durch das entsprechende Modalverb**“. In der Hauptschulprüfung weisen die Arbeitsanweisungen in Bezug auf das Üben der Modalverben im Lehrwerk eine völlige Kongruenz auf. Die Arbeitsanweisung ist für den Lerner, der für ein Jahr Deutsch gelernt hat schwer zu verstehen. Die Lehrer in der, im Rahmen dieser Forschung, durchgeführten Befragung haben angegeben, dass sie die Arbeitsanweisungen auf einer anderen Sprache als Deutsch erstellen würden.

Was die Übersetzungsaufgaben anbetrifft, kommen im Lehrwerk die Arbeitsanweisungen wie folgt vor:

„**Übersetzen Sie ins Arabische!**“

---

<sup>155</sup> Ebd, S: 115

Es wird festgestellt, dass auch die Arbeitsanweisungen in der Prüfung mit denen des im algerischen Gymnasium geltenden Lehrwerks „**Vorwärts mit Deutsch**“ für das 3. Studienjahr völlig kongruieren.

Bezüglich des 27. Kriteriums, ob die Lerner bei der Erstellung der Hauptschulprüfung gefragt wurden, welchen Aufgabentyp sie für angemessen halten und in dem sie bessere Leistungen haben, stehen in der Prüfungsanleitung sowie im Curriculum keine Hinweise darauf, dass die Lerner bei der Erstellung dieser Prüfung mitbeteiligt haben. Auch keine Anweisungen, dass die Lehrkraft in den Lernkontrollen die Lerner fragt, welchen Aufgabentyp sie bevorzugen, sind vorgefunden, obwohl es bekannt ist, dass die Aufgabenvielfalt die Performanz der Lerner verbessert.

Die durchgeführte Analyse des Testimpacts hat Folgendes ergeben:

Das Lesen des Textes und das Lösen der Aufgabe „Fragen zum Text“ könnten innerhalb der Prüfungserfahrung einen Lernprozess erlösen. Bezüglich der Feedbackfunktion, haben alle befragten Lehrer angegeben, dass sie eine Verbesserungsstunde vor der Noteneingabe organisieren. So kann mit Vorbehalt festgestellt werden, dass das Feedback seine Funktion erfüllt.

Was das 26. Kriterium anbetrifft, steht in der Prüfungsanleitung der Hauptschulprüfung für algerische Gymnasiasten fest, dass die Lehrkraft sich bei der Testerstellung nur auf die Themen, die im Programm festgelegt sind, beschränken muss.

Dies ist für die Versicherung eines hohen Interaktivitätsgrad, und für die Verstärkung der Lernerfahrung mit den gelernten Inhalten von extremer Wichtigkeit. Durch die Einsetzung ihres funktionalen Wissens bzw. des Leseverstehens können die Schüler neues Wissen erwerben, Es wurde auch festgestellt, dass das Feedback theoretisch seine Funktion erfüllt, da die Testinhalte den Lerninhalten entsprechen. Was die Mitbeteiligung der Lerner während der Testerstellung angeht, ist in der Prüfungsanleitung sowie im algerischen aktuellen Curriculum nicht darauf hingewiesen worden, dass die Schüler bei der Testentwicklung befragt wurden oder dabei beteiligt haben. Bezüglich des 28. Kriteriums, hat die Analyse ergeben, dass den Lernern wegen eines Defizits in der Punkteskala kein qualitatives Feedback dargeboten wird.

### **2.3 Ergebnisse und Auswertung der Lernerbefragung**

Um die Lage der Leistungsbeurteilung im algerischen DaF-Unterricht im Gymnasium praxisbezogen zu analysieren, wurde eine Umfrage unter 100 Gymnasiasten im Fach

„Deutsch als dritte Fremdsprache“ durchgeführt. Der Fragebogen wurde in einem einfachen Deutsch erstellt und darüber hinaus den Lernern bei jeder Frage weiter auf Arabisch erklärt worum es geht.

Die Analyse der Lernerantworten hat ergeben, dass 95% der Lerner die Tests und Prüfungen positiv einsehen. 3% haben angegeben, dass sie die Prüfungen negativ betrachten, 2 % der befragten Lerner haben angegeben, dass die Prüfungen für sie eine Sanktion bedeuten. Bezüglich der zweiten Frage, haben 98% angegeben, dass der Lehrer ihnen erklärt, wie bei Lernkontrollen und Prüfungen die Note zustande kommt. 2% haben mit „trifft nicht zu „ geantwortet, es soll hierbei erwähnt werden, dass diese Quote „2%“ in der Statistik als „insignifikant“ oder unbedeutend angesehen wird, wenn sie mit 98% verglichen wird.

Bezüglich der dritten Frage, haben 57% der befragten Lerner angegeben, dass ihre Noten wirklich ihre Leistungen auf dem Blatt widerspiegeln. 43% der Lerner haben angegeben, dass ihre Noten ihrem tatsächlichen Leistungsstand nicht entsprechen. Was die vierte Frage anbetrifft, haben 8% angegeben, dass die Hauptschulprüfung schwer sei. 82% der befragten Lerner haben angegeben, dass die Prüfung von mittelmäßigem Schwierigkeitsgrad sei. Bezüglich der Arbeitsanweisungen haben 68% der befragten Lerner angegeben, dass die Arbeitsanweisungen verständlich seien. 32% haben dagegen angegeben, dass sie die Arbeitsanweisungen nicht verstehen. Ein Lerner hat dazu geschrieben, dass er die Arbeitsanweisungen nicht liest und, dass er direkt antwortet. Dies ist dazu zu führen, dass die Prüfung für die Lerner vertraut ist, so dass sie je nach der Aufgabenform antworten. Was den Einfachheitsgrad der vierten Frage in der Aufgabe „Textverständnis“ anbetrifft, hat keiner der befragten Schüler angegeben, dass er die Frage einfach findet. 94% der befragten Lerner haben angegeben, dass sie die Frage schwer finden. 6% der befragten Lerner haben angegeben, dass die Frage von mittelmäßigem Schwierigkeitsgrad sei.

80% der Lerner haben angegeben, dass die Zeitanweisung in den Lernkontrollen ungenügend finden. 77% der Lerner haben angegeben, dass die Zeitanweisung in der Hauptschulprüfung genügend sei. Auf die Frage, ob es vor der Noteneingabe eine Verbesserungsstunde gibt, haben 100% der Lerner mit „Ja“ geantwortet. 84% der befragten Lerner haben angegeben, bezüglich dieser Verbesserungsstunde, dass sie es für wichtig halten, an der Verbesserungsstunde teilzunehmen. Ein Schüler hat geschrieben, dass er dadurch seine Lücken und Fehler erkennt. 16% haben angegeben, dass sie die Verbesserungsstunde langweilig finden. Keiner der befragten Lerner hat bei dem Umfeld „unnützlich“ angekreuzt. Die Ergebnisse zeigen, auffallend, dass die Lerner einen bedeutenden Bewusstsein besitzen, was das Thema Prüfungen und ihre Wichtigkeit anbelangt.

## **2.4 Ergebnisse und Auswertung der Lehrerbefragung**

Bei der Durchführung haben mir verschiedene Schwierigkeiten begegnet, u.a. dass die Deutschlehrer meist in zwei Gymnasien arbeiten, so dass es schwer ist, eine große Stichprobe zu sammeln. Darum musste die Befragung während der Abiturkorrektur durchgeführt werden. Dies ist für den Befragungsinhalt ein entsprechender Kontext. darüber hinaus wurden Lehrer aus verschiedenen Gegenden in Algerien, besonders aus Osten Algeriens (Constantine, Skikda, Bordj Bou Arreridj, Bouira.) befragt. Die Ergebnisse des Lehrerfragebogens waren wie folgt:

Bezüglich der Frage, „wo haben Sie ihre Ausbildung erhalten?“ haben 66,66% der befragten Stichprobe angegeben, dass sie ihre Ausbildung an der Universität und auch an dem ITE bekommen.

Die Lehrer, die in dem Umfeld „ ITE „ angekreuzt haben, haben auch angegeben, dass die Ausbildungsdauer genügend war. Es wurde hierzu auch festgestellt, dass die Lehrer, die angegeben hat, dass sie sowohl in dem ITE und auch an der Universität ihre Ausbildung bekommen haben, haben auch angekreuzt, dass die Ausbildung theoretisch und praktisch war. 33,33% haben angegeben, dass sie spezifische Informationen über die Leistungsbeurteilung bekommen haben.

Es wird anhand der Lehrerfragebogenanalyse festgestellt, dass die Lehrer, die angegeben haben, ihre Ausbildung nur an der Universität gemacht zu haben 66,66% der ausgewählten Stichprobe“, geben auch an, dass die Ausbildungsdauer für sie ungenügend sei.

Auf die Frage, die die Gründe der Lernerfchlleistungen im Fach Deutsch betrifft, haben 16,66% der Lehrer angegeben, dass die Gründe lernerbezogen sind. Eine Lehrkraft hat darüber hinaus geschrieben, dass die Lerner keine Lust am Lernen haben. Nur 2 Lehrer haben angegeben, dass die Prüfungsform inadäquat ist. 66,66% haben angegeben, dass das Problem in dem Lehrwerk liegt. Bezüglich der Frage, ob es eine Verbesserungsstunde vor der Noteneingabe gibt, haben alle Lehrer mit „trifft zu“ geantwortet. Überraschenderweise haben 11,11% (2 Lehrer) der ausgewählten Stichprobe auf die Frage“ Würden Sie Dafür sein, dass die Arbeitsanweisungen auf einer anderen Sprache als Deutsch erstellt werden?“ mit „Nein“ geantwortet. Es soll darauf hingedeutet werden, dass die 2 Lehrer der Gruppe gehören, die auch darauf angekreuzt haben, ihre Ausbildung nur an der Universität erhalten zu haben. 88,88 % der Lehrer haben angegeben, dass sie es besser finden, die Arbeitsanweisungen auf einer den Lernern verständlicheren Sprache zu formulieren. Diese Ergebnisse deuten darauf hin, dass die Lehrer, die an dem Fragebogen teilgenommen haben, hinsichtlich der Problematik der Leistungsbeurteilung im algerischen DaF-Unterricht eine bestimmte Haltung haben.

Was die Lehrerstrategien bei der Erstellung der Testaufgaben anbetrifft haben 16,66% angegeben, dass sie die Testaufgaben nach dem Einfachheitsgrad organisieren. Zwei andere haben angegeben, dass sie die Testaufgaben nach den getesteten Kompetenzen und Fertigkeiten einteilen. 72,22% der Lehrer haben angegeben, dass sie die Testaufgaben nach der Ordnung der Lerninhalte im Programm einreihen. Bezüglich der Frage, worauf sich die Lehrer bei der Auswahl der Testinhalte stützen, haben 55,55% angegeben, dass sie die Testinhalte auswählen, dass sie den Lerninhalten entsprechen. Zwei Lehrer haben angegeben, dass sie die Testinhalte wählen, die den Lernerinteressen entsprechen. 33,33% haben angegeben, dass sie dabei berücksichtigen, dass die Testinhalte sowohl den Lerninhalten als auch den Lernerinteressen entsprechen. Diese Frage zielte auf die Einschätzung des Prüfungsverhaltens der Lehrer in Bezug auf die Kriterien Authentizität und Interaktivität. Was die Benotungsprobleme anbelangt, haben 66,66% angegeben, dass die vierte Frage der zweiten Aufgabe des ersten Sub-Tests, genauerweise die Frage“ Erklären Sie diesen Satz“ für die Lerner schwer sei. 27,77% haben angegeben, dass sie die Frage unnützlich finden. Eine Lehrkraft hat angegeben, dass die Frage einfach sei. Bezüglich der Beurteilungsprobleme im Subtest Grammtik, haben 83,33% angegeben, dass sie im Fall der Antwort auf einen einzigen Teil der Struktur den halben Punktwert eingeben. Nur drei Lehrer haben angegeben, dass sie keinen Punktwert geben. Dies drückt eine Lehrerfreundliche Haltung gegenüber der Lernerperformanz aus, um den Schwierigkeitsgrad der Aufgabe abzumildern. Andererseits

drücken diese Ergebnisse Defizite in der Punkteskala aus, die dazu führen, dass die Lehrer die Lernerleistungen verschiedenwertig beurteilen. Bezüglich der Frage, ob die Lehrer die Prüfung für valide halten, haben 16, 66% der Lehrer angegeben, dass sie Prüfung unvalid sei. Ein Lehrer hat zwar alternative Aufgabentypen vorgeschlagen, wie Lückentexte, Substitutionübungen und Zusammenfassungen. Dies ist jedoch positiv zu bewerten.

## **2.5 Analyse und Auswertung der Schülerkopien**

Die Lernerkopien werden als Teil der Prüfung betrachtet. Das Antwortblatt kann das Testblatt selbst sein, oder die Testfragen begleiten. Bei der Analyse der Lernerkopien wird ein anderer Aspekt der Validierung im Vordergrund gerückt. Die Prüfung wird im Hinblick auf die Lernerreaktionen auf die Testaufgaben analysiert und im Fall der Kongruenz mit den Gütekriterien validiert. Die Validierung wird auf der Basis der Lernerreaktionenanalyse auf den Test erfolgen bzw. es wird erhoben, in welchen Aufgaben waren die Lerner leistungsstark.

Die Lernerkopienanalyse ermöglicht auch die Messung des Schwierigkeitsgrads der Aufgaben. Die Kopienanalyse dient auch zur Erforschung der wahren Gründe hinter den Fehlleistungen der Lerner. Die Analyse der Lernerkopien ist für die Validierung des Tests wichtig, da eine Testerprobung bei jeder Testerstellung unentbehrlich ist. Zusammenfassend ist die Lernerkopienanalyse eine empirische Validierung der vorab eingeführten Prüfungsanalyse anhand der Kriterien von BACHMAN und PALMER (1996).

Der Korpus meiner Analyse stellt 200 Schülerkopien bzw. Antworten auf die Prüfung für algerische Gymnasialen Fremdsprachenlerner dar. Das Thema der Prüfung wurde aus der fünften Lektion ausgewählt, und betrifft den Einfluss der Informationsmittel oder Medien auf den Konsum. Die Analyse der Kopien wird im größten Teil auf die Sub-Tests Textverstehen und Schreibkompetenz fokussieren, die die rezeptiven und produktiven Kompetenzen darstellen, während der Subtest Sprachkompetenz passive Kenntnisse misst. Es wird im Nachfolgenden überprüft werden, welche Reaktionen die Lerner auf die Testaufgaben erweisen. Es wird ebenfalls der Schwierigkeitsgrad der Prüfung gemessen. Die Lerner werden nach einem in der Prüfungsanleitung vorgegebenen Testmodell geprüft, um das Validierungsverfahren abzuschließen. Die experimentellen Verfahren fanden zwischen Februar und Mai 2010 statt.

Die Lerner werden zwei Hauptschulprüfungen ausgesetzt, Gründe dieser Auswahl ist die Vermeidung eventueller Einflüsse der Prüfung auf die Lernerkompetenz, die signifikanten Divergenzen sind nur diejenigen, die in den meisten Lernerkopien rekurren. Abschließend, soll hierzu erwähnt, dass die hier erbrachten Resultate keinen Anspruch auf absolute Vollständigkeit, bzw. Richtigkeit erheben. Sie treffen vielmehr die 4. Gymnasien zu, in denen die empirischen Untersuchungen durchgeführt wurden.

Im Nachfolgenden werden die Lernerergebnisse in der Prüfung Deutsch für algerische Gymnasiasten des dritten Studienjahres, Fachbereich Fremdsprachen, analysiert. Ziel der Analyse ist die Feststellung des Schwierigkeitsgrads der Prüfung, sowie die Validierung der bei der Kriterienanalyse gesammelten Feststellungen. Die Analyse zielt überdies auf die empirische Erforschung der Lernerreaktionen auf die Testaufgaben. Was dieses Kapitel betrifft, wird vielmehr auf die Analyse und Erforschung der Lehrerreaktionen auf die Lernerperformanz anhand der vorgegebenen Punkteverteilung fokussiert. Eine Fehleranalyse ist jedoch nicht das Ziel dieser Testexperimentation. Es ist an dieser Stelle zu erwähnen, dass die Beurteilung der Lernerantworten im Fach Deutsch im algerischen Prüfungssystem auf eine Punkteskala von 0 bis 20 stützt. 20 stellt den vollen Punktwert (100%) dar. Der in der Hauptschulprüfung erhaltene Punktwert entscheidet über den Übergang in die nächste Klasse. Im Fall des dritten Studienjahres ist die Prüfung vorwiegend eine Lernerfolgskontrolle und zielt darauf ab, die Lerner für die Abiturprüfung vorzubereiten. Im vorangegangenen Kapitel wurde festgestellt, dass bezüglich der erhaltenen Gesamtnoten bzw. des globalen Punktwerts haben 2, 28% die Note zwischen 18 bis 20 erhalten, die dem Punktwert 80 bis 100% entsprechen. 1,42% der Lerner haben den Punktwert von 14 bis 17 erhalten. 28,57% der Lerner haben den Punktwert von 11 bis 14 erhalten. 12,85% der Lerner erlangen den Punktwert von 9 bis 10. Der Anteil der Lerner, die den Punktwert von 4 bis 8 erhalten haben beträgt 34, 28% der Lernergesamtzahl.

Die Lerner, die eine der 4. Schwelle niedrige Note erhalten haben, stellen auffallend 18,57% der Lernergesamtzahl dar. Wie aus den vorab eingeführten Statistiken, ergibt sich, dass nur 38,57% der erforschten Lernerstichprobe den Durchschnitt bekommen haben.

Aus den oben vorgelegten Ergebnissen stellt sich heraus, dass der Schwierigkeitsgrad der Prüfung insgesamt hoch ist, und dies ohne die regionalen Besonderheiten bzw. die Lernereigenschaften im Betracht zu nehmen. Bezüglich der ersten Aufgabe zum Überprüfen des Leseverstehens haben nur 35,71% der Lerner den vollen Punktwert erhalten trotz der auffallenden Einfachheit dieser Aufgabe. Anhand der Analyse wird erhoben, dass die

Lehrkraft bei der Benotung die Richtigkeit bzw. die Unrichtigkeit der Begründung in der ersten Aufgabe des Tests nicht berücksichtigt hat, sondern nur ob der Lerner bei der Ankreuzung falsch bzw. richtig liegt. Eine weitere Überprüfung des Korpus hat ergeben, dass nur 15% der Lerner die den vollen Punktwert erhalten haben, auch bei der Begründung richtig begründet haben. Es wurde festgestellt, dass die Lehrkraft nur die richtige Ankreuzung im Betracht nimmt, da nur ein kleiner Anteil der Lerner sowohl bei der Ankreuzung als auch bei der Begründung richtig antwortet. Dies bestätigt was bezüglich der Benotungsprobleme in dieser Aufgabe festgestellt wurde. Fraglich ist, ob die Reliabilität der Aufgabe dadurch nicht beeinträchtigt wird. Bezüglich der zweiten Aufgabe des Testteils Textverständnis, hat die Analyse folgende Anteile ergeben: 95% der Lerner haben dem Text einen Titel gegeben. Es wird bei der Analyse festgestellt, dass der Anteil 95% nur mit Vorbehalt angenommen werden kann. 90% der Lerner d. h. 180 der ausgewählten Lernerstichprobe haben einen dem Text entsprechenden Titel gegeben. Die verbleibende Rate haben zwar einen Titel eingegeben, aber der Thematik nicht treu entspricht. Die meisten Lerner haben bei der Eingabe des Titels sich von dem Text inspiriert. Bezüglich der Aufgabe "Fragen zum Text" haben alle Lerner der ausgewählten Stichprobe zumindest eine Frage zum Text richtig gelöst.

Es sollte erwähnt werden, dass die Benotungsprobleme dort eintreten, wo die Lerner ganze Abschnitte aus dem Text abschreiben, insbesondere wenn dieselben Wörter in der Frage auch im Text rekurren. Bei der Analyse haben 90% der Lerner ganze Abschnitte aus dem Text entnommen; in der ausgewählten Stichprobe.

Was die vierte Frage des ersten Prüfungsteils „ Erklären Sie diesen Satz“ anbetriift, antworten nur 2,5% richtig auf die Frage.

7,5 der Lerner antworten auf die Frage nicht.

90% der Lerner geben falsche Antworten an.

In dem Subtest „ Grammatik“ haben 33, 33% der Lerner 2, 5 oder mehr (50%), von den 5 Punkteinheiten, die die Aufgabe gibt, bekommen.

Keiner der Lerner in der ausgewählten Stichprobe hat in der Grammatik den vollen Punktwert (5/5) erlangen.

Bezüglich der Lehrerreaktion auf die Lernerantworten, wird anhand der Analyse festgestellt, dass die Lehrkraft den halben Punktwert eingegeben hat, im Fall der richtigen Antwort auf einen einzigen Teil der Struktur.

16, 12 % der Lerner in der ausgewählten Stichprobe überschreiten die zwei Testteile bzw. sie antworten lediglich auf dem ersten Subtest „ Textverständnis“.



Diese Lerner wurden auch bei der Errechnung der Lernerleistungen im Subtest „ Grammatik“ miteinbezogen.

In dem Testteil „ Schreibkompetenz“ erreichen 51, 61% der Lerner einen Punktwert, der 50% oder mehr von dem allgemeinen Punktwert für diesen Testteil entspricht.

29, 03 % der Lerner versuchen nicht auf die entsprechende Frage zu antworten bzw. einen Aufsatz zu schreiben.

### **Zusammenfassung**

Aus dem praktischen Teil hat es anhand der empirischen Forschung folgendes Bild ergeben:

Das Testmodell für den algerischen Deutschunterricht im Gymnasium erfüllt die Formkriterien. Die Einflüsse, die die Testmethode auf die Lernerperformanz haben, kommen aus einem Überschreiten der Testkriterien und nämlich aus dem Nicht- Respektieren der Validitäts -und Reliabilitätskriterien.

Das Kriterium der Validität ist nur mit Vorbehalt erfüllt, eine Verletzung dieses Kriteriums kommt zustande, wenn die Testinhalte den im Curriculum festgelegten Richtlinien widersprechen. Dies ist in dem zweiten Teil der Prüfung „ Grammatik“ ersichtlich, da der algerische Daf- Unterricht sich dem kommunikativen Ansatz verpflichtet.

Darüber hinaus wurde die Validität des Tests abgenommen wegen der Inszenierung der Kombiniertenaufgabe im Subtest Textverständnis, die die Lerner ohne Zurückgreifen auf den Text lösen können und somit das Textverständnis nicht misst bzw. die Erhebung der beabsichtigten Daten über die Lernertextverständnis nicht erlaubt.

Die vierte Frage der Testaufgabe „Textverständnis“ weist einen sehr niedrigen Validitätsgrad auf in Bezug auf ihren hohen Schwierigkeitsgrad.

Das Kriterium der Reliabilität wird verletzt im Testteil „**Grammatik**“ in dem die Lehrkraft keinen detaillierten Richtigkeitskriterien bekommt.

16, 12% der Lerner schreiten die Testteile Grammatik und Schreibkompetenz über. Es wurde im praktischen Teil der vorliegenden Arbeit demonstriert, anhand der Testerprobung, dass die Inszenierung der Aufgaben zum Testen des Regelsprachwissens im zweiten Sub- Test „ Grammatik“ und das Testen der grammatischen Regeln auch im Sub- Test Schreibkompetenz hinter den Fehlleistungen der Lerner liegen. Die Lerner werden wegen dieser doppelten Grammatikbenotung zwei Mal in den gleichen Strukturen getestet und somit benachteiligt.

Im empirischen Teil der vorliegenden Arbeit wurde klar, dass eine perfekte Gewichtung der Gütekriterien fast utopisch ist. Im Sub- Test Textverständnis realisieren die Fragen zum Text

einen hohen Validitätsgrad aber hinsichtlich des Reliabilitätskriteriums wurden Benotungsprobleme notiert.

### **Zusammenfassung und Ausblick**

Die vorgelegte Magisterarbeit erforschte den potenziellen Grund der Lernerfehleistungen im gymnasialen Deutschunterricht in Algerien. Es wird bei der Forschung auf den Benotungsfehler, die der Prüfungsform zurück zu führen könnten, betont. Hauptanliegen war ein Versuch die Prüfung anhand bestimmter Kriterien zu validieren. Es wurde von einem Prinzip der Testtheorie ausgegangen, dass eine Prüfung nicht nur das einzelne Prüfungsblatt darstellt, vielmehr ist sie eine Fülle von Bezugsdokumenten und äußeren Bedingungen, die die Prüfungsidentität ausmachen.

Aus dem theoretischen Teil hat sich ergeben, dass die Leistungsbeurteilung ein wichtiges Unterrichtsvehikel ist; die Funktionen, die Tests und Prüfungen in einem Sprachunterricht haben, können die Selektionsfunktion, die Einstufungsfunktion, die Lernstandskontrolle usw. sein.

Die Testmethode ist nicht von der Lehrmethode zu trennen, deshalb sind Tests heutzutage kommunikativ orientiert und spiegeln die curricularen Richtlinien wider.

Die Testergebnisse könnten für die Lernerlaufbahn entscheidend sein. Daher ist die Erstellung eines validen Tests vor allem eine ethische Problematik. Die Erstellung eines validen Tests ist überdies eine weitgehend komplexe Aufgabe, die ein Zurückgreifen auf die Bezugsdokumente wie das Curriculum und das Lehrwerk benötigt. In diesem Kontext müssen die Prüfungszwecke, sowie die Zielgruppe klar beschrieben werden. Die Testgütekriterien, Reliabilität, Konstruktvalidität und Authentizität müssen bei der Testerstellung respektiert werden. Die Prüfungserstellung muss sich darüber hinaus an den allgemeinen und spezifischen Zielsetzungen der Schulprogramme und Lerninhalte orientieren.

Im theoretischen Teil wurden die spezifischen Testdesignprinzipien aufgestellt, u.a. dass die Testaufgaben je nach ihrem Einfachheitsgrad vom Einfachen zum Schweren graduiert werden müssen. Seitdem die Testergebnisse über die Lerner entscheiden, müssen die Testgütekriterien im Vordergrund stehen. Dies lässt sich dadurch begründen, dass die Ergebnisse in einem unvaliden Test vielmehr der Resultat der Messfehler sind und weniger des getesteten Konstrukts.

Die theoretische Forschung führte zur Schlussfolgerung, dass die Gütekriterien in einem engen Zusammenhang realisierbar sind, so dass die Beeinträchtigung eines einzigen Gütekriteriums zur Beeinträchtigung der allgemeinen Testgültigkeit führt. Dies schließt die

Tatsache nicht aus, dass eine absolute oder perfekte Gewichtung der Gütekriterien jedoch utopisch ist.

Der empirische Teil hat ergeben, dass sich die Hauptschulprüfung für algerische Gymnasiasten im Fach Deutsch hinsichtlich der Form lernerfreundlich aufweist, indem sie zwei Prüfungsvariablen vorgibt. Die dritte Testaufgabe „Schreibkompetenz“ als eine offene Form realisiert einen hohen Interaktivitätsgrad mit den Schülern und erfüllt darüber hinaus die Testgütekriterien Validität und Authentizität. 43% der befragten Lerner haben angegeben, dass ihre Noten ihren tatsächlichen Leistungen nicht entsprechen. 32% der Lerner verstehen die Arbeitsanweisungen nicht. Die Lehrer ihrerseits haben angegeben, dass sie die Arbeitsanweisungen auf einer anderen Sprache als Deutsch zu erstellen vorziehen. Die befragten Lehrer haben die Fehlleistungen nicht der Testpraxis zurückgeführt, sondern dem Lehrwerk und anderen Faktoren, die lernerbezogen sind. Die Erprobung der Hauptschulprüfung in zwei algerischen Gymnasien hat ergeben, dass die Lernerleistungen schwach waren (38,57% der ausgewählten Stichprobe erhalten den Durchschnitt). Dies ist u.a. einem hohen Schwierigkeitsgrad der Testaufgaben zurückzuführen. Der Schwierigkeitsgrad der vierten Frage des ersten Testteils ist von 97,5%. Aus der Analyse des vom Ministerium vorgegebenen Testmodells wurde festgestellt, dass die Prüfungsanleitung dem kommunikativen Ansatz des algerischen Deutschunterrichts widerspricht, indem die Prüfung ein wichtiger Teil für die Sprachanalyse vorgibt. Überdies erhalten die Lerner 11 Punkte von 20 für die sprachlichen Mittel allein. Darüber hinaus wird die grammatische Kompetenz der Lerner im Subtest „Grammatik“ und auch in der Testaufgabe „Schreibkompetenz“ gemessen, was eine Beeinträchtigung eines didaktischen Prinzips stellt, und zur Benachteiligung der Lerner führt. Der Subtest „Sprachkompetenz“ erfüllt somit das Kriterium der Validität nicht, da der kommunikative Ansatz, woran sich der algerische Deutschunterricht orientiert, die Sprachanalyse nicht fördert. Zusammenfassend kann dazu gesagt werden, dass die Inszenierung der Grammatikaufgaben zum Teil hinter einer Schülerfehlleistung liegt. Es hat sich auch herausgestellt, dass der Prüfungsteil „Textverständnis“ Defizite in Bezug auf die Gütekriterien aufweist. An dieser Stelle sei die Kombinationsaufgabe erwähnt, die der Lerner ohne Zurückgreifen auf den Text lösen kann und somit das Textverständnis nicht misst, sondern die Vokabel.

Die Testaufgabe Richtig/ Falsch ist wenig zuverlässig für das Testen des Leseverstehens, und erschwert die Benotungsprozeduren. Als alternative Formen werden offene und halboffene Aufgaben wie Substitutionsübungen vorgeschlagen. In diesem Kontext ist es wichtig die Lerner, sowohl im Gymnasium als auch an der Universität, bei der Auswahl der

Testformen durch Befragungen und Kommunikation einzubeziehen, um zu erheben, welche Testformen an denen sie bessere Leistungen erbringen.

Die Analyse des Lehrwerks im Licht der Leistungsbeurteilung hat gezeigt, dass die Lerner auf eine Frage antworten müssen, die sie nur einmal und in einem einzigen Text im Lehrbuch behandeln, und als feste und rekurrierende Aufgabe in der Prüfung steht. Es handelt sich um die vierte Frage im Subtest „Fragen zum Text“. Außer der vierten Frage im ersten Prüfungsteil sind alle Aufgaben im Laufe des Schuljahrs behandelt worden. Das Lehrwerk enthält keine Checklisten und fördert die Selbstbeurteilung nicht.

### **Ausblick**

Die empirische Arbeit hat Anregungen zu zukünftigen Forschungen im Bereich des Testens und Prüfens im algerischen DaF-Unterricht angegeben.

-Die Aktualisierung des Lehrplans in Bezug auf die neuen Beurteilungsformen wie Portfolios und Lerntagebücher.

-Beitrag des Lehrwerks „Vorwärts mit Deutsch“ zur Optimierung der Selbstbeurteilung.

-Die Entwicklung einer neuen Hauptschulprüfung, die mit einem Design-Statement und einer detaillierten Prüfungsanleitung begleitet wird.

Die Testpraxis im algerischen Daf- und Germanistikunterricht sollte aufgabenorientiert erfolgen, d. h. der Lerner wird die Testerfahrung als die Lösung einer sprachlichen Aufgabe, die dem Zielsprachgebrauch entsprechen muss, um den Authentizitätsgrad der Testformen zu erhöhen.

Die Hochschullehrpläne sollten aktualisiert werden, dass die Lehr-und Forschungsstudenten mit dem Thema Leistungsbeurteilung im Sprachunterricht sowohl theoretisch als auch praktisch durch Hospitationen in den Gymnasien vertraut werden. Zusammenhängend sollten den Lehrern ernste Weiterbildungsseminaren zum Thema Leistungsbeurteilung veranstaltet werden, um die Inszenierung anderer Beurteilungsformen wie Portfolios und Lerntagebücher zu begleiten. Welchen Stellenwert diese neuen Beurteilungsformen zur Verbesserung der Lernerperformanz haben könnten, ist anderen Forschungen überlassen.

## **Zusammenfassung und Ausblick**

Die vorgelegte Magisterarbeit erforschte den potenziellen Grund der Lernerfehlleistungen im gymnasialen Deutschunterricht in Algerien. Es wird bei der Forschung auf den Benotungsfehler, die der Prüfungsform zurück zu führen könnten, betont.

Aus dem theoretischen Teil hat sich ergeben, dass die Leistungsbeurteilung ein wichtiges Unterrichtsvehikel ist; die Funktionen, die Tests und Prüfungen in einem Sprachunterricht haben, können die Selektionsfunktion, die Einstufungsfunktion, die Lernstandskontrolle usw. sein. Die Testmethode ist nicht von der Lehrmethode zu trennen, deshalb sind Tests heutzutage kommunikativ orientiert und spiegeln die curricularen Richtlinien wider. Der empirische Teil hat ergeben, dass sich die Hauptschulprüfung für algerische Gymnasiasten im Fach Deutsch hinsichtlich der Form lernerfreundlich aufweist, indem sie zwei Prüfungsvariablen vorgibt. Die dritte Testaufgabe „Schreibkompetenz“ als eine offene Form realisiert einen hohen Interaktivitätsgrad mit den Schülern und erfüllt darüber hinaus die Erprobung der Hauptschulprüfung in zwei algerischen Gymnasien hat ergeben, dass die Lernerleistungen schwach waren (38,57% der ausgewählten Stichprobe erhalten den Durchschnitt). Aus der Analyse des vom Ministerium vorgegebenen Testmodells wurde festgestellt, dass die Prüfungsanleitung dem kommunikativen Ansatz des algerischen Deutschunterrichts widerspricht, indem die Prüfung ein wichtiger Teil für die Sprachanalyse vorgibt. Zusammenfassend kann dazu gesagt werden, dass die Inszenierung der Grammatikaufgaben zum Teil hinter einer Schülerfehlleistung liegt. Es hat sich auch herausgestellt, dass der Prüfungsteil „Textverständnis“ Defizite in Bezug auf die Gütekriterien aufweist. An dieser Stelle sei die Kombinationsaufgabe erwähnt, die der Lerner ohne Zurückgreifen auf den Text lösen kann und somit das Textverständnis nicht misst, sondern die Vokabel.

Mots clés :

Leistungsbeurteilung; Algerischer Daf-Unterricht; Blue-Print; Kriterienkatalog von Bachman und Palmer; Test; Hauptschulprüfung für das algerische Gymnasium; Textverständnis nicht misst.