



universität
wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Möglichkeiten und Grenzen einer standardisierten Leistungsbeurteilung“

Einstellungen Französischlehrender zur Leistungsbeurteilung schriftlicher
Performanzen mittels Beurteilungsraster auf Sprachniveau B1

verfasst von / submitted by

Manuel Roth, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Philosophie (Mag. phil.)

Wien, 2017 / Vienna, 2017

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 190 333 347

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Lehramtsstudium UF Deutsch UF Französisch

Betreut von / Supervisor:

Univ.-Prof. Mag. Dr. Eva Vetter

Mitbetreut von / Co-Supervisor:

Mag. Dr. Marie-Luise Volgger-Bahensky

„Es gibt Leistung ohne Erfolg,
aber keinen Erfolg ohne Leistung.“

- LA ROCHEFOUCAULD

Danksagung

An dieser Stelle möchte ich mich herzlich bei all jenen bedanken,
die mich im Laufe meines Lebens und Studiums bis zu diesem Punkt
geleitet und begleitet haben.

Enge Freunde sowie Kommilitoninnen und Kommilitonen, die zu liebenswürdigen Begleitern
wurden, erheiterten den Studienalltag.

Vor allem aber gebührt ein großer Dank meine Familie, die mich in all meinen Taten
stets unterstützt, gestärkt und motiviert hat.

Danke dafür!

Besonderer Dank zur Betreuung dieser Diplomarbeit gilt Frau Prof.ⁱⁿ Mag.^a Dr.ⁱⁿ Eva Vetter
sowie Frau Mag.^a Dr.ⁱⁿ Marie-Luise Volgger-Bahensky, die mit viel Engagement und
geistreichen Denkanstößen das Verfassen dieser Arbeit ermöglichten.

Ebenfalls möchte ich mich gebührend bei den kooperativen Lehrpersonen bedanken, die einen
Einblick in ihre Gedanken gewährt und ihre Erfahrungen geteilt haben, wodurch sie einen
wertvollen Beitrag zu meinen Forschungszwecken geleistet haben.

Danke! Merci!

Inhaltsverzeichnis

Abstract	1
I. Einleitung	3
II. Theoretischer Kontext	5
1 Assessment	5
1.1 Assessment im Fremdsprachenunterricht.....	7
1.1.1 Funktionen	8
1.1.2 Formen.....	10
1.2 Measurement und Testing	13
1.2.1 Testarten	14
1.2.2 Gütekriterien.....	16
1.3 Evaluation.....	17
1.4 Herausforderungen des Fremdsprachenunterrichts.....	19
2 Leistungsbeurteilung im österreichischen Schulsystem.....	23
2.1 Das Leistungsprinzip	23
2.1.1 Der Leistungsbegriff.....	24
2.1.2 Leistungsformen	26
2.2 Leistungsbeurteilung	28
2.3 Bezugsnormen der Leistungsbeurteilung	29
3 Standardisierung des Fremdsprachenunterrichts.....	33
3.1 Bildungsstandards	33
3.2 Standardisierungsprozesse hinsichtlich schriftlicher Produktion.....	35

4	Standardorientierte, kriterienbasierte Leistungsbeurteilung schriftlicher Performanzen im Französischunterricht	37
4.1	Besonderheiten bei schriftlichen Performanzen	38
4.2	Bewertungskriterien	41
4.3	Beurteilungsraster	42
4.3.1	Arten von Beurteilungsrastern	44
4.3.2	Ziele und Funktionen von Beurteilungsrastern	45
4.3.3	Problematik der Interpretationsobjektivität.....	47
4.4	Beurteilungsraster des BIFIE auf Niveau B1	50
5	Zusammenfassung	55
III.	Empirische Untersuchung	57
6	Forschungsdesign.....	57
6.1	Forschungsfragen	57
6.2	Methodenwahl.....	58
6.2.1	Leistungsbeurteilung einer Performanz	58
6.2.2	Interviews	60
6.2.3	Auswertungsmethode	61
6.2.4	Lehrpersonen.....	62
7	Präsentation und Analyse der Ergebnisse	64
7.1	Beurteilungen der Performanz	64
7.2	Einstellungen der Lehrpersonen dem Beurteilungsraster gegenüber	66

7.2.1	Beurteilung schriftlicher Performanzen.....	66
7.2.2	Beurteilungsraster	67
7.2.3	Bezugsnormen	77
7.2.4	Performanz.....	79
7.2.5	Handlungsorientierung	79
7.2.6	Anregungen und Wünsche.....	80
8.	Zusammenfassung	82
IV.	Conclusio.....	83
V.	Résumé en langue française	88
VI.	Bibliographie.....	99
VI.1	Sekundärliteratur	99
VI.2	Internetquellen.....	107
VII.	Abbildungsverzeichnis	109
VIII.	ANNEX	110
VIII.1	Schriftliche Performanz	110
VIII.2	Intervieworganisation.....	114
VIII.3	Transkriptionen	120

Abstract

Diese Diplomarbeit wendet sich dem Themengebiet der Leistungsbeurteilung schriftlicher Performanzen mittels Beurteilungsraster im Französischunterricht zu und untersucht die Einstellung der Lehrenden diesem Beurteilungsinstrument gegenüber. In einem ersten Teil, der theoretischen Kontextualisierungen, wird der grundlegende Bereich des Assessments aufgearbeitet, bevor es zu definieren gilt, was eine Schülerleistung im Kontext des schulischen Fremdsprachenunterrichts darstellt. Darüber hinaus ist es unabdingbar, die standardorientierte und kriterienbasierte Leistungsbeurteilung näher zu beleuchten und in diesem Rahmen das Instrument eines Beurteilungsrasters analytisch zu betrachten. Darauf aufbauend wird der im österreichischen Französischunterricht zur Anwendung kommende Beurteilungsraster auf Sprachniveau B1 in den Fokus der Arbeit gerückt. Auf diese theoretische Annäherung folgt eine empirische Untersuchung basierend auf qualitativen Methoden der Sozialforschung, wobei mit Hilfe problemzentrierter Interviews mit Lehrpersonen, deren Einstellung und Haltung dem Raster gegenüber sowie Informationen hinsichtlich dessen Anwendung in ihrer Bewertungspraxis erhoben werden. Diese verbalen Daten werden inhaltsanalytisch ausgewertet präsentiert und diskutiert. Auf Basis dieser Resultate kann festgestellt werden, dass grundsätzlich einheitliche, als Leitfäden dienende Bewertungshilfen positiv angenommen werden, wohingegen der vorliegende Raster im Speziellen eher auf negative Ressentiments stößt, da er in der Anwendung zu wenig präzise und effektiv sowie zu zeitintensiv in seiner Handhabung sei.

Abstract (en langue française)

Ce travail ci-présent traitera du sujet de l'évaluation des performances, à l'aide d'une grille d'évaluation dans le contexte du cours de français du niveau de langue B1 en Autriche. Dans la première partie nous aborderons tout d'abord la base du sujet de l'évaluation et dans un second temps, l'évaluation dans le cadre scolaire. Non seulement la performance d'un/e élève est influencée par des facteurs externes mais aussi par l'évaluation du/de la professeur/e, c'est pourquoi il semble essentiel d'aborder les valeurs de références jouant un rôle central dans ce contexte. Dans le cadre de la standardisation du système scolaire en Autriche, le but était de créer une évaluation commune en définissant pour cela une grille d'évaluation déterminée par rapport aux standards minimums acquis en fin d'étude et s'appuyant sur des critères concrets et unitaires. Cette grille d'évaluation du BIFIE sera analysée. La deuxième partie de ce travail s'appuiera sur une étude empirique afin de pouvoir explorer les avis des enseignant(e)s sur ce domaine de travail, et plus particulièrement sur la grille d'évaluation au niveau de langue B1 ainsi que de la manière dont elle est utilisée pour l'évaluation. Les résultats montrent que la position des enseignant(e)s vis-à-vis de la grille d'évaluation du BIFIE est assez critique en raison de plusieurs défauts. Elle est appliquée de manière très variée et différente parmi les professeur(e)s de français en Autriche. Ce que nous pouvons remarquer dans le cadre de cette étude, est le souhait d'une nouvelle édition remaniée de la grille qui a été manifesté : la grille d'évaluation devrait être plus simple dans l'application et plus précis dans la formulation.

Abstract (in English)

This thesis discusses the assessment of written texts based on analytic rubrics in teaching French as a foreign language on level B1. It seems therefore necessary to illustrate initially the domain of foreign language assessment in Austrian schools in a theoretical way. Subsequently the appertaining rubric will be examined before the data of the empirical part, which consisted of a qualitative study in the form of interviews with French teachers, is presented. The findings prove that the teachers appreciate guidelines with benchmarks while this rubric of the BIFIE is severely criticized regarding different matters, especially the expenditure of time and the imprecise notion of the descriptors.

I. Einleitung

In der vorliegenden Diplomarbeit mit dem Titel *„Möglichkeiten und Grenzen einer standardisierten Leistungsbeurteilung. Einstellungen Französischlehrender zur Leistungsbeurteilung schriftlicher Performanzen mittels Beurteilungsraster auf Sprachniveau B1.“* soll ein Bereich des Assessments im schulischen Fremdsprachenunterricht unter Berücksichtigung von stattgefundenen Standardisierungsprozessen nicht nur theoretisch näher beleuchtet, sondern auch durch empirische Forschung mittels qualitativer Methoden untersucht werden.

Die Motivation meinerseits ergab sich durch ein großes Interesse an der Bewertung sprachproduktiver SchülerInnenperformanzen sowie an dem damit einhergehenden Unbehagen bei definitiven Urteilsbildungen hinsichtlich Leistungsbeurteilung, welches bei angehenden Lehrpersonen zu beobachten ist. Aus diesem Grund schien auch die Auseinandersetzung mit dem Beurteilungsraster für schriftliche Leistungen auf Sprachniveau B1 äußerst spannend, da dieser ein Instrument zur Bewertung repräsentiert und somit als Hilfestellung und Richtlinie fungiert, die prinzipiell das Unbehagen zu mildern vermag. In dieser Hinsicht soll die subjektive Wahrnehmung dieses Beurteilungsrasters bei verschiedenen Lehrpersonen untersucht und Einblick in deren Einstellungen diesbezüglich erlangt werden, um folgende forschungsleitende Fragestellung beantworten zu können: *„Wie lassen sich die Einstellungen Französischlehrender dem Beurteilungsraster des BIFIE¹ auf Sprachniveau B1 gegenüber beschreiben?“*.

Zunächst soll sich im Zuge dieser Arbeit dem Themengebiet theoretisch angenähert werden, indem unerlässliche das Assessment im Fremdsprachenunterricht betreffende Punkte erläutert werden, von Formen und Funktionen zu Gütekriterien und besonderen Herausforderungen des Französischunterrichts. In weiterer Folge wird die Leistungsbeurteilung im österreichischen Schulsystem behandelt, um rechtliche Rahmenbedingungen und Grundsätze, die dieser zugrunde liegen, aufzuklären. Da in der Diplomarbeit die Leistungsbeurteilung unter Berücksichtigung der in den letzten Jahren verstärkten Standardisierung des österreichischen Schulwesens aufgearbeitet wird, soll dieser Standardisierungsprozess in Hinblick auf den Fremdsprachenunterricht dargelegt

¹ Das Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens ist eine Institution des Bundesministeriums für Bildung zur Vermittlung zwischen Schulverwaltung, Forschung und schulischer Praxis (vgl. <https://www.bifie.at/>).

werden, um Punkte wie die Kompetenz- und Handlungsorientierung im Französischunterricht näher zu beleuchten. Diesbezüglich ist der Gemeinsame Europäische Referenzrahmen für Sprachen eine wichtige Quelle, die in direktem Zusammenhang mit der Standardisierung des Fremdsprachenunterrichts steht. Ebenso sollen Merkmale der standardisierten Leistungsbeurteilung im Fremdsprachenunterricht erarbeitet werden, bevor auf die speziellen Besonderheiten bei der Bewertung schriftlicher Performanzen übergeleitet wird. In diesem Kontext wird das Bewertungsinstrument eines Rasters eingeführt und untersucht. Hierbei sind unterschiedliche Arten sowie Funktionen und Ziele von Rastern aufzuzeigen, bevor im speziellen auf den im österreichischen Französischunterricht auf Sprachniveau B1 zur Anwendung kommenden Beurteilungsraster des BIFIE eingegangen wird. Diesem wird sich zunächst deskriptiv angenähert, worauf folgend eine auf die Empirie vorbereitende, analytische Betrachtung stattfindet, wobei einige Formulierungen der den Niveaustufen zuzuordnenden Deskriptoren betrachtet sowie voneinander abgegrenzt werden.

In einem zweiten Teil der Diplomarbeit soll die Leistungsbeurteilung mittels Beurteilungsraster empirisch erforscht werden. Der Praxisteil wird sich auf qualitative Methoden der Sozialforschung, die im Kontext des zweiten Teiles konkret erläutert werden, stützen. In einem ersten Zug wurde eine bestimmte schriftliche Performanz eines Schülers anonymisiert an unterschiedliche Lehrpersonen mit dem Auftrag um Beurteilung mittels Beurteilungsraster ausgehändigt. Die stattgefundenen Beurteilungen der Performanz wurden meinerseits analysiert und in Hinsicht auf Punktevergabe bei der Beurteilung ausgewertet sowie interpretiert. Darauf folgte ein problemzentriertes Interview mit diesen Französischlehrenden, in dem grundsätzlich ihre Einstellungen zur Leistungsbeurteilung aber auch konkret ihre Haltung dem Beurteilungsraster gegenüber erhoben wurden. Weiters wollte auf das Korrekturverhalten und die Anwendung des Rasters bei der Bewertung eingegangen werden, wobei die unterschiedlichen im Schulunterricht aufzufindenden Bezugsnormen mit dem Raster in Verbindung gesetzt wurden. In dem Interview, welches auf Grundlage der qualitativen Inhaltsanalyse nach MAYRING ausgewertet wurde, sollten die Lehrpersonen ebenfalls erläutern, wie die Note beziehungsweise Punktevergabe für sie zustande kam und begründen, worauf sie sich stützen. In diesem Sinne war die Einstellung zu dieser Bewertungshilfe, deren Handhabung sowie die Interpretation der einzelnen Deskriptoren auf den verschiedenen Niveaustufen im Zentrum des Forschungsinteresses.

II. Theoretischer Kontext

Im Rahmen dieser Diplomarbeit ist es notwendig, in einem ersten theoretischen Teil den grundlegenden Bereich des Assessments und seiner Teilbereiche näher zu erläutern, damit in weiterer Folge eine Auseinandersetzung mit der Leistungsbeurteilung schriftlicher Performanzen im österreichischen Schulwesen unter Berücksichtigung der stattgefundenen Standardisierungsprozesse möglich scheint und ein Basiswissen zum Verständnis der präsentierten Forschungsergebnisse schafft. Dafür wird in einem ersten Kapitel über das Assessment allgemein sowie im Kontext des Fremdsprachenunterrichts informiert, bevor die damit in Verbindung stehenden Subbereiche und Disziplinen in Hinsicht auf das Schulwesen, wie die Messung von Ergebnissen, die Testtheorie und das Feld der Evaluation, erläutert werden. Daran schließt eine Darlegung der Leistungsbeurteilung an, wobei zunächst das Prinzip der Leistung im Schulkontext betrachtet und definiert wird, ehe mögliche Bezugsnormen aufgezeigt werden. Bevor zum grundlegenden Bereich der Leistungsbeurteilung schriftlicher Performanzen mit Hilfe des Beurteilungsrasters übergegangen werden kann, muss die Standardisierung des Fremdsprachenunterrichts, welche einen maßgeblichen Einfluss auf die Kompetenzorientierung ausübte, behandelt werden. Im Anschluss wird die Besonderheit von schriftlichen Performanzen sowie das Beurteilungsinstrument eines Kompetenzrasters kritisch betrachtet und bearbeitet. Ein besonderer Fokus wird hierbei auf den im österreichischen Französischunterricht Anwendung findenden Beurteilungsraster des BIFIE auf Sprachniveau B1 gelegt, der im Zuge der standardisierten Reife- und Diplomprüfung als Beurteilungsmaßstab dient.

1 Assessment

Der aus dem Englischen entlehnte Terminus ‚Assessment‘ wird im Deutschen hinsichtlich seiner Bedeutung mit den Begrifflichkeiten ‚Bewertung‘ und ‚Einschätzung‘ gleichgestellt (vgl. DUDENREDAKTION ⁹2007a: 97). In diesem Bereich werden unterschiedliche Termini im Sprachgebrauch als synonym angesehen, ebenso in einschlägiger Fachliteratur (vgl. KLEBER 1992: 77), weshalb zunächst Beschreibungen des Alltagsverständnisses aus Wörterbüchern präsentiert werden, bevor eine fachspezifische Untergliederung, in Anlehnung an anglophone Fachwerke, stattfindet. Im Cambridge Dictionary wird *assessment* definiert als: „*the act of judging or deciding the amount, value, quality, or importance of something, or the judgment or decision that is made*“ (<http://dictionary.cambridge.org/dictionary/english/assessment>).

Demnach bedarf es eines Objektes, welches hinsichtlich möglicher verschiedener Kriterien bewertet oder eingeschätzt wird, weshalb Assessment im Allgemeinen als ein Prozess zu verstehen ist, der von einem aktiven Subjekt auf ein passives Objekt² angewandt wird.

Das Bedürfnis des Menschen sein Umfeld nach unterschiedlichen Maßstäben einzuschätzen und Verschiedenes zu bewerten, ist auf ein bewusst zielorientiertes Handeln, welches durch einen Optimierungsgedanken motiviert ist, zurückzuführen (vgl. WOTTAWA/THIERAU³2003: 27). Dieser Prozess ist notwendig, damit komplexe Umweltsituationen und insbesondere Interaktionssituationen bewältigt werden können und eine angemessene Reaktion darauf folgen kann. Mit Hilfe der kognitiven Kompetenz zur Informationssammlung und -reduktion sind Bewertungen mit vordiagnostischem Charakter möglich, die nicht in ihrer Summe ganzheitlich verarbeitet, sondern in simplen Einheiten zusammengefasst werden, wobei bereitgestellte Schemata zur Hilfe zu ziehen sind, wodurch sich der menschliche Klassifizierungsgedanke begründet (vgl. KLEBER 1992: 78 f).

„Le classement est un concept d’origine sociologique. À chaque instant nous classons les gens que nous connaissons et, donc les jugeons, émettons à leur propos une évaluation, c’est-à-dire les classons et leur conférons une identité. Eux aussi font de même à notre égard. Nous sommes classeurs et classés. [...] Or, la distinction (être capable de ne pas confondre et de juger comme jugent les instances de légitimation ; être capable d’être jugé comme quelqu’un d’adéquat) est le fondement du jugement que la société porte sur nous. En résumé, nous sommes tous « des classeurs classés par nos classements ».“ (CUQ 2003: 43)

Bewertungen und Klassifizierungen werden im Zusammenhang mit gesellschaftlichem Fortschritt und im Dienste einer Aufklärung vorgenommen. Deshalb gilt es abzuwägen, ob ein dienlicher Beitrag zur Klärung eines problematischen Sachverhalts gegeben ist oder die Besserung eines vorherrschenden Status in Aussicht steht. Der stetige Optimierungsgedanke und eine erwartete Problemlösungskompetenz sind demnach erhoffte Hauptziele (vgl. STOCKMANN/MEYER 2010: 17).

Diese Notwendigkeit der Bewertung als allgemeiner und alltäglicher Prozess nimmt im Schulkontext eine besondere Stellung ein, da seitens Lehrpersonen Unmengen an

² Eine Sonderform stellt das *self-assessment*, Autoevaluation, dar, wobei eine Selbstreflexion eines Individuums zum eigenen Lernverhalten sowie -fortschritt stattfindet (vgl. WEEDEN/WINTER/BROADFOOT 2002: 73). Näheres dazu: Verweis auf Kapitel 1.1.2 dieser Arbeit.

komplexen sozialen Interaktionen im Unterrichtsgeschehen verarbeitet werden müssen, damit eine darauf bezugnehmende adäquate Handlung stattfinden kann (vgl. KLEBER 1992: 78 f). Im Folgenden werden mit Bewertungen in Verbindung stehende Begrifflichkeiten geklärt, voneinander abgegrenzt sowie miteinander in Beziehung gesetzt, damit eine professionelle Auseinandersetzung mit dieser Thematik im Kontext Schule möglich wird.

1.1 Assessment im Fremdsprachenunterricht

BROWN und ABEYWICKRAMA definieren im Glossar ihres Werkes ‚*language assessment*‘ die Begrifflichkeit des Assessments als „*an ongoing process of collecting information about a given object of interest according to procedures that are systematic and substantively grounded*“ (BROWN/ABEYWICKRAMA 2010: 346). Der Bereich des Assessments ist in das gesamte Unterrichtsgeschehen eingegliedert. Innerhalb dieser Disziplin sind weitere untereinander interagierende Subbereiche, wie das Measurement, das Testing sowie die Evaluation, einzuordnen und abzugrenzen, wie aus der folgenden Grafik ersichtlich ist (vgl. BROWN/ABEYWICKRAMA 2010: 6). Diese einzelnen in direktem Zusammenhang mit dem Assessment stehenden Gebiete werden folgend präziser definiert und abgegrenzt sowie untereinander in Verbindung gesetzt.

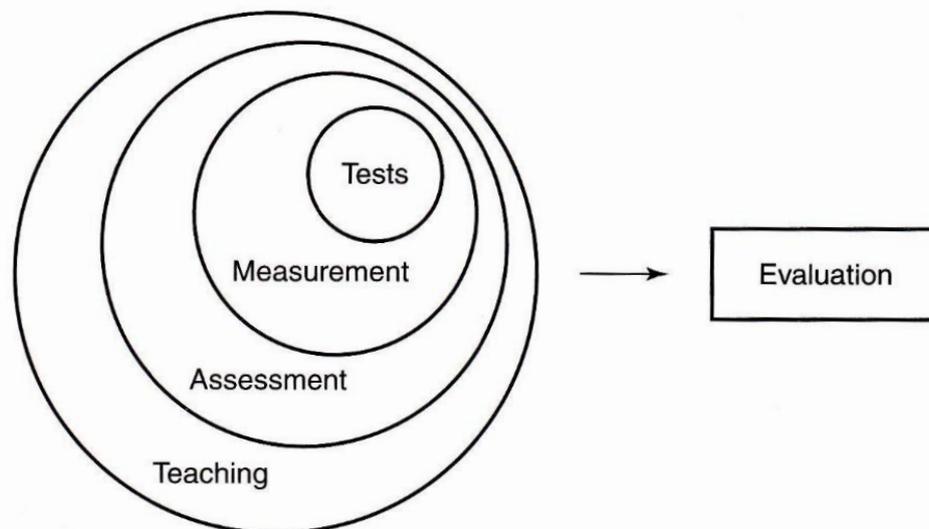


Abbildung 1 (aus: BROWN/ABEYWICKRAMA 2010: 6)

Zunächst soll allerdings der Frage nachgegangen werden, weshalb dem Assessment eine essentielle Rolle im Unterrichtsgeschehen beigemessen wird. In Bezug auf den

schulischen Kontext wird der Bewertung in erster Linie eine einzige Funktion zugeschrieben: die Überprüfung von Kompetenzen. Hinsichtlich des Fremdsprachenunterrichts wird dies meist präzisiert als Überprüfung der sprachlichen Kompetenzen (vgl. GROTHJAHN/KLEPPIN 2015: 12). Diesem allgemeinen, reduzierenden Verständnis von Assessment im Fremdsprachenunterricht gilt es entgegenzuwirken und die mannigfaltigen Funktionen, die dieser Bereich erfüllt, aufzuzeigen, wie es im nachstehenden Unterkapitel erfolgt.

1.1.1 Funktionen

Das Assessment erfüllt eine Vielzahl an wichtigen Aufgaben im Unterrichtsgeschehen, auch wenn diese nicht immer wahrgenommen werden. So tragen auch die im Folgenden genannten bedeutenden Funktionen dazu bei, dass schulisches Assessment hohen Qualitätsansprüchen genügen muss (vgl. GROTHJAHN 2010: 214).

Prinzipiell können die Funktionen des Assessments in drei übergeordnete Bereiche gegliedert werden, je nach Perspektive der Beteiligten. Zu unterscheiden sind die Perspektive der Lehrenden, jene der Lernenden sowie die bildungspolitische Perspektive (vgl. GROTHJAHN/KLEPPIN 2015: 12).

In Bezugnahme auf die Bildungspolitik leistet das Assessment einen wichtigen Beitrag zur Überprüfung bildungspolitischer Zielsetzungen, beispielsweise die Erreichung gewisser formulierter Standards, um in Folge realistische Lernziele in Curricula aufstellen zu können. Weiters dient es einem Vergleich von Bildungseinheiten, seien dies einzelne Schulstandorte, Regionen einer Nation oder gar internationale Vergleiche (vgl. GROTHJAHN/KLEPPIN 2015: 13).

Für Lernende besteht Interesse an der Information über den eigenen Leistungsstand sowie an der Kenntnisnahme über die persönlichen Lernfortschritte. Dies kann durch Autoevaluationen, die den Lernenden zugänglich sind, ermöglicht werden. Außerdem ist es von Bedeutung, dass sie diagnostische und reflexive Kompetenzen entwickeln, welche eine Voraussetzung für erfolgreiches autonomes Lernen darstellen. Letztlich kann als bedeutendes Interesse an Bewertungen aus Perspektive der Lernenden auch die Erlangung eines Zertifikats genannt werden, um persönliche Kompetenzen offiziell zu bestätigen (vgl. GROTHJAHN/KLEPPIN 2015: 13).

Eine Vielzahl an Funktionen erfüllt das Assessment für die Lehrenden, weshalb es als ein stetig stattfindender Prozess im Unterrichtsgeschehen anzusehen ist, der eine weite

Bandbreite an methodischen Umsetzungen beinhaltet und selbst unbewusst Eindrücke aus Perspektive der Lehrenden nach sich zieht (vgl. BROWN/ABEYWICKRAMA ²2010: 3).

Folgende sind häufige Funktionen, die dem Assessment zugeschrieben werden können:

Eine der Hauptfunktionen repräsentiert die Diagnose individueller Stärken und Schwächen der Lernenden, welcher eine Einstufung und Klassifizierung, beispielsweise hinsichtlich des Kompetenzniveaus, folgen kann. Zu dieser diagnostischen Funktion zählen außerdem die Feststellung des Erreichens von Lernzielen, Minimalstandards oder anderen curricularen Vorgaben sowie die damit einhergehende Vergabe von Noten und die individuelle Förderung durch besondere Unterstützungsmaßnahmen wie auch die Beobachtung von Fortschritten einzelner Personen oder Gruppen, oftmals in einem Vergleich Lernender untereinander oder mit anderen Gruppen, wie Parallelklassen.³

Darüber hinaus findet eine Rückmeldefunktion statt, sowohl an Lehrende, als auch an Lernende und weitere Interessierte, wobei diese über den Lernstand informiert werden, wohingegen Lehrpersonen aus Ergebnissen Rückschlüsse auf ihren Unterricht ziehen können und eben jenen dahingehend reflektieren und überarbeiten. Diese Auswirkung wird als Washback-Effekt⁴ bezeichnet und übt eine wichtige Funktion hinsichtlich der Unterrichtsevaluation aus.

Die Berechtigungsfunktion ist vor allem für Zulassungen von Bedeutung. Als solche differenziert sie Fremdsprachenlernende hinsichtlich bestimmter Ergebnisse und erfüllt eine Selektionsfunktion.

Im Schulkontext kommt dem Assessment ebenfalls eine erzieherische Funktion zu, da mit Hilfe von Evaluationen einerseits die Motivation verstärkt, andererseits die Disziplin erhöht werden kann (vgl. GROTHJAHN/KLEPPIN 2015: 12 f).

³ Der Vergleich einer Leistung mit jener anderer Lernender wird noch genauer in Unterkapitel 2.3 dieser Arbeit behandelt.

⁴ Der Washback-Effekt, oder auch Backwash-Effekt genannt, bezeichnet den Einfluss sowie die Wirkung von Tests und anderen Assessmentmethoden auf das Lernen und den Unterricht allgemein. Dies beinhaltet einerseits sowohl beabsichtigte als auch unbeabsichtigte Auswirkungen sowie andererseits auch positive und negative Einflüsse auf den Schulalltag. Festgestellt wurde, dass jedenfalls Folgen in Bezug auf den Lerninhalt zu konstatieren sind, wobei Uneinigkeit herrscht, ob der Effekt auch die Methodologie des Unterrichts und Aufgabenformate nachhaltig beeinflusst, wovon allerdings einige ForscherInnen (beispielsweise ALDERSON/HAMP-LYONS 1996, STATCHER/CHUN/BARRON 2004) ausgehen (vgl. BARNES 20016: 158f). Näheres zu diesem Rückwirkungseffekt siehe:

WALL, Diane (2012): Washback. In: FULCHER, Glenn/DAVIDSON, Fred (Hrsg.): The Routledge Handbook of Language Testing. S. 79 - 92. London [u.a.]: Routledge.

„Die unterschiedlichen Funktionen der Leistungsbewertung lassen sich schwerlich unter einen Hut bringen. Es ist ausgeschlossen, den Lernenden eine helfende Hand entgegenzustrecken, ihnen Tipps zu geben und sie andererseits (wenn auch einige Zeit später) mit einer Abstufung zu bedrohen, falls sie das Lernziel verfehlen.“ (STERN 2010: 31)

Diese Vielfalt an Funktionen kann das Assessment durch verschiedenartige Formen erfüllen, die im nächsten Unterkapitel erläutert werden, wobei folgend keinerlei Anspruch auf Ganzheitlichkeit besteht, da die Erläuterung die bedeutendsten Formen beinhaltet.

1.1.2 Formen

Zunächst lässt sich das Assessment hinsichtlich seines Planungscharakters und seiner Intention in ein informelles und formelles Assessment unterteilen. Das informelle Assessment kann in einer Vielzahl von Arten auftreten und umfasst beispielsweise spontanes, beiläufiges, ungeplantes Nicken, Kommentieren sowie sonstige Möglichkeiten des Feedbacks.⁵ Abgegrenzt davon findet das formelle Assessment stets systematisch geplant statt, um der Lehrperson oder den Lernenden Auskunft über den Lernfortschritt beziehungsweise Leistungsstand zu geben (vgl. BROWN/ABEYWICKRAMA ²2010: 6 f).

Ebenso wird zwischen direktem und indirektem Assessment unterschieden, wobei ein Lerninhalt entweder direkt abgefragt wird und die Lernenden genau jene Kompetenzen und Handlungen unter Beweis stellen, die bewertet werden wollen, oder indirekt überprüft wird, indem bei einer Aufgabe sämtliche Fertigkeiten, die der abzurufenden Fähigkeit zugrunde liegen, integrativ angewandt werden müssen und so einer Bewertung unterliegen (vgl. HUGHES ²2003: 17 ff).

⁵ Im allgemeinen und generellen Verständnis wird Feedback als eine Rückmeldung verstanden (vgl. BROOKHART (2010: 7). In der Grundform handelt es sich demnach um einen Regelkreis, bei dem Objekt A einen Impuls an Objekt B sendet und B an A eine Rückmeldung tätigt (vgl. BUHREN 2015: 11). Im Gegensatz zu den technischen Regelungsprozessen unterliegt das Feedback im zwischenmenschlichen Bereich keiner objektivierbaren Festlegung, sondern verhält sich variabel abhängig von den betroffenen Interaktionspersonen sowie der Situation (vgl. FENGLER ³2004: 12). Feedback kann wie auch das Assessment in verschiedenen Formen auftreten, intern sowie extern, verbal oder nonverbal sowie intentional beziehungsweise nicht intentional (vgl. SALZMANN 2015: 23). BUHREN hält fest, dass Feedback und Evaluation nahe beieinander liegende Konzepte sind, die es dennoch zu unterscheiden gilt, denn Evaluation entspricht den Kategorien von Reliabilität und Validität, wohingegen dies bei Feedback nicht zwangsweise Voraussetzung ist, da es sich oftmals um eine direkte und unmittelbare Rückmeldung handelt, die verbal, spontan und möglicherweise auch unsystematisch geäußert wird. Noch hinzu kommt, dass Evaluation auf bestimmte Bewertungskategorien und Bewertungsmaßstäbe zurückgreift, wohingegen Feedback nicht vorrangig als Bewertungsinstrument zu sehen ist (vgl. BUHREN 2015: 16 f).

In einer weiteren Instanz findet eine Unterteilung in Bezug auf die Funktion und Richtungswirkung statt. Hierbei wird zwischen formativen und summativen Bewertungen unterschieden. Formativ bedeutet, dass der Prozess der Kompetenzaneignung und -erweiterung im Fokus steht. Informelles Assessment im Unterrichtsgeschehen unterstützt eben diese formative Herangehensweise, da der Fortschritt des Lerner beziehungsweise der Lernerin im Vordergrund steht und dieser Entwicklungsprozess durch die Lehrperson unterstützt wird, indem Fehlerhaftes korrigiert und Positives hervorgehoben wird. Im Gegensatz dazu zielt summatives Assessment auf eine abschließende Messung der Lernziele ab, beispielsweise am Ende einer Lerneinheit oder eines Kurses. Diesbezüglich wird kein Prozess berücksichtigt, sondern rein die momentane erbrachte Leistung und ist aus diesem Grund als produktorientiert zu bezeichnen. Aus summativer Bewertung folgt häufig eine Evaluation⁶, also eine Entscheidung auf Basis einer Urteilsbildung, was allerdings nicht zwingend notwendig ist (vgl. BROWN/ABEYWICKRAMA²2010: 7).

Ähnlich wie die vorangegangene Dichotomie werden kontinuierliche und punktuelle Beurteilung gesondert betrachtet. Während bei der Kontinuität eine temporäre Gesamtheit eines Prozesses von der Bewertung berücksichtigt wird, meint punktuell lediglich die Beurteilung einer Aufgabe in einem bestimmten Moment (vgl. Europarat 2001: 180).

Bewertungen werden im österreichischen Schulkontext zumeist intern getroffen, also von den Lehrpersonen, den Unterrichtenden, selbst. Eine Alternative stellt das externe Assessment dar, welches allerdings informelle und formative Bewertungen ausschließt und nur summative Beurteilungen zulässt. Externes Assessment würde allerdings das Dilemma der Leistungsbewertung für Lehrpersonen verringern, welches vorhergehend bei den Funktionen aufgezeigt wurde (vgl. STERN 2010: 31):

„Deswegen plädieren viele Expert/inn/en für externe Tests, was die Lehrer/innen von ihrer Doppelrolle als „Trainer/innen“ und „Schiedsrichter/innen“ befreien würde. [...] Andererseits haben externe Tests auch Nachteile. Erstens sind sie nur punktuell, während Lehrer/innen sich bei ihren Bewertungen auf wesentlich vielfältigere Befunde stützen können. Zweitens lösen sie, vor allem, wenn viel von ihnen abhängt [...], einen „Backwash-Effekt“ aus, d. h. Lehrer/innen sehen sich genötigt, auf wichtige Lernimpulse zu verzichten, wenn diese nicht auch zur Verbesserung der Testergebnisse beitragen.“ (STERN 2010: 31)

⁶ Der Terminus Evaluation wird in Unterkapitel 1.3 dieser Arbeit näher betrachtet und erläutert.

Weiters muss eine Bewertung im Schulkontext nicht immer von einer Lehrperson ausgehen, denn neben der Fremdevaluation, sei es die Lehrperson oder eine externe Institution wie bei Zertifikatsprüfungen, ist es Lernenden möglich, eine Selbstevaluation durchzuführen, um Auskunft über den persönlichen Lernstand zu erhalten oder den weiteren Lernprozess zu planen sowie eine Reflexion darüber auszulösen. Ferner sind in LernerInnengruppen Peer-Evaluationen durchführbar, deren Hauptfunktion vor allem darin zu sehen ist, dass gleich gestellte Lernende sich gegenseitig unterstützen ihre Sprachkompetenzen zu reflektieren, weshalb sie einen bedeutenden Prozess für selbstreflexives Lernen darstellen (vgl. GROTH/JAHN/KLEPPIN 2015: 37 f). Im Fremdsprachenunterricht werden hierfür Sprachenportfolios⁷ herangezogen, um die bisher erreichten Kompetenzen dokumentieren zu können (vgl. FÄCKE 2010: 230). Andere Techniken stellen beispielsweise Lerntagebücher und Autotests dar. Die Selbstevaluation ist als vergleichsweise neue Methode im Schulkontext zu sehen und wuchs aus der Orientierung an qualitativen Methoden heraus. Grundlegend basiert sie auf dem Verständnis, dass effektives Lernen auf einer aktiven Einbindung der Lernenden in den gesamten Prozess basiert (vgl. OSCARSON 1997: 175).

Eine weitere Dichotomie wird betreffend der Referenzbezüge⁸ getroffen. In diesem Fall wird das normorientierte Assessment vom kriteriumsorientierten abgegrenzt. Hinsichtlich ersterem werden Ergebnisse von Lernenden in Relation zu denen anderer gesetzt und in eine gewisse Rangordnung eingestuft, wodurch nicht genau hervorgeht welche Kompetenzen erfüllt werden, sondern lediglich ein Vergleich zu anderen Personen gezogen wird (vgl. WAUGH/GRONLUND¹⁰2013: 26). Bei kriterienorientiertem Assessment steht die Bewertung der Fähigkeiten, gemessen an bestimmten, im Vorhinein festgesetzten Kriterien beziehungsweise Standards, im Vordergrund (vgl. BROWN/ABEYWICKRAMA²2010: 8).

⁷ Sprachenportfolios begleiten und dokumentieren einen Sprachlernprozess, indem sie eine Vorzeigefunktion, eine Reflexionsfunktion sowie eine Planungsfunktion erfüllen. Sie bestehen meist aus einem Sprachenpass, einer Sprachenbiografie und einem Dossier mit gesammelten Arbeiten (vgl. BURWITZ-MELZER 2010: 232 ff).

⁸ Hinsichtlich zulässiger Bezugssysteme im österreichischen Schulwesen wird in diesem Kontext auf Unterkapitel 2.3 dieser Arbeit verwiesen.

„The purpose of criterion-referenced tests is to classify people according to whether or not they are able to perform some task or set of tasks satisfactorily. The tasks are set, and [...] those who perform them satisfactorily ‘pass’; those who don’t, ‘fail’. This means that students are encouraged to measure their progress in relation to meaningful criteria [...]. Criterion-referenced tests therefore have two positive virtues: they set meaningful standards in terms of what people can do, which do not change with different groups of candidates, and they motivate students to attain those standards. (HUGHES 2003: 21).“

Beurteilungen sind ebenso zu unterscheiden hinsichtlich der Aspekte, die sie berücksichtigen. Entweder handelt es sich um eine ganzheitliche, holistische Beurteilung, welche eine globale, synthetische Einschätzung zur Folge hat, beispielsweise bezüglich der schriftlichen Sprachproduktion allgemein, oder analytisches Beurteilen, wobei einzelne Aspekte getrennt betrachtet und bewertet werden (vgl. EUROPARAT 2001: 185).

Auf einer letzten Ebene ist noch eine Unterteilung in subjektive und objektive Bewertungsverfahren zu treffen. Von objektivem Assessment wird gesprochen, solange nur eine vorgegebene und unveränderbare richtige Antwort vorliegt, wohingegen bei subjektiven Bewertungsmethoden dies nicht der Fall ist, sondern die Beantwortung einem Urteil der Lehrperson bedarf (vgl. BROWN/ABEYWICKRAMA 2010: 351 ff).

„If no judgement is required on the part of the scorer, then the scoring is *objective*. A multiple choice test, with the correct responses unambiguously identified, would be a case in point. If judgement is called for, the scoring is said to be *subjective*. There are different degrees of subjectivity in testing. The impressionistic scoring of a composition may be considered more subjective than the scoring of short answers in response to questions on a reading passage. (HUGHES 2003: 22).“

Um eine möglichst hohe Objektivität gewährleisten zu können, werden Leistungen anhand bestimmter Kriterien gemessen und quantifiziert (vgl. BROWN/ABEYWICKRAMA 2010: 350), wie im folgenden Subbereich des Assessments aufgezeigt wird.

1.2 Measurement und Testing

Wie aus Abbildung 1 im Unterkapitel 1.1 hervorgeht, ist in den Bereich des Assessments das Measurement eingebettet, worin wiederum das Testen eingegliedert ist. Im deutschen Sprachgebrauch werden im Alltagsverständnis eine Vielzahl an Begrifflichkeiten, wie Bewertung, Beurteilung, Diagnose, Evaluation und weitere, die zwar in engem Verhältnis zueinander stehen aber sich dennoch voneinander unterscheiden, synonym verwendet

(vgl. KLEBER 1992: 77), weshalb der Terminus *technicus* des Measurements in dieser Arbeit beibehalten wird und einer Abgrenzung bedarf.

Zu definieren ist dieser als „*a process of quantifying a test-taker's performance according to explicit procedures of rules*“ (BROWN/ABEYWICKRAMA ²2010: 350). Im Fokus stehen also eine Quantifizierung und eine Messung von Leistungen, die bestimmten Regeln, sprich Gütekriterien, folgt, weswegen von Bewertung durch Messung gesprochen wird (vgl. <http://dictionary.cambridge.org/dictionary/english/measurement>).

Wie schon das Measurement eine Subform des Assessments darstellt (vgl. Abbildung 1), so sind Tests als untergeordnete Technik oder Methode in diesem Verfahren anzusehen, denn ein Test ist „*a method or procedure for measuring a person's ability, knowledge, or performance in a given domain*“ (BROWN/ABEYWICKRAMA ²2010: 353). Assessmentinstrumente wie Tests können in verschiedene Arten klassifiziert werden, je nachdem welchen Zweck sie verfolgen. Dabei können sie den unterschiedlichen Formen des Assessments unterliegen (vgl. LIENERT/RATZ ⁶1998: 14 f).

1.2.1 Testarten

Sogenannte *Achievement Tests* erheben die Fähigkeiten eines Schülers beziehungsweise einer Schülerin innerhalb eines bestimmten Bereiches und zielen darauf ab, ob gewisse Lernziele erreicht wurden oder nicht, woraus auch eine gewisse diagnostische Funktion resultieren kann und anzeigt woran noch gearbeitet werden muss. Solche Leistungstests unterliegen zumeist dem summativen Assessment, da sie zum Abschluss einer Lektion stattfinden, nehmen aber dennoch eine wichtige formative Rolle ein, wenn ein Test Feedback anbietet. Sie sollten allerdings nicht mit diagnostischen Tests gleichgesetzt werden, wie im weiteren Verlauf aufgeklärt wird (vgl. BROWN/ABEYWICKRAMA ²2010: 9).

Proficiency Tests zielen hingegen darauf ab, die Globalkompetenz in einer Sprache zu bewerten, weshalb sie von einzelnen Kursen oder Curricula losgelöst sind. Das Messen umfassender und übergreifender Sprachkompetenzen steht im Vordergrund dieser meist standardisierten Tests, die selten darauf ausgelegt sind, Feedback zu erteilen, sondern lediglich eine erbrachte Leistungsfähigkeit bescheinigen (vgl. HUGHES ²2003: 11 f).

Die beiden bereits erwähnten Testarten können im Sinne eines *Placement Tests* eingesetzt werden, welche Lernende einem bestimmten Level oder einer Gruppe zuordnen sollen. Dabei findet stets eine Orientierung an Referenzen, die einzelne Gruppen oder Niveaus

beschreiben, beziehungsweise an Fähigkeiten, die diese beherrschen müssen, statt (vgl. BROWN/ABEYWICKRAMA ²2010: 10 f).

Progress Tests finden Anwendung innerhalb eines Sprachlernprozesses, um die Fortschritte bewerten zu können. Sie ähneln den beschriebenen Leistungstests, welche allerdings am Ende einer Lektion, Einheit oder eines Kurses stattfinden, wohingegen *Progress Tests* in unterschiedlichen Stadien des Lernprozesses stattfinden (vgl. ALDERSON/CLAPHAM/WALL 1995: 11 ff).

Aptitude Tests, die eine generelle Eignung zu fremdsprachlichem Lernen messen sollen, sind kaum mehr üblich, denn eine im Vorhinein grundsätzliche Einschätzung abzugeben, welche Leistung eine Person erbringen werden kann, ist nicht haltbar (vgl. BROWN/ABEYWICKRAMA ²2010: 11 f).

„Instead, attempts to measure language aptitude more often provide learners with information about their preferred styles and their potential strengths and overcoming the weaknesses [...]. Any test that claims to predict succes in learning a language is undoubtedly flawed, because we now know that with appropriate self-knowledge , active strategic involvement in learning, and/or strategies-based instruction, virtually everyone can eventually succeed (BROWN/ABEYWICKRAMA ²2010: 12).“

Ziel diagnostischer Tests ist es, Stärken und Schwächen von Sprachlernenden auszumachen und positive Änderungen für das Lehren und Lernen vorzunehmen. Trotz häufiger Umsetzung dieses positiven Lerneffekts durch formative Methoden in der Praxis, darf umgekehrt nicht davon ausgegangen werden, dass formatives Assessment stets diagnostisch sei (vgl. JANG 2012: 120 f).

„Diagnostic assessment focuses on the common sources of error encountered by students, so that the learning difficulties can be pinpointed and remedied. [...] Diagnosing learning problems is a matter of degree. Formative assessment determines whether a student has mastered the learning tasks being taught and, if not, prescribes how to remedy the laearning failures. Diagnostic assessment is designed to probe deeper into the causes of learning deficiencies that are left unresolved by formative assessment. (WAUGH/GRONLUND ¹⁰2013: 7f).“

Im Schulkontext finden hauptsächlich Leistungstests, *progress tests* und diagnostisches Assessment Anwendung, zwischen denen häufig die Grenzen verschwimmen, wobei die Hauptfunktion ersterer das Aufzeigen des Ausmaßes an Beherrschung des bisher Gelehrten ist, wohingegen Diagnosen aufdecken sollen, woran in Zukunft noch verstärkt gearbeitet werden muss (vgl. BROWN/ABEYWICKRAMA ²2010: 10).

1.2.2 Gütekriterien

Testgütekriterien sind erforderlich, um Qualitätsbeurteilung sowie Wissenschaftlichkeit zu gewährleisten. Hierfür sind drei Hauptgütekriterien, Objektivität, Reliabilität und Validität, zu unterscheiden (vgl. POSPESCHILL 2010: 17).

Während unter Objektivität der Grad an Unabhängigkeit von jeglichen Beeinflussungen, wie Situationsbezug, durchführende Person, Auswertung und weitere, verstanden wird, bezieht sich die Reliabilität auf die Messgenauigkeit im Sinne einer Zuverlässigkeit des Tests (vgl. ROST ²2004: 33). Das Kriterium der Validität gibt den Grad der Gültigkeit hinsichtlich der Aussagefähigkeit an und setzt voraus, dass ein Test misst, was er zu messen behauptet (vgl. CHAPELLE 2012: 21).

Neben diesen Hauptkriterien werden weitere Nebenkriterien formuliert, welche die Bereiche Normierung, Skalierung, Wirtschaftlichkeit, Nützlichkeit, Zumutbarkeit, Unverfälschbarkeit, Fairness, Vergleichbarkeit und Transparenz umfassen (vgl. POSPESCHILL 2010: 17). Zusätzlich werden abseits der klassischen Testtheorie in Bezug auf den Fremdsprachenunterricht weitere spezifische Kriterien definiert, die die Authentizität, den Rückwirkungseffekt, die Schwierigkeit von Aufgaben, die Trennschärfe dieser sowie ihre Unabhängigkeit voneinander und die Handlungsorientierung umfassen (vgl. CASPARI/GROTJAHN/KLEPPIN 2010: 60), wodurch eine Wandlung von psychometrischen zu kommunikativen Testaufgaben im Fremdsprachenunterricht auszumachen ist (vgl. EYNAR 2010: 417).

Die Messung von Leistung ist somit als Subform und integraler Bestandteil des Assessments anzusehen, da Tests eine gewisse Technik in diesem Prozess der Einschätzung darstellen und deshalb zusammenfassend als Bewertungsmethode zu bezeichnen sind. Aus den Ergebnissen dieser Tests resultierende Entscheidungen und Urteilsbildung unterliegen dem Bereich der Evaluation (vgl. BROWN/ABEYWICKRAMA ²2010: 3 ff), der im folgenden Verlauf dieses Kapitels genauer untersucht werden soll, um die Grundlage zur Annäherung der Leistungsbeurteilung im österreichischen Schulsystem zu schaffen.

1.3 Evaluation

In gewisser Abgrenzung zum Assessment wird Evaluation im Glossar von BROWN und ABEYWICKRAMA mit folgender Begriffserläuterung definiert: „*making decisions and/or value judgements based on the results of tests, other assessments, and/or teacher's reports*“ (BROWN/ABEYWICKRAMA ²2010: 348). Demnach steht sie in einem besonderen Zusammenhang mit der Entscheidung und einer Urteilsbildung, die auf einer Bewertung basiert. Diese Basis bilden unterschiedliche Arten von Tests, in denen bestimmte Kriterien gemessen werden sollen (vgl. BROWN/ABEYWICKRAMA ²2010: 3 ff).

„Exams are the important assessment tools that are utilized at every phase of the education system. Assessment is digitalization of the qualifications, expressing the observed qualifications via numbers and symbols. Evaluation, on the other hand, is a decision making proecess relating to the assessed qualification, by comparing the results obtained from the assessment process with certain criteria. (KARA/ÇELIKLER 2015: 21).“

Während der übergeordnete Terminus Assessment den gesamten Prozess der Bewertung an sich meint, wird in Bezug auf eine einhergehende Urteilsbildung und Beurteilung, im Sinne einer Diagnose, vom Teilbereich der Evaluation gesprochen. Diese ist untrennbar mit dem Fachgebiet des Assessments verbunden (vgl. BROWN/ABEYWICKRAMA ²2010: 3 ff), wie auch aus Abbildung 1 hervorgeht. Evaluationsergebnisse rufen in ihrer Folge ein Urteil hervor. Eben solche Werturteile, die nach Einschätzungen gebildet werden, können den unterschiedlichen Formen des Assessments unterliegen. So besteht die Möglichkeit formeller oder informeller, subjektiver oder objektiver sowie fremder oder eigener Urteile, abhängig der zugrundeliegenden Bewertungsinstrumentarien und Assessmentmethoden (vgl. STOCKMANN/MEYER 2010: 17).

Als weitläufiger Terminus, der in unterschiedlichen Disziplinen zu verankern ist und aus mehreren Zugängen erschlossen werden kann, liegen vielfältige der jeweiligen Fachwissenschaft zugewandte Definitionen vor (vgl. WOTTAWA/THIERAU ³2003: 13), weshalb laut Duden Evaluation allgemein als „sach- und fachgerechte Bewertung“ (DUDENREDAKTION ⁶2007b: 531) definiert wird. Die Evaluation und die damit einhergehende Forschung sind also der zutreffenden Disziplin entsprechend anzupassen. Eine übereinstimmende Auffassung herrscht darüber, dass jeglicher Evaluation in ihrem Grundschema drei Aspekte zugrunde liegen, weswegen festgehalten werden kann,

dass „Evaluation ein *Instrument zur empirischen Generierung von Wissen* ist, das mit einer *Bewertung verknüpft* wird, um *zielgerichtete Entscheidungen* zu treffen (STOCKMANN/MEYER 2010: 64).“

Die Evaluationsforschung erhebt als praxis- und anwendungsbezogene Disziplin einen Anspruch auf Problemlösungsfähigkeit, wofür sie auf das umfassende Repertoire an sozialwissenschaftlichen Theorien, Konzepten und Methoden zurückgreift, um diese im Sinne der wissenschaftlichen Praxis zu Forschungszwecken heranzuziehen. Neben dieser Problemlösungskompetenz ist ihr ferner aufgrund der Zweckorientierung eine Nützlichkeit zuzuschreiben, da sich Evaluation an konkret vorgegebenen Zielen orientiert und diese verfolgt (vgl. STOCKMANN/MEYER 2010: 57).

Der Gestaltungsaspekt nimmt in der Evaluationsforschung eine bedeutende Rolle ein, da Vorbedingungen zu definieren sind. Hinsichtlich der Evaluation ist zu beachten, um welches Evaluationsobjekt, in welchem Evaluationsbereich es sich handelt, das anhand welcher Methoden und Modelle, zu welchem Zwecke, von wem evaluiert werden soll. Die durchzuführende Evaluation muss auf jene Bedingungen angepasst und abgestimmt sein (vgl. WOTTAWA/THIERAU ³2003: 55 f), wie es CUQ und GRUCA im Folgenden hinsichtlich schulischer Evaluation im Fremdsprachenunterricht zusammenfassen:

„Évaluer signifie déterminer précisément ce qu'on veut évaluer et mesurer, c'est-à-dire définir au préalable les objectifs, les situations et les tâches, dresser un inventaire d'actes de paroles, définir les niveaux à atteindre ou une hiérarchie de niveaux, définir les critères d'évaluation et les critères de passation des tests, etc. L'évaluation d'une langue est donc un processus complexe, d'autant plus qu'il faut évaluer la compétence des apprenants non seulement en fonction de la maîtrise du code linguistique et des savoir-faire communicatifs, mais aussi en fonction d'une composante socio-culturelle et des composantes disciplinaires. Mais, l'évaluation est avant tout un acte formatif, qui se situe dans le continuum de l'apprentissage et qui apporte une contribution importante à l'appropriation d'une langue étrangère. Évaluer ne doit pas être une fin, mais un moyen.“ (CUQ/GRUCA 2005: 219)

1.4 Herausforderungen des Fremdsprachenunterrichts

„Die Grenzen meiner Sprache sind die Grenzen meiner Welt.“
(WITTGENSTEIN, zitiert nach DECKE-CORNILL/KÜSTER³2015: 18)

Wie aus dem im vorherigen Unterkapitel angeführten Zitat von CUQ und GRUCA zu entnehmen ist, stellt das Assessment im Fremdsprachenunterricht einen außerordentlich komplexen Prozess dar, denn Gegenstand ist die Sprache in all ihren Facetten und die mit ihr in Zusammenhang stehenden Faktoren, die für die Befähigung, in kommunikativen Situationen handlungsorientiert agieren zu können, nötig sind. Lernen an sich ist als komplexes Ereignis zu verstehen, wobei das Erlernen einer Sprache eine besondere Herausforderung darstellt. Im Bereich dieser Arbeit wird vom systematischen, gesteuerten Erlernen einer Fremdsprache⁹ im schulischen Kontext ausgegangen. Da Sprache als menschliches Agieren zu verstehen ist und viele unterschiedliche Funktionen in Interaktionssituationen erfüllt, ist sie als Instrument für funktionales, kommunikatives Handeln¹⁰ anzusehen, wodurch eine Reihe an Fertigkeiten benötigt wird (vgl. BUDDE/RIEGLER/WIPRÄCHTIGER-GEPPERT²2012: 29 ff), um dies im Sinne des Lehrplans umzusetzen, denn oberstes Ziel des Fremdsprachenunterrichts ist demnach „situationsadäquates Kommunizieren“, welches in allen Fertigkeiten, mündlich und schriftlich sowie rezeptiv und produktiv, stattfinden soll (vgl. Lehrplan HAK 2014: 2 ff).

„Kommunikative Aufgaben sind ein Merkmal des alltäglichen Lebens im privaten, öffentlichen und beruflichen sowie im Bildungsbereich. Die Bewältigung [...] beinhaltet die strategische Aktivierung spezieller Kompetenzen, um innerhalb eines bestimmten Lebensbereichs eine Gruppe zielgerichteter Handlungen mit einem klar definierten Ziel und einem speziellen Ergebnis auszuführen.“ (EUROPARAT 2001: 153)

⁹ Das komplexe Konstrukt Sprache kann hier nicht ausführlich erläutert werden, ist aber als ein zur Kommunikation verwendetes Zeichensystem zu beschreiben. Unter einer Fremdsprache wird grob umrissen verstanden, dass diese, in Abgrenzung zur Erst- sowie Zweitsprache, gesteuert, explizit und bewusst gelernt wird und keine zentrale Bedeutung für die Bewältigung des Alltagslebens einnimmt (vgl. EDMONDSON/HOUSE³2006: 7 ff).

¹⁰ Die aus der Reformpädagogik stammende Handlungsorientierung bezieht sich auf die Bedeutung der Äußerungsfähigkeit der Lernenden, um in möglichst authentischen Kommunikationssituationen die Zielsprache für sprachliches Handeln zweckmäßig nutzen zu können (vgl. CASPARI 2008: 26).

Diese Fertigkeiten¹¹, deren sich SprachbenutzerInnen in Kommunikationssituationen bedienen, umfassen eine Vielfalt an unterschiedlichen Einordnungen, da nicht ausschließlich sprachliche und damit in direktem Zusammenhang stehende, sondern in einer gewissen Weise alle menschlichen Kompetenzen zur Kommunikationsfähigkeit beitragen, weshalb in einer ersten Instanz ganz generell zwischen allgemeinen und sprachlichen Kompetenzen unterschieden wird. Weiters lassen sich diese allgemeinen Kompetenzen erneut in vier Unterkategorien differenzieren. Hierbei ist zunächst das deklarative Wissen (*savoir*), das von Sprachlernenden kumulativ angehäuft wird, anzuführen, welches sich aus allgemeinem Weltwissen verknüpft mit soziokulturellem Wissen und interkulturellem Bewusstsein zusammensetzt. Abzugrenzen ist hiervon das prozedurale Wissen (*savoir-faire*), wobei das Hauptaugenmerk auf die praktische Umsetzung gelegt wird und somit praktische Fertigkeiten in den Vordergrund rücken (vgl. EUROPARAT 2001: 103 ff).

„Generell können wir sagen: **Deklaratives Sprachwissen** ist eher auf stabile Gegebenheiten bezogen (*know what* [...]), während **prozedurales Wissen** auf Vorgänge gerichtet ist (*know how*, eher ein „Können“ wie z.B. das kontextuelle Erschließen [...]).“ (DECKE-CORNILL/KÜSTER ³2015: 44)

Persönlichkeitsbezogene Kompetenzen (*savoir-être*), wie Motivation, Werte und Überzeugungen oder weitere Persönlichkeitsfaktoren, und die Lernfähigkeit (*savoir-apprendre*), zu der Sprach- und Kommunikationsbewusstsein, Lerntechniken sowie heuristische Fertigkeiten zählen, sind weitere allgemeine Kompetenzen, auf die Fremdsprachenlernende zurückgreifen (EUROPARAT 2001: 103 ff).

Neben diesem allgemeinen Wissen, ist das spezifisch sprachliche hervorzuheben, welches als kommunikative Sprachkompetenz zu bezeichnen ist, und in sprachwissenschaftlicher Hinsicht einer weiteren Gliederung in linguistische, soziolinguistische und pragmatische Kompetenzen bedarf (vgl. FÄCKE 2010: 118). Die linguistische Fertigkeit ist zu verstehen als „Kenntnis der formalen Mittel, aus denen wohlgeformte, sinnvolle Mitteilungen zusammengesetzt und formuliert werden können, und als die Fähigkeit, diese Mittel auch zu verwenden.“ (EUROPARAT 2001: 110). Diesbezüglich lässt sich die Unterteilung in

¹¹ Kompetenzen sind vorhandene oder erlernbare kognitive Fähigkeiten und Fertigkeiten, die dazu beitragen gewisse Probleme zu lösen, sowie innere persönliche Fähigkeiten und die Bereitschaft, diese Problemlösungsstrategie der gegebenen Situation anzupassen und erfolgreich und verantwortungsvoll auf die verfügbaren Ressourcen zurückzugreifen (vgl. WEINERT 2001: 27 f).

deklaratives und prozedurales Wissen erkennen, da es sich um die Kenntnis der formalen Mittel handelt aber auch um die Verwendung dieser in den unterschiedlichen Fertigkeiten, wie Hören, Sprechen, Lesen und Schreiben. In diesen linguistischen Bereichen fallen lexikalische, grammatische, semantische sowie phonologische Kompetenzen hinsichtlich der Fremdsprache (vgl. FÄCKE 2010: 118). Alleinig sprachlich formales Wissen ist nicht ausreichend, um erfolgreich kommunizieren zu können, weshalb die soziolinguistische Kompetenz, damit soziale Interaktionen adäquaterweise eingegangen und aufrechterhalten werden können, und die pragmatische Kompetenz, welche die Kommunikation im Sinne der Sprecherabsicht gestaltet, anzufügen sind (vgl. EUROPARAT 2001: 109 ff). Als weitere Besonderheit ist anzuführen, dass Sprache nicht nur Gegenstand des Unterrichts, sondern auch zentrales Medium im Unterrichtsgeschehen ist, weshalb dem eine bedeutende Rolle im Lehr- und Lernprozess beizumessen ist (vgl. FELDER 2003: 42 ff), denn die Unterrichtssituation ist häufig eine der wenigen authentischen Kommunikationssituationen (mögliche Auslandsaufenthalte und [e-]Tandems exklusive) im schulischen Fremdsprachenunterricht (vgl. VOSS³2001: 106).

Sprachenlernen und Sprachgebrauch steht unter einem immensen Einfluss von Emotionen, die häufig eine gewisse Selektionsfunktion im Verarbeitungsprozess erfüllen.

„In allen kommunikativen Kontexten sind selbst dort, wo scheinbar Sachinformationen im Vordergrund stehen, emotionale Anteile konnotiert. [...] Ob im Unterrichtsgeschehen oder in sog. ‚natürlicher‘ Umgebung, immer schwingen emotionale Anteile mit. [...] Für manche Lernende steht gewiss der Gegenstand des Lernens, also die Sprache, die mit ihr verbundenen Kulturen und die mit ihr sich eröffnenden Kommunikationsmöglichkeiten, im Vordergrund, während andere sich eher vom sozialen Kontext des Lernens [...] leiten lassen.“ (DECKE-CORNILL/KÜSTER³2015: 50 f)

Somit kommt der eigenen Person im Fremdsprachenunterricht ein wesentlicher Beitrag zu, wobei das Eigene im Fremdsprachenlernen einem Fremden gegenübersteht, denn ein Erfahren einer anderen Kultur mit bestimmten Wertvorstellungen und Denkweisen findet statt, was eine Herausforderung hinsichtlich der Lerner Identität wird (vgl. BRUSCH 2009: 20 f). Humboldt erläuterte dies bereits folgendermaßen: „Die Erlernung einer fremden Sprache sollte die Gewinnung eines neuen Standpunkts in der bisherigen Weltansicht sein.“ (HUMBOLDT zitiert nach BRUSCH 2009: 20) Im Kontext von Interkulturalität kommt dem Fremdsprachenunterricht hier eine Sonderstellung zu, da der

Erfahrungsbereich der SchülerInnen erweitert wird, wodurch nicht nur das Verstehen sowie Akzeptieren der Fremdheit, sondern ebenso die eigene Fremdheit in der Zielsprache bewusstgemacht wird (vgl. BREUGNOT 2000: 288 f), wodurch es gilt, WITTGENSTEINS Zitat zu Beginn des Unterkapitels kritisch zu betrachten:

„Das Erlernen einer Fremdsprache kann dem Lerner helfen, die Grenzen seiner lebensweltlichen Erfahrung und seiner Erziehung zu überwinden und ihn ermutigen, seine eigene Identität aus zahlreichen Quellen zu speisen. [...], das individuelle Selbst- und Weltverständnis [...] zu entfalten und zu bereichern und gleichzeitig Toleranz gegenüber anderen Menschen zu fördern.“ (BRUSCH 2009: 21)

Allerdings sind nicht nur bereits erwähnte Kompetenzen und interne Faktoren ausschlaggebend für den Fremdsprachenunterricht, sondern in entscheidender Weise ebenfalls äußere Faktoren, angefangen bei Merkmalen und Herausforderungen, die das Klassenzimmer betreffen (vgl. HAB 2010: 156 ff), bis hin zur weitreichenden Faktorenkomplexion des Fremdsprachenunterrichts, wobei eine Vielzahl an äußeren Begebenheiten zu berücksichtigen sind, wie wissenschaftliche Diskurse, Sprachen- und Bildungspolitik und vieles mehr (vgl. FÄCKE 2010: 6 f).

In Bezug auf den Französischunterricht im Spezifischen wird dieser vergleichsweise als herausfordernd eingeschätzt, was auf objektive sprachinhärente Schwierigkeiten zurückzuführen ist, wobei konkrete Probleme hinsichtlich Nasalität, Homophonie und den lautlichen Bereich im Allgemeinen, Tempus-Aspekt-Beziehungen und weitere anzuführen sind, weswegen konstatiert wird, dass für den Französischunterricht ein erhöhter Lernaufwand zu Beginn erforderlich ist, jedoch bei zunehmenden Erfolgserlebnissen die Schwierigkeiten abnehmen (vgl. RAABE ⁴2003: 536 f).

Nach dieser kurzen Darlegung der Herausforderungen im und Besonderheiten des Französischunterrichts, ist die Komplexität des Assessments und des weitreichenden Bereichs, der zu bewerten versucht wird, diesbezüglich grob umrissen, weshalb in einem nächsten Schritt die Schulleistung in den Fokus gerückt wird, bevor die Beurteilung solcher im Französischunterricht behandelt wird.

2 Leistungsbeurteilung im österreichischen Schulsystem

Nachdem in einem ersten Kapitel zunächst allgemein über Bewertung und in weiterer Folge über Assessment im Unterrichtskontext informiert wurde, wobei unterschiedliche Formen und Funktionen sowie die den Unterricht betreffende Testtheorie dargelegt wurden, wird in diesem Kapitel die bereits umrissene Theorie des Beurteilens mit dem österreichischen Schulsystem in Zusammenhang gebracht, weshalb zunächst der Begriff der Leistung einer Definition bedarf, bevor die zulässigen und gängigen Beurteilungsmöglichkeiten präsentiert werden. Hierbei wird bereits in Hinblick auf die Empirie Bezug zu schriftlichen Performanzen genommen, da geschriebene Texte und deren Beurteilung mit Hilfe standardisierter Bewertungshilfen den Hauptfokus dieser Arbeit einnehmen.

2.1 Das Leistungsprinzip

„Es gibt Leistung ohne Erfolg, aber keinen Erfolg ohne Leistung.“
(LA ROCHEFOUCAULD, zitiert nach HILGERS 2008: 9)

Ausgehend von einer negativen Behaftung von Arbeit im antiken Altertum, welche Sklaven vorbehalten war, wandelte sich im Zusammenhang mit dem calvinistischen Protestantismus und der bürgerlichen Emanzipationsbewegung die Auffassung von und die Einstellung zu Arbeit. Sie wurde fortan mit einem hohen ethischen Wert verbunden. Soziale Mobilität und gesellschaftliches Vorankommen ist seit Abschaffung der Ständegesellschaft an Arbeit und ihr resultierendes Produkt, die Leistung, geknüpft. Leistung wurde somit ein gesellschaftliches Selektionskriterium (vgl. KLEBER 1992: 79 ff), weshalb das Leistungsprinzip ein Gestaltungsprinzip, das die Verteilung von materiellen und sozialen Möglichkeiten gerecht regeln soll, für Gesellschaften repräsentiert, weswegen von einer sogenannten Leistungsgesellschaft gesprochen wird (vgl. BENISCHEK 2006: 81). Dadurch wurden Faktoren wie Herkunft, Gruppenzugehörigkeit und Geschlecht unbedeutender und eine Demokratisierung der Gesellschaft durch das Leistungsprinzip setzte ein (vgl. CHUDASKE 2012: 13).¹²

¹² Für vertiefende Ausführungen zum Verhältnis zwischen Leistung und Demokratie unter Berücksichtigung des Bildungskontextes wird auf SALDERN (2011) verwiesen.

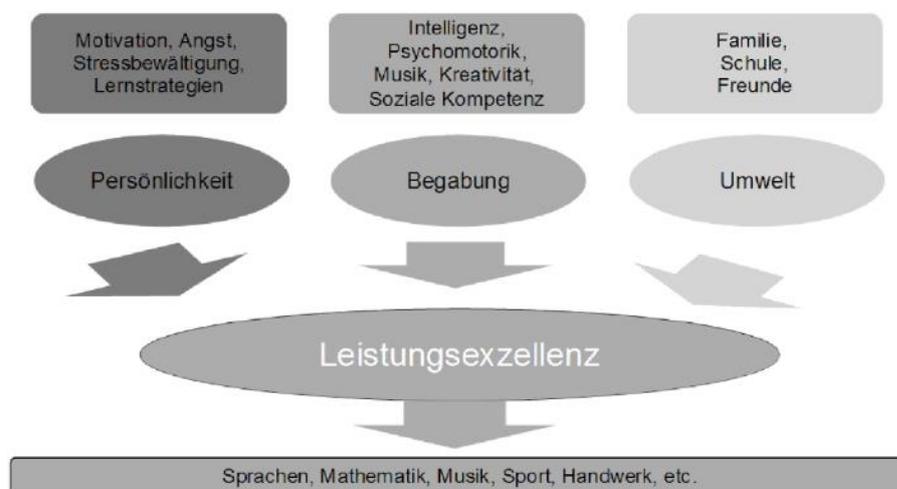
2.1.1 Der Leistungsbegriff

Problemlos definiert werden kann Leistung in den Naturwissenschaften, wo sie in der Disziplin der Physik das Verhältnis der Arbeit zur Arbeitszeit angibt (vgl. RENNERT/SCHMIEDEL 1995: 28). In den Sozialwissenschaften hingegen ist eine allgemeingültige Begriffserläuterung nicht möglich (vgl. BENISCHEK 2006: 81), denn Handlungen und die dadurch hervorgerufenen Produkte, sprich Leistungen, werden abhängig von externen Faktoren unterschiedlich bewertet (vgl. SALDERN 1999: 74).

„Leistung ist ein bewußt [sic] hervorgerufener Handlungsvorgang, der ein allgemein als bedeutsam anerkanntes Produkt hervorbringt. [...] In einem Bewertungsprozeß [sic] von Handlungserlebnissen durch eine Gruppe, die Gesellschaft oder ein Individuum wird bestimmt, was als Leistung angesehen wird und was nicht.“ (KLEBER 1992: 82)

Leistung ist somit nicht nur als zweckbezogen, sondern auch als normbezogen anzusehen, da sie stets in Verhältnis zu einem Bezugssystem zu betrachten und bewerten ist. Weiters unterliegt sie einer gewissen Anlage- und Umweltabhängigkeit, da sie das Zusammenwirken von Fähigkeiten eines Individuums und den äußeren Umweltfaktoren darstellt (vgl. JÜRGENS 1992: 21 ff).¹³

¹³ Anzuführen ist hierbei das Leistungsmodell nach HELLER und HANY, wodurch veranschaulicht wird, welche Faktoren Einfluss auf menschliche Leistung haben und wie facettenreich sich diese zusammensetzt. Abbildung 2: Zusammensetzung von Leistungsexzellenz (zitiert nach SALDERN 2011: 39):



Für vertiefende Ausführungen zu den externen Umweltfaktoren in Bezug auf Schulleistungen wird auf CHUDASKE (2012) sowie HELLER (2001) verwiesen.

„[Leistung ist zu beschreiben als, M.R.] Ergebnis oder Vollzug einer Tätigkeit, die mit Anstrengung und gegebenenfalls Selbstüberwindung verbunden ist und für die Gütemaßstäbe anerkannt werden.“ (KLAFKI ⁵1996: 124)

KLAFKIS Leistungsverständnis zufolge ergibt sich menschliche Leistung¹⁴ demnach aus dem Zusammenwirken der Fähigkeit einer Person, ihren Bemühungen und der aufgebrauchten Anstrengung sowie der Annahme der Herausforderung durch die Aufgabe in Relation zu bestimmten Maßstäben (vgl. KÖNIGSLEHNER 1996: 661). Diese Beschreibungen treffen auf den Leistungsbegriff im sozialwissenschaftlichen Sinne zu, allerdings wirft der Terminus Schulleistung erneut eine bestimmte Problematik auf, da seine Definition ebenfalls nicht einheitlich ist. Die bereits dargebotenen Begriffserläuterungen auf den Schulkontext zu beziehen, ist wenig sinnvoll, da somit sämtliche solcher in der Schule durchgeführten Tätigkeiten als Schulleistung anzusehen wären (vgl. CHUDASKE 2012: 14). Ausschlaggebend für den Schulbezug ist hingegen die Zweckorientierung, denn im Fokus stehen die Lernprozesse und -ergebnisse, die von der Schule und den Curricula initiiert werden (vgl. INGENKAMP/LISSMANN ⁶2008: 131). Nicht nur sozialwissenschaftlicher Auffassung zufolge, sondern vor allem im Schulkontext ist Leistung als ein Prozess und ein Produkt zu verstehen. Beiden Ebenen ist eine gleichwertige Bedeutung zuzumessen:

„Leistung ist ein Prozess und Produkt (Vollzug, Ergebnis, Prozess, Ausführung). Schüler und Eltern erwarten klare Bewertungsgrundlagen und Erkenntnisse über den Leistungsstand. Leistung wird von Lehrkräften häufig am „Produkt“ gemessen. Dem Prozess muss aber eine gleichrangige Bedeutung zugemessen werden. [...] Leistung kann somit als Prozess und Produkt menschlichen Handelns im Kontext von Selbst- und Fremdbewertung gesehen werden.“ (SALDERN 2011: 34)

Im Schulunterrichtsgesetz (SchUG) wird Leistung an sich nicht definiert, jedoch wird erläutert, dass Schulleistungen unter Berücksichtigung des Unterrichtsfaches die selbständige Umsetzung von Arbeiten, die Aufnahme und Anwendung der Lehrinhalte,

¹⁴ Aufgrund der mehrdimensionalen Umsetzung menschlicher Leistung wird diese in Taxonomien, einer Aufschlüsselung von Kategorien hinsichtlich der Einstufung von Lernzielen, klassifiziert. Taxonomien können psychologischer, inhaltlich-thematischer oder kultureller Art sein. Die bekannteste ist die Lernzieltaxonomie nach BLOOM, wobei der kognitive Bereich, welcher in sechs aufeinander aufbauende Anforderungsbereiche unterteilt ist (Wissen, Verstehen, Anwendung, Analyse, Synthese und Evaluation), der affektive Bereich (Aufnahme, Reaktion, Wertung, Organisation und Charakterisierung) und der psychomotorische Bereich (Imitation, Manipulation, Präzision, Handlungsgliederung und Naturalisierung) unterschieden werden (vgl. BENISCHEK 2006: 82 ff). Weiterführendes siehe: BLOOM (²1973).

die Durchführung der Aufgaben sowie eigenständiges Arbeiten der SchülerInnen umfassen (vgl. SchUG § 18 Abs. 3). Eine Definition von Schulleistung für den österreichischen Schulkontext ist auch in der Leistungsbeurteilungsverordnung (LBVO) nicht gegeben (vgl. LBVO). STERN liefert folgende begriffliche Klärung:

„Was als gute Leistung zählt, [...] liegt in unserem Schulsystem weitgehend in der Hand der Lehrer/innen. Lernleistungen von Schüler/inne/n sind Fortschritte in Bezug auf die fachlichen und überfachlichen Lernziele. [...] Was als Lernleistung zählt und überprüft wird, ist meistens die Wiedergabe von Wissen [...], das Vorführen von Fertigkeiten [...] und fachliches Verständnis. [...] Lernfortschritt wird nicht mehr nur am Zuwachs von Wissen abgelesen [...], sondern am Zuwachs an Können.“ (STERN 2010: 25)

Zusammenfassend lässt sich aus den bislang dargebotenen Erläuterungen folgern, dass Schulleistung sowohl der Prozess als auch das Produkt aus der Zusammenwirkung der Fähigkeiten sowie der Anstrengung von SchülerInnen ist, die darauf abzielt, von der Schule definierte (Lern-) Ziele zu erreichen. WEINERT unterstreicht diesbezüglich noch die Bedeutung des Lernens¹⁵ sowie die Differenzierung von Individuen in Hinsicht auf die Schulleistung und schlussfolgert:

„Unabhängig von den unterschiedlichen Fähigkeiten und Talenten der Schüler muss alles gelernt werden, was später gewusst und gekonnt wird. Lernen ist der wichtigste Mechanismus der kognitiven Entwicklung. [...] Unter vergleichbaren schulischen Lernbedingungen ist es nicht möglich, die individuellen Lern- und Leistungsunterschiede generell aufzuheben. Obwohl die individuellen Fähigkeits-, Lern- und Leistungsunterschiede über die Zeit hinweg relativ stabil bleiben, sind die (individuell variablen) Lern- und Leistungsfortschritte eine Funktion der Quantität und Qualität des Lernens und werden mehr oder minder stark von der Wirksamkeit des Unterrichts beeinflusst. Schulleistungen sind also stets Leistungen der Schüler, die durch die Schule begünstigt oder erschwert werden.“ (WEINERT 2002: 85)

2.1.2 Leistungsformen

Durch den Fakt, dass Leistung eine komplexe, mehrdimensionale Handlung darstellt, lassen sich weitere Differenzierungen in Bezug auf die in den Fokus gestellten Gegebenheiten dieses Handlungsvorganges treffen (vgl. BENISCHEK 2006: 85):

¹⁵ Im Allgemeinen wird unter Lernen der Erwerb von Kenntnissen und Fertigkeiten verstanden. In einem weitgefassten, wissenschaftlichen Sinne ist der Prozess des Lernens nötig, um Lebewesen eine Anpassung an sich verändernde Umweltbedingungen (hier der Kontext Schule – vorgegebene Sachverhalte und Lehrziele) zu ermöglichen. Damit dies umgesetzt werden kann, bedarf es eines Speichers, im menschlichen Fall das Gedächtnis, aus dem Gelerntes wieder aktiviert werden kann (vgl. TREMEL 2006: 288 ff).

Produktorientierte Leistung stellt das Ergebnis einer Anstrengung in den Mittelpunkt. Das Ausmaß an Anstrengung ist nicht bedeutend für die Beurteilung, lediglich das Resultat, das entstandene Produkt also, wird berücksichtigt (vgl. FLOREK 1999: 23).

In Abgrenzung dazu wird bei prozessorientierter Leistung nicht das Produkt, sondern die Handlung der Ausführung beachtet. Hierbei spielt die Aktion, die zur Problemlösung angewandt wird, die tragende Rolle (vgl. FLOREK 1999: 23).

Bei Bezug auf die individuelle Leistung rückt die persönliche Anstrengung, die ein Individuum aufbringt, um eine Aufgabe erfolgreich zu verrichten, in den Vordergrund. Neben den persönlichen Voraussetzungen werden die gegebenen externen Umstände, die die Leistungshandlung beeinflussen, einbezogen (vgl. BENISCHEK 2006: 86).

Gegenteilig dazu stellt die kollektive Leistung, das Produkt oder den Prozess eines Kollektiv dar. Sämtliche Leistung, die durch die Gemeinschaft einer Gruppierung erzeugt wird, sei es eine Kleingruppe oder eine große Gesellschaft, wird als kollektive Leistung bezeichnet (vgl. BENISCHEK 2006: 86).

Wird eine Leistung, entweder produktorientiert oder prozessorientiert sowie individuell oder kollektiv, in ein Verhältnis zu einer anderen oder mehreren unterschiedlichen Leistung gesetzt und steht dieser Vergleich im Fokus, wird von wettbewerbsbeziehungsweise konkurrenzorientierte Leistung gesprochen. Im Kontext Schule werden SchülerInnenleistungen untereinander verglichen und unterschiedliche Gesichtspunkte betreffend (Qualität, Quantität, Anstrengung) abgewogen (vgl. FLOREK 1999: 24).¹⁶

Handlungsaktionen sozialer Natur werden als soziale Leistung bezeichnet (vgl. BENISCHEK 2006: 86). Im Schulkontext gewannen solch überfachliche Kompetenzen an Bedeutung und sind in das Unterrichtsgeschehen integriert:

„Sie [die sozialen und personalen Kompetenzen, M.R.] legen den Fokus ihres Mehrwerts auf die Persönlichkeit der Schülerin, des Schülers in der Auseinandersetzung mit sich selbst und der Umwelt. Qualitäten wie Reflexionsfähigkeit, bewusstes Einbringen in die Gemeinschaft, Wachsen an Aufgaben, bewusstes Aneignen von Verhaltensregeln im sozialen Miteinander sind nur einige wesentliche Aspekte. Die Förderung sozialer und personaler Kompetenzen kann als Querschnittsaufgabe schulischer Ausbildung gesehen werden, die sich auf den gesamten Schulalltag und die gesamte Schule bezieht.“
(<https://www.bmb.gv.at/schulen/unterricht/uek/sozial.html>)

¹⁶ Verweis auf die soziale Bezugsnorm. Näheres siehe Unterkapitel 2.3

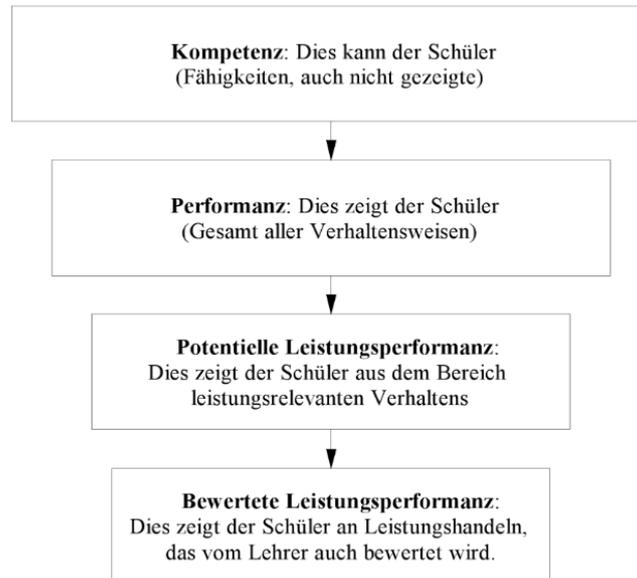
Als dynamische Leistung gelten Aufgaben von hoher Originalität. Dies umfasst theoretische Ideen, Handlungen oder produzierte Werke. Im Mittelpunkt steht der Erkenntnisfortschritt und die hohe Problemlösungskompetenz, um soziale sowie technische Herausforderungen bewältigen zu können (vgl. FLOREK 1999: 25).

2.2 Leistungsbeurteilung

Wie bereits im ersten Kapitel erarbeitet wurde, wird im Zuge des Assessments Schulleistung durch die evaluierende Person wahrgenommen, erkannt, eingeordnet oder bewertet und in weiterer Folge, im Sinne einer Evaluation, beurteilt, wobei ein solches Urteil in einer gewissen Weise einem Gutachten ähnelt. Die Leistungsbeurteilung ist demnach als ein Prozess zu verstehen, der sich in zwei Phasen gliedert: Einerseits wird die Schulleistung erhoben und festgestellt, weshalb die erste Phase als Leistungsfeststellung bezeichnet wird, andererseits erfolgt eine Beurteilung nach vorgegebenen Kriterien in Form einer Einordnung oder Zuordnung (beispielsweise in eine Notenskala) der Leistung, welche die Leistungsbeurteilung repräsentiert (vgl. NEUWEG⁵2014: 4).

Leistungsfeststellungen sind das Aufzeigen von Informationen den Wissens- und Kenntnisstand von Schülerinnen betreffend. Hierfür finden unterschiedliche Arten von Aufgaben Anwendung, um mit Hilfe dieser Messinstrumente, die Schulleistungen festhalten zu können (vgl. STERN 2010: 28). Ergebnisse von Leistungsmessungen können in vielfältigen Formen angegeben werden, wie beispielsweise in Absolutpunkten, Prozentpunkten, Beschreibung der demonstrierten Fertigkeiten samt Ausprägungsgrad und viele weitere (vgl. NEUWEG⁵2014: 4). Eine GERS-konforme Leistungsfeststellung von Sprachleistungen bezieht sich stets auf die Qualitätskriterien der Validität, Reliabilität und Durchführbarkeit (vgl. ÖSZ 2008: 7). Erfasst werden kann nur, was beobachtbar ist, weshalb in diesem Kontext eine Unterscheidung zwischen Kompetenz und Performanz vorgenommen werden muss (vgl. SALDERN 2011: 36), denn „jeder Leistung liegt ein entsprechendes Potenzial zugrunde, die Umsetzung eines Potenzials in aktuelle Leistungen hängt aber von zahlreichen Drittvariablen ab“ (ROST/BUCH 2010: 258). Die Kompetenz umfasst die Fertigkeiten eines Schülers/einer Schülerin, somit die Gesamtheit dessen, was geleistet werden kann. Die Performanz ist davon abzugrenzen, da sie eine konkrete Realisierung darstellt, der gewisse Fertigkeiten zugrunde liegen (vgl. SALDERN 2011: 36).

Aus folgender Abbildung geht der Prozess hervor, wie Kompetenzen in Performanzen festgehalten werden können und folglich als Leistungsfeststellung einer Bewertung der beurteilenden Person unterzogen werden, wodurch die Leistungsbeurteilung zustande kommt.



Von der Kompetenz zur bewerteten Leistungsperformanz

Abbildung 3 (SALDERN 2011: 38).

„Leistungsbeurteilung ist die im Anschluss an die Leistungsfeststellung vorgenommene Bewertung des Messergebnisses durch den Vergleich mit einem Beurteilungsmaßstab. Das Ergebnis der Leistungsbeurteilung ist in Österreich durch die vom Gesetzgeber definierten Beurteilungsstufen (Noten) auszudrücken.“ (NEUWEG ⁵2014: 4)

2.3 Bezugsnormen der Leistungsbeurteilung

Wie bei der Begriffserläuterung von Leistung dargelegt wurde, ist diese immer als normbezogen anzusehen, da sie stets in Verhältnis zu einem Bezugssystem zu bewerten ist, denn eine Gruppe, Gesellschaft oder ein Individuum entscheidet, was Leistung ausmacht und was als solche bezeichnet wird beziehungsweise ihren Ausprägungsgrad (vgl. KLEBER 1992: 82).

„Ein Muster oder Leitbild, das als objektives Maß für die Bewertung von Leistungen festgelegt wird, hat im allgemeinen nur einen zeitlichen Wert. Selbst ein idealer Vergleichswert ist ein menschliches Konstrukt. Es wird hier und heute als gültig angesehen, weil es nach dem neuesten Stand der Forschung als begründet erscheint und weil es in Übereinstimmung mit einem bestimmten Bewußtsein [sic] der Menschen steht.“ (FLOREK 1999: 104)

Die Bezugsnormorientierung ist somit laufender Überarbeitung ausgesetzt (vgl. FLOREK 1999: 104) und berücksichtigt die Präferenzen einer Person, Gruppe oder Gesellschaft, die eine Leistung einer Bewertung in einem Bezugssystem unterzieht. Eine bedeutende Funktion ist dieser im Schulkontext beizumessen, da die Bezugsnormorientierung von Lehrpersonen wegweisend dazu beiträgt, wie Beurteilungen von Leistungen interpretiert, eingeordnet und begründet werden. Erfolge sowie Misserfolge von Lernenden wollen laut MAIER auf ihre Anstrengung beziehungsweise auf ihre unzureichende Anstrengung zurückgeführt werden, um die Leistungsbeurteilung auf personenabhängige Faktoren und keine äußeren Umstände zu basieren, sodass Leistungserfolge als internal erklärt werden und eine erhöhte Motivation bezüglich der LernerInnenanstrengung zustande kommt (vgl. MAIER 2015: 85 f). Eine Diskussion über die mittlerweile klassische dreifache Gliederung von Bezugssystemen ist bis auf die Zeit um 1900 zurückzuführen (vgl. FLOREK 1999: 106). Diese Bezugsnormen, die als Vergleichsperspektiven aufzufassen sind, stellen allerdings keineswegs ein theoretisches Konstrukt, das aus einem psychologischen oder bildungswissenschaftlichen Kontext hervorging, dar, sondern sind vielmehr alltägliche Vergleichsmaßstäbe. Hierbei ist eine individuelle von einer sozialen Bezugsnorm zu differenzieren, je nachdem ob eine Person mit einer anderen beziehungsweise mehreren anderen verglichen wird oder ob ein Individuum mit eigenen früheren Leistungen als Maßstab bewertet wird, wodurch ein temporaler, diachroner Aspekt miteinbezogen wird (vgl. RHEINBERG ²2002: 61). Neben den beiden personenbezogenen Bezugssystemen ist als drittes die kriteriale oder sachliche Bezugsnorm anzuführen, die an gesetzten Lehrzielen festgemacht wird und somit einen Inhaltsbezug aufweist (vgl. NEUWEG ⁵2014: 76 ff).

Die soziale Bezugsnorm bezieht sich auf das Verhältnis einer Schulleistung zu denen der anderen in einer LernerInnengruppe, beispielsweise in einer Klasse. Hierbei sind gewisse Vor- sowie auch Nachteile auszumachen, wenn durchschnittsorientiert beurteilt wird, denn Beurteilungsmaßstäbe werden nicht im Vorhinein festgelegt, sondern erst die Ergebnisse ergeben die Wertbestimmung der Bewertung. Vorteile sind, dass eine Falscheinschätzung bezüglich der Schwierigkeit einer Aufgabe seitens der Lehrperson oder etwaigen Vermittlungsproblemen des Lehrinhalts die daraus resultierenden mangelhaften Leistungen angepasst werden und das Beurteilungsmaß gemildert wird (vgl. NEUWEG ⁵2014: 76 f). Darüber hinaus ist leicht erkennbar, welche Gruppenzugehörigen überdurchschnittlich gut oder schlecht abschneiden, da meist das

gesamte Ausmaß einer Beurteilungsskala zur Anwendung kommt. Als wesentliche Nachteile sind zu nennen, dass Beurteilte keinerlei Information darüber erhalten, inwieweit die Lernziele erfüllt wurden, dass durch die zufällige Angehörigkeit zu einer Bezugsgruppe ein und dieselbe Leistung unterschiedlich eingeordnet werden kann, sowie dass schwächere Gruppenmitglieder in wiederholtem Ausmaße schlechte Beurteilungen hinnehmen werden müssen (vgl. STERN 2010: 29). Weiters gilt es, die Förderung von kompetitiven Verhaltensweisen zu beachten, denn gute Schulleistungen können in einem solchen Bezugssystem nur erzielt werden, wenn jene der anderen verhältnismäßig weniger steigen oder gar sinken (vgl. SCHMIDINGER 2012: 13).

„Wird der Standpunkt der Teilnehmer an einem Wettbewerb dagegen von Konkurrenz geprägt, sind die Handlungen des einzelnen auf das vordringliche Ziel gerichtet, dem Partner zuvorzukommen, den Gegner zu übertreffen und den größeren Gewinn zu haben. Man möchte vor allem besser als andere sein, um dadurch Ansehen zu erlangen oder seine soziale Stellung zu wahren. [...] Wichtig ist nur, daß [sic] das eigene Werk den Vergleich mit dem aushält, was die Gegner zustande gebracht haben („soziale Bezugsnorm“).“
(FLOREK 1999: 77)

Bei einer zweiten personenbezogenen Bezugsnorm steht ein einziges Individuum im Fokus, weshalb von einer individuellen Bezugsnorm zu sprechen ist. Hierbei sind Schulleistungen eines Schülers beziehungsweise einer Schülerin im Verhältnis mit vorangegangenen Resultaten zu vergleichen, wodurch der Lernentwicklung die zentrale Rolle beigemessen wird. Dies ist allerdings nur möglich, sollten bereits Leistungsergebnisse vorliegen (vgl. MAIER 2015: 88), da ein temporaler, diachroner Ansatz Grundlage ist (vgl. RHEINBERG ²2002: 61). Im Schulwesen muss eine solche Lernverlaufsdiagnostik fortführend über einen längerfristigen Rahmen durchgeführt werden, weil Kompetenzen und Wissen kumulativ¹⁷ aufgebaut werden (vgl. MAIER 2015: 88). Als besonders positiv ist die individuelle Bezugsnorm vor allem im Anfängerunterricht sowie in stark heterogenen Gruppen anzusehen. Unabhängig von jeglicher Ausgangssituation wirkt sich dieses Bezugssystem motivationsfördernd und begünstigend auf den Lernprozess aus (vgl. SCHMIDINGER 2012: 12), denn nur der eigene Lernzuwachs steht im Vordergrund völlig frei vom Vergleich oder Verhältnis zu anderen Leistungen. Jedoch ist kein Bezug zu anderen Leistungen oder allgemeinen

¹⁷ Der Lehrplan ist im österreichischen Schulkontext als Spirallehrplan zu verstehen (vgl. Lehrplan HAK 2014: 4), wobei Lehrinhalte immer wieder in komplexerer Form aufgegriffen werden, wodurch sich Fertigkeiten und Wissen systematisch erweitern und anhäufen (vgl. WOOLFOLK 2008: 425).

Leistungserwartungen durchführbar, weshalb die gesellschaftliche und prognostische Funktion einer beurteilenden Leistungsüberprüfung nicht gegeben ist (vgl. NEUWEG⁵2014: 76 f).

Von diesen beiden personenbezogenen Bezugssystemen kann der kriterienorientierte Maßstab abgegrenzt werden¹⁸ (vgl. STERN 2010: 29 f). Bezieht sich die Überprüfung einer Leistung auf festgesetzte Kriterien, die im Vorfeld der Beurteilung definiert wurden, so ist Ziel einer solchen Leistungsbewertung die Ausprägung der Lernzielerfüllung anzugeben (vgl. NEUWEG⁵2014: 80). Die Orientierung an einer Sachnorm, die im österreichischen Schulkontext der Lehrplan darstellt, ist „eine plausible, aber nicht immer die lernförderlichste Bezugsnorm“ (SCHMIDINGER 2012: 12). Laut LBVO sind der „Maßstab für die Leistungsbeurteilung [...] die Forderungen des Lehrplanes unter Bedachtnahme auf den jeweiligen Stand des Unterrichtes“ (LBVO § 11, Abs. 1). Rechtlich tragbar ist im wesentlichen die Lehrzielorientierung unabhängig von jeglicher sozialen Bezugsnorm, wohingegen die individuelle Orientierung in Ausnahmefällen zur Anwendung kommen darf (vgl. NEUWEG⁵2014: 76 ff).

„Lehrzielorientiert beurteilen kann nur, wer über klare sachliche Beurteilungskriterien verfügt. [...] Gegenwärtig ist aber weder den Notendefinitionen noch den Lehrplänen eindeutig zu entnehmen, welche Leistung ein Schüler [...] konkret zu erbringen hat. Dass Lehrer sich häufig am Leistungsdurchschnitt der Klasse orientieren, ist insofern verständlich. Dringend erforderlich ist daher die Ausarbeitung von Kompetenzrastern [...], die mit klaren Beurteilungsmodellen verbunden werden.“ (NEUWEG⁵2014: 82).

Eine solche Vereinheitlichung der Leistungsbeurteilung mit Hilfe von Kompetenzrastern findet zunehmend häufiger Einzug in das schulische Assessment, wie auch im österreichischen Schulwesen (vgl. NEUWEG⁵2014: 15). Das Instrument des Beurteilungsrasters wird im Zuge dieser Arbeit genauer erläutert und untersucht werden, besonders jener, der im romanischen Fremdsprachenunterricht in österreichischen Schulen auf Sprachniveau B1 Anwendung findet. Davor ist es allerdings unerlässlich, die stattgefundenen Standardisierungsprozesse im fremdsprachlichen Kontext kurz zu umreißen und auf die Schriftproduktion im Französischunterricht einzugehen.

¹⁸ Hierbei ist ein Verweis auf Unterkapitel 1.1 dieser Arbeit möglich, wo bereits in Bezug auf die unterschiedlichen Formen von Bewertung normorientiertes von kriterienorientiertem Assessment abgegrenzt wurde.

3 Standardisierung des Fremdsprachenunterrichts

Die stattgefundenen Standardisierungsprozesse in der Schulentwicklung sind auf das Interesse, vorgegebene Ziele und gewisse Standards flächenübergreifend zu erreichen, damit eine Qualitätssicherung und Vergleichbarkeit im Bildungswesen erfolgen kann, zurückzuführen (vgl. HILLBRANDT/SINTZEN-KÖNIGSFELD 2009: 15). In diesem Zusammenhang muss der Gemeinsame Europäische Referenzrahmen für Sprachen des Europarats (GERS) angeführt werden, denn

„diese Publikation hat nicht nur [...] eine intensive fachdidaktische Diskussion zum Fremdsprachenunterricht ausgelöst, sondern ihre inhaltlichen Anregungen und Vorgaben sind in zahlreichen europäischen Ländern aufgegriffen und in Lehrplanvorgaben eingearbeitet worden.“ (LEUPOLD 2010: 40)

Schulen vergeben an ihre AbsolventInnen öffentlich bescheinigte Atteste, die gewisse Leistungen und Kompetenzen anerkennen, wodurch sicherzustellen ist, dass diese tatsächlich vergleichbar sind, wo gleiche Zertifikate ausgestellt werden (vgl. HILLBRANDT/SINTZEN-KÖNIGSFELD 2009: 15). Diesbezüglich ist es ein Ziel des GERS, verschiedenen Instanzen eine Hilfestellung bezüglich dem Vergleich von Qualifikationssystemen zu ermöglichen, wofür Referenzniveaus mit Kann-Beschreibungen in den unterschiedlichen Teilkompetenzen heranzuziehen sind (vgl. LEUPOLD 2010: 44 ff). Der Kompetenzbegriff im Fremdsprachenunterricht wurde ebenso maßgeblich durch den GERS beeinflusst, wie bereits die Erläuterung des komplexen Verständnisses von kompetenzorientiertem Fremdsprachenunterricht (vgl. LEUPOLD 2010: 202) in Unterkapitel 1.4 dieser Arbeit zeigte, wobei die unterschiedlichen zu einer situationsadäquaten, erfolgreichen Kommunikation nötigen Fertigkeiten dargelegt wurde.

3.1 Bildungsstandards

„Die im deutschsprachigen Raum so bezeichneten ‚Bildungsstandards‘ sollen dazu beitragen, dass erbrachte Leistungen an entsprechenden kriteriumsorientierten Bezugsnormen in Form von Standards gemessen und bewertet werden. Diese lassen sich als verbindliche Festlegungen charakterisieren, die jene zentralen Kompetenzbereiche beschreiben, die Schüler/innen bis zu einer bestimmten Jahrgangsstufe erworben haben sollen; sie stellen somit Ergebniserwartungen dar.“ (HILLBRANDT/SINTZEN-KÖNIGSFELD 2009: 15 f)

Diese Richtlinien, wobei bestimmte Kompetenzen mit einem gewissen Punkt in der Schullaufbahn in Verbindung stehen, wurden unter Bezugnahme auf den GERS verfasst. Diese konkret formulierten Lernergebnisse bezeichnen also die erwünschten

grundlegenden Lernstände der SchülerInnen zu gewissen Zeitpunkten in ihrer schulischen Entwicklung (vgl. <https://www.bifie.at/bildungsstandards>).

Eine Ausdifferenzierung kann hierbei hinsichtlich des Anforderungsniveaus vollzogen werden, da zwischen Mindest-, Regel- und Maximalstandards unterschieden wird. In Bezug auf das schulische Assessment sind vor allem die Mindest- und Regelstandards von Bedeutung, denn SchülerInnen müssen ein bestimmtes Minimum an Fertigkeiten aufweisen, um einen Bildungsabschnitt positiv absolvieren zu können, wohingegen Regelstandards auf jene Kompetenzen verweisen, „die im Durchschnitt von den Schülerinnen und Schülern bis zu einem bestimmten Bildungsabschnitt erreicht werden sollten“ (LEUPOLD 2009: 35). Aus dieser Standardisierung von Bildungsinhalten ergab sich die Notwendigkeit einer zentralen, standard-orientierten Leistungsüberprüfung (vgl. HILLBRANDT/SINTZEN-KÖNIGSFELD 2009: 26 f), die in Österreich die standardisierte Reife und Diplomprüfung repräsentiert (vgl. https://www.bmb.gv.at/schulen/unterricht/ba/reifepruefung_bhs.html). Sie „ermöglicht eine Vergleichbarkeit der Leistungen unter denselben Bedingungen in ganz Österreich.“ (https://www.bmb.gv.at/schulen/unterricht/ba/reifepruefung_bhs.html). Aus diesen Gegebenheiten der Standardisierungsprozesse wurde in Anlehnung an die Gütekriterien der Testtheorie versucht, Standards für die Leistungsbewertung zu definieren, die als sinnvolle Prinzipien lernförderlich wirken und die Unterrichtsplanung positiv beeinflussen sollen. Die amerikanische Mathematiklehrvereinigung NCTM formulierte sechs Prinzipien für diese *assessment standards*, zu denen Leistungstransparenz, Lernförderung, Chancengerechtigkeit, Offenheit, Schlüssigkeit sowie Kohärenz zählen (vgl. STERN 2010: 95 f).

Neben diesen positiven Effekten der Vergleichbarkeit und Bildungsgerechtigkeit wurden allerdings einerseits die Reduzierung und Trivialisierung des Unterrichts, andererseits die Veränderung der Testkultur mit Hauptfokus auf jene Teilbereiche, die sich auch psychometrisch messen lassen (vgl. LEUPOLD 2009: 35), wodurch sich ein *Teaching-to-the-test-Effekt* ergeben kann (vgl. HILLBRANDT/SINTZEN-KÖNIGSFELD 2009: 29), als negative Aspekte geäußert. Weiters lässt sich ein Dilemma ausmachen, welches die Gegenüberstellung und Konkurrenz von Standardisierung und Individualisierung im Schulunterricht betrifft¹⁹. Dies bezieht sich vor allem auf die Zwiespältigkeit zwischen offenen Formen sowie neuen Lernkulturen, wobei auf individuelle Formate Rücksicht

¹⁹ Näheres hierzu siehe: BUSCHKÜHLE, Carl-Peter/DUNCKER, Ludwig/OSWALT, Vadim (2009).

genommen wird, und einer zeitgleichen Vorbereitung auf standardisierte Testformate, die sprachliche Teilfähigkeiten messen wollen und das daraus resultierende Problem keinen der beiden Wege zu vernachlässigen (vgl. REZAT 2014: 97 ff). Im wissenschaftlichen Diskurs stieß die Standardisierung auf einigen Seiten prinzipiell auf harte Kritik (vgl. LEUPOLD 2009: 35), wobei es im Zuge dieser Arbeit nicht möglich scheint die Standardisierung und ihre Auswirkungen, abgesehen von der Leistungsbeurteilung, ausführlicher zu thematisieren.²⁰

Im Zuge der Standardisierung wurden, wie bereits erwähnt, Standards festgelegt, die jeder Schüler und jede Schülerin erreicht haben müsste, um eine positive Bestätigung der Absolvierung des jeweiligen Kompetenzniveaus in einer Fremdsprache zu erhalten, weshalb es einer zentralen, standardorientierten Leistungsüberprüfung bedurfte, die die vorab vereinbarten Standards abprüft, weshalb gewissermaßen eine Kanonisierung im Schulbetrieb Einzug hielt (vgl. HILLBRANDT/SINTZEN-KÖNIGSFELD 2009: 15 ff). Dass der gesamte Standardisierungsprozess im Kontext Schule im Rahmen dieser Arbeit nicht vollständig erarbeitet werden kann, ist nur verständlich, weshalb in Hinblick auf den Fokus, die Leistungsbeurteilung von schriftlichen Performanzen, näher auf die Standardisierung dieser eingegangen werden soll.

3.2 Standardisierungsprozesse hinsichtlich schriftlicher Produktion

Die schriftliche Produktion betreffend können in Hinsicht auf eine gewisse Kanonisierung die Textsorten als wichtiges Merkmal der Standardisierung angeführt werden. Im Gemeinsamen Europäischen Referenzrahmen wird an verschiedenen Stellen auf die Textarbeit eingegangen, jedoch will der GERS keine verpflichtenden Vorgaben diktieren, sondern vielmehr didaktische Hilfestellungen und übergreifende Normierungen von Kompetenzniveaus zur Verfügung stellen, um die Planung des Fremdsprachenlernprozesses aus allen Perspektiven zu erleichtern (vgl. EUROPARAT 2001: 17 ff). Da die Realisierungen schriftlicher Kompetenzen äußerst vielfältig sein können (vgl. LEUPOLD 2009: 282 ff), war es nötig, die möglichen abprüfbaren Textsorten einzuschränken und die Formate der Prüfung genau zu definieren. Die vorgegebenen Textsorten werden im Textsortenkatalog des BIFIE angeführt. Für Französisch auf Sprachniveau B1 sind die Textsorten Artikel, Bericht, E-Mail, Brief, Blog, Antrag und

²⁰ Für weiterführende Informationen wird verwiesen auf: TESCH/LEUPOLD/KÖLLER 2009 sowie RAITH 2011.

Broschüre in Abhängigkeit des besuchten Schultyps vorgesehen. Die einzelnen Textsorten werden konkret definiert und weiters wird erläutert, welche Ansprüche zu erfüllen sind, um einen adäquaten Text produzieren zu können (vgl. <https://www.bifie.at/node/1462>). Einhergehend mit der Vorgabe von bestimmten Textsorten werden Sprachhandlungen, die diesem Kompetenzniveau zugeschrieben werden, vorgegeben, denn jede Textsorte verlangt nach spezifischen sprachlichen Handlungen, um den Zweck und die Funktion des Schreibens umsetzen zu können (vgl. KRAUSE 2005: 3 ff). Durch die Vorgabe von Inhalten werden in Konkurrenz gleichermaßen die Kriterien der Überprüfung festgesetzt, denn eine kulturspezifische Textsortenkompetenz²¹ umfasst sprachliche Aspekte sowie Wissen hinsichtlich Aufbau und Layout (vgl. <https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>).

„Um Kandidatinnen und Kandidaten eine optimale Vorbereitung auf die standardisierte kompetenzorientierte Reife- und Diplomprüfung in den lebenden Fremdsprachen zu ermöglichen, sollten die zur Anwendung kommenden Testmethoden im Unterricht frühzeitig thematisiert werden. Die Lehrerinnen und Lehrer sollten ihre Schülerinnen und Schüler spätestens am Beginn des vorletzten Lernjahres mit dem Beurteilungsraster der jeweiligen Zielkompetenzstufe vertraut machen, um ihnen zu verdeutlichen, welche qualitativen Anforderungen bei der Textproduktion erfüllt werden müssen.“ (<https://www.srdp.at/schriftliche-pruefungen/lebende-fremdsprachen/allgemeine-informationen/>)

Angeleitet wird die zu überprüfende Schreibhandlung durch eine standardisierte Aufgabenstellung, die durch das BIFIE erstellt und mit Hilfe von Operatoren, Handlungsanweisungen, die zur Erfüllung der Aufgabenstellung von Nöten sind, formuliert wird (vgl. CEBS 2013: 23 ff). Für diese Diplomarbeit relevant sind allerdings weniger die Aufgabenstellungen, sondern die standardisierte Leistungsbewertung, die der österreichischen Abschlussprüfung von Fremdsprachen durch einen Beurteilungsraster zu erfolgen hat und ebenso für die Beurteilung von sonstigen schriftlichen Arbeiten, wie Hausübungen und Portfolios empfohlen wird. (vgl. <https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>). Nach diesem äußerst kurzen Abriss, der die wichtigsten Punkte der erfolgten Standardisierungen hinsichtlich der Schriftproduktion offenlegt, wird der Fokus auf den in diesem Prozess entwickelten Beurteilungsraster gelegt, wobei dieser zunächst näher theoretisch untersucht und analysiert wird, woran eine empirische Untersuchung zu diesem Beurteilungsinstrument anschließt.

²¹ Für ausführlichere Erläuterungen zur Textsortenkompetenz im Fremdsprachenunterricht wird auf ADAMZIK (2005) verwiesen.

4 Standardorientierte, kriterienbasierte Leistungsbeurteilung schriftlicher Performanzen im Französischunterricht

Ein in den vorherigen Kapiteln dargelegter kompetenzorientierter Unterricht fordert eine ihm gerecht werdende kompetenzorientierte Leistungsüberprüfung (vgl. LUCYSHYN 2007: 574), die sich auf im Vorhinein festgesetzte Kriterien bezieht, um lehrzielorientiert im Sinne der Standards testen zu können (vgl. GROTHJAHN 2008: 175 f).

„Sowohl bei den Bildungsstandards-Tests als auch bei der Standardisierten Kompetenzorientierten Reifeprüfung wird eine „Positivbeurteilung“ angewandt [...]. Sowohl im internationalen Kontext als auch in Österreich setzt sich immer mehr diese kriterienbezogene Leistungsbeurteilung durch, da sie klarer, transparenter und vergleichbarer ist. [...] Testaufgaben müssen immer ganz klar auf eine der Fertigkeiten fokussiert sein.“
(ISAK 2010: 86)

Der Hauptzweck, den standardisierte fremdsprachliche Leistungstests verfolgen, ist die Sprachfertigkeiten des beziehungsweise der Geprüften möglichst objektiv zu bewerten, wobei individuelle Lern- und Leistungsfortschritte unberücksichtigt bleiben, da es gilt, gänzlich unabhängig die gesetzten Standards zu überprüfen, um eine valide Aussage über die Sprachfertigkeiten in einem weitläufig anerkannten Kontext tätigen zu können (vgl. BRUSCH 2009: 203). Solche Leistungsfeststellungen, wie die standardisierte Reife- und Diplomprüfung, sind formelle, punktuell stattfindende, summative und direkte Tests, die im Schulwesen intern durchgeführt aber bei Sprachzertifikaten beispielsweise von externen Prüfern abgenommen werden (vgl. DECKE-CORNILL/KÜSTER³2015: 259 ff).

Die Standardorientierung bei der Leistungsbeurteilung bezieht sich auf die Bildungsstandards, die sich im Lehrplan und sonstigen curricularen Vorgaben als Kompetenzstandards niederschlagen, die es zu beherrschen und anwenden zu können gilt. Diese wiederum gehen auf die allgemein akzeptierte Niveauskalierung des GERS (A1, A2, B1, B2, C1, C2) für die einzelnen Teilkompetenzen zurück. Das Gesamt der unterschiedlichen Beschreibungen, die beherrscht werden können sollen und demnach als Regelstandards aufzufassen sind, bildet den Standard für ein Referenzniveau, beispielsweise B1. Solche Kompetenzstufenmodelle erfüllen die Zuordnung einer Lernleistung zu einem Referenzniveau (vgl. LEUPOLD 2010: 412 ff)

„In der Leistungsdiagnostik sollen Bildungsstandards somit als *kriteriale, schulübergreifende Bezugsnorm* fungieren [...]. Erst auf ihrer Grundlage kann die individuelle Ausprägung einer Kompetenz unabhängig vom jeweiligen Klassenniveau vergleichbar identifiziert werden. Standardbasierte Lernstandsbestimmungen sind auf Kompetenzstufenmodelle angewiesen, die *Niveaustufen* einer Kompetenz angeben [...].“ (DRIESCHNER 2011: 114)

Somit ist festzuhalten, dass durch die Standardorientierung automatisch eine kriterienbasierte Leistungsfeststellung erfolgt, da die Standards die Kriterien in einem gewissen Maße vorgeben und bestimmen. Die im österreichischen Französischunterricht berücksichtigten Kriterien werden im Folgenden genauer behandelt, wovor allerdings die schriftliche Leistungsüberprüfung mit Beurteilungsrastern genauer betrachtet wird.

4.1 Besonderheiten bei schriftlichen Performanzen

Sprachliche Leistungsbewertung stellt grundsätzlich eine Herausforderung aufgrund des äußerst komplexen Prozesses dar, denn es gilt, die vielfältigen Realisierungsmöglichkeiten von Sprache und die damit in Zusammenhang stehenden Fertigkeiten zu überprüfen (vgl. REZAT 2014: 97 f). Eine Sprachperformanz dient immer der Kommunikation und ist somit situationsgebunden sowie adressatenspezifisch, weshalb dies schwer zu erfassen ist und ganzheitliche integrative Aufgabenstellungen benötigt (vgl. BRUSCH 2009: 204).

„Oberstes Ziel ist, dass der Schüler bestimmten Situationen kommunikativ gewachsen ist. Auf der zweiten Lernzielebene beweist der Schüler seine Performanzfähigkeit, kommunikativ handelnd oder diskursiv reflektiert sprachliche Mittel erfolgreich zu verwenden. Erst an dritter Stelle, wenngleich nicht unwichtig für die curriculare Planung, wird die linguistische Kompetenz, also die Problematik grammatischer und stilistischer Normen, genannt“ (PIEPHO 1974: 335 f)

Eine schriftliche Performanz repräsentiert eine konkrete Realisierung, der gewisse Fertigkeiten aus der Gesamtheit an zur Verfügung stehenden Kompetenzen zugrunde liegen (vgl. SALDERN 2011: 36), wie aus dem Zitat nach PIEPHO ersichtlich ist. Schriftliche Leistungen scheinen im Vergleich zur mündlichen Sprachproduktion doch greifbarer und objektivierbarer überprüfbar (vgl. NIEWELER 2010: 219), dennoch handelt es sich um einen Bereich, der bei Lehrpersonen, wie die Beurteilungspraxis zeigt, seit Jahrzehnten immer noch Unbehagen auslöst²² (vgl. REZAT 2014: 100).

²² Näheres hierzu: vgl. INGENKAMP (1995) und BIRKEL (2003)

„Die sprachliche Leistungsmessung anhand geschriebener Texte ist wegen deren Komplexität nur sehr eingeschränkt zuverlässig und objektiv“ (BÖRNER ³2001: 293) postuliert BÖRNER, wobei er darauf verweist, dass eine Bewertungsgerechtigkeit bei Textperformanzen schwer zu erreichen ist²³ (vgl. BÖRNER ³2001: 293).

In diesem Kontext können weitere auf die Bewertung einfließende Faktoren genannt werden, denn ein Ergebnis und Werturteil sprachlicher Leistungen ergibt sich aus der Zusammensetzung vielseitiger Einflüsse. Prinzipiell werden diese in drei Überkategorien unterteilt, welche die Aufgabe, die Testperson und die beurteilende Person umfassen. In standardisierten Testungen ist der Faktor der Aufgabenstellung, der am wenigsten ausschlaggebend für schwankende Ergebnisse, da eine Gruppe eine einheitliche Instruktion bekommt, die nur mehr in der Art der Testdurchführung variiert. Die Testperson und all ihre Merkmale, die das Weltwissen generell und sonstige persönliche Daten, wie Erstsprache, Alter sowie auch auf die Prüfung bezogene Sachverhalte, wie Vertrautheit mit dem Aufgabenformat beispielsweise, umfassen, sind als bedeutende Einflussgröße zu berücksichtigen, die sich im Produkt der Performanz niederschlagen. Direkte Einwirkung auf das Resultat verübt der Evaluator oder die Evaluatorin, weil es sich hierbei um jene Person handelt, die das Urteil hinsichtlich einer schriftlichen Performanz fällt, somit ist diese zu einem bedeutenden Teil von Merkmalen des Beurteilers beziehungsweise der Beurteilerin und der Beurteilung an sich abhängig. Auch hier sind persönlich Merkmale sowie Einstellungen und momentane Befindlichkeiten aber auch Bezugsnormen ausschlaggebend, denn nach der Beurteilung einiger herausragenden Leistungen, wirkt eine durchschnittliche Performanz weniger kompetent als in anderen Vergleichssituationen, wobei eine objektive Distanz der beurteilenden Person gefragt ist. Diesbezüglich ist auch der Faktor der Beurteilung anzugeben, da die verwendete Beurteilungsskala, abhängig von ihrem subjektiven Auffassungsgrad und Interpretationsraum, Schwankungen und negative Einflüsse in der Bewertung verringern kann (vgl. GROTHJAHN/KLEPPIN 2008: 188 f).

„Sicherlich lassen sich vor allem im Fall komplexer produktiver sprachlicher Leistungen Verzerrungen nie völlig vermeiden. Nichtsdestoweniger sollte man auch im schulischen Kontext danach trachten, den Einfluss verzerrender Faktoren möglichst gering zu halten.“
(GROTHJAHN/KLEPPIN 2008: 189)

²³ Verweis auf HUGHES (²2003): Kapitel 9: Testing writing.

Höchstmögliche Objektivität kann durch geschlossene Aufgabenformate, wie Multiple-Choice, garantiert werden, da hierbei die beurteilende Lehrperson keine Einflussgröße darstellt, denn diese Testtypen haben eine eindeutige Antwort ohne Spielraum und sind maschinell auswertbar (vgl. GROTHJAHN 2010: 213), wobei andere Nachteile zu beachten sind, beispielsweise ein erhöhtes Raterisiko, das Abprüfen von Faktenwissen und weniger sprachliche Kompetenz sowie mangelnde Berücksichtigung eines handlungs- und kompetenzorientierten Fremdsprachenkenntnis berücksichtigen (vgl. SACHER 2011: 33 f).

Während diese objektiven, geschlossenen Testtypen speziell bei den rezeptiven Fertigkeiten Anwendung finden, sind im produktiven sprachlichen Bereich offene Aufgabenformate gezwungenermaßen nötig, um kommunikatives Handeln bewerten zu können (vgl. LEUPOLD 2010: 420). Bei individueller Sprachproduktion in offenen Aufgaben sind unterschiedliche und vielfältige Lösungen möglich, weshalb eine subjektive Einschätzung des/der Beurteilenden nötig ist. Hierbei können Antwortmöglichkeiten nicht nur *richtig* oder *falsch* sein, sondern es bedarf einer persönlichen Bewertung des Ausprägungs- und Erfüllungsgrades der vorliegenden Performanz (vgl. ALDERSON/CLAPHAM/WALL 1995: 107). Das Ausmaß der Subjektivität im Bewertungsprozess ist abhängig vom Aufgabenformat, denn Kurzantworten bieten meist weniger Freiraum und Interpretationsraum als freie essayhafte Textproduktionen (vgl. HUGHES 2003: 22). Im Falle der österreichischen standardisierten Reife- und Diplomprüfung werden die Schreibaufträge mit zu behandelnden Unterpunkten, welche Schreibhandlungen anweisen, versehen und in bestimmten Textsorten angeleitet (vgl. Bericht zur Entwicklung der SRDP: 28 - 118), weswegen von eingeschränkten Schreibaufträgen, *restricted response questions*, gesprochen werden kann, wodurch die Umsetzungsmöglichkeiten eingegrenzt werden (vgl. WAUGH/GRONLUND 2013: 132 f). Einer ausgeprägten subjektiven Einschätzung kann mit Hilfe von Mehrfachbeurteilungen vorgebeugt werden, wodurch diese relativiert werden, allerdings weist dieses Verfahren wenig Transparenz auf (vgl. REZAT 2014: 100 f).

„Bei direkten Messverfahren (Performanztests) werden Lernende im Allgemeinen auf Basis eines subjektiven Urteils eingestuft. Das bedeutet, dass die Entscheidung über die Qualität der Leistung des Lernenden subjektiv gefällt wird, indem relevante Faktoren berücksichtigt werden und man Bezug auf irgendwelche Richtlinien oder Kriterien sowie auf Erfahrungen nimmt. Um der Fairness willen sollte jedoch jede Beurteilung so objektiv wie möglich sein“ (EUROPARAT 2001: 183)

4.2 Bewertungskriterien

Wie bereits erwähnt ergeben sich die Kriterien zur Bewertung aus den gesetzten Standards heraus (DRIESCHNER 2011: 114) und werden benötigt, um die Subjektivität der Beurteilerin beziehungsweise des Beurteilers zu mindern, damit eine möglichst objektive und faire Bewertung stattfinden kann (vgl. EUROPARAT 2001: 183). Neben der Option einer Beurteilung mehrerer EvaluatorInnen eines Textes stellen spezifische Kriterien, anhand derer die Einschätzung eines Textes verübt wird, ein wichtiges praktikables und aussagekräftiges Bezugssystem dar (vgl. BÖRNER ³2001: 293). In Bezug auf die Vieldimensionalität schriftlicher Leistungen ist es notwendig, mehrere Kriterien heranzuziehen (vgl. SACHER 2011: 37 f), welche gänzlich unterschiedliche aber auch teilweise überlappende Bereiche abdecken (vgl. GROTHJAHN/KLEPPIN 2008: 187).

Bewertungskriterien erfüllen eine wichtige Funktion und haben daher gewissen Qualitätsanforderungen zu entsprechen. Zunächst ist es wichtig, dass die SchülerInnenleistung an expliziten und transparenten Kriterien, in einem lehrzielorientierten an vorgegebenen Standards gemessenen Sinne, gemessen werden und nicht rein aufgrund eines sozialen Bezugssystems erfolgen, denn durch eine hohe Transparenz der Bewertungskriterien ist den Geprüften bewusst, welche Anforderungen an sie gestellt werden und zu erfüllen sind, wodurch auch ein erhöhtes Maß an Fairness entsteht. Darüber hinaus gilt es, wie es im Gemeinsamen Europäischen Referenzrahmen für Sprachen beschrieben ist, die Kriterien in positiven Kann-Beschreibungen zu formulieren, um in einem lernförderlichen Prozess, der durch Lernfortschritte einen nicht unbeachtlichen Motivationsfaktor darstellt, SchülerInnenperformanzen anhand dieser Deskriptoren einzuschätzen. Unerlässlich und von besonderer Bedeutung ist eine breite Ausdifferenzierung der zu berücksichtigenden Kriterien. Dabei muss darauf geachtet werden, dass sowohl eine horizontale Gliederung hinsichtlich der inhaltlichen Ebene der Kriterien als auch eine vertikale sich auf die Niveaustufen beziehende Aufgliederung gegeben ist, damit eine schriftliche Leistung umfassend bewertet werden kann und in diesem Prozess auch kleinere Lernfortschritte zum Vorschein kommen (vgl. GROTHJAHN/KLEPPIN 2008: 200 f).

Die Leistungsbeurteilungsverordnung im österreichischen Schulwesen definiert einige fachliche Aspekte für den Fremdsprachenunterricht, die bei der Beurteilung schriftlicher SchülerInnenleistungen zu berücksichtigen sind. Hinsichtlich sprachlicher Fertigkeiten

zählen dazu die idiomatische Ausdrucksweise, grammatische Korrektheit, ein ausreichender Wortschatz sowie Schreibrichtigkeit und die Angemessenheit des Stils und Ausdrucks. Weiters sind die Formvorschriften, das Layout also, und der Inhalt und Aufbau für die Beurteilung maßgebend heranzuziehen (vgl. LBVO § 16 Abs. 1).

In dem bei der standardisierten Reife- und Diplomprüfung verpflichtend zu verwendenden und sonst empfohlenen Beurteilungsraster auf Niveau B1 des BIFIE werden diese fachlichen Aspekte aufgegriffen und in vier Kriterien präsentiert. Diese umfassen die Erfüllung der Aufgabenstellung, den Aufbau und das Layout, das Spektrum sprachlicher Mittel sowie die Sprachrichtigkeit, welche allesamt als gleichwertig in ihrer Gewichtung anzusehen sind. In ihrer vertikalen Ausdifferenzierung umfassen sie für eine ausreichende Streuung elf Niveaustufen, wobei jeweils sechs dieser Ausprägungsgrade Deskriptoren aufweisen (vgl. <https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>). Die ausgewählten Bewertungskriterien zeigen auf, welche Bedeutung den einzelnen Bereichen in beigemessen wird (vgl. GROTHJAHN/KLEPPIN 2008: 199).

Bei ganzheitlichen und qualitativ erfassbaren Leistungen, wie es bei Sprachperformanzen der Fall ist, bedarf es präziser Bewertungskriterien und einer Aufgliederung in Teilleistungen, sodass eine möglichst hohe Objektivität im Zuge der Auswertung geschaffen werden kann (vgl. SACHER 2011: 37 f),

Die festgesetzten Kriterien, welche die Basis für eine an Standards orientierte Leistungsbeurteilung bilden, können in unterschiedlichen Beurteilungsinstrumenten Anwendung finden, wie im folgenden Unterkapitel erläutert wird.

4.3 Beurteilungsraster

Die an den Lehrzielen und Standards orientierten Bewertungskriterien werden in einem Kriterienkatalog zusammengetragen, sodass sie als Beurteilungsinstrument den Evaluationsprozess unterstützen und kriterienbasiert anleiten (vgl. REZAT 2014: 101). Die Auswahl der Kategorien, die dabei berücksichtigt werden, ergibt sich aus den Kompetenzmodellen, wobei zumeist, wie es im österreichischen Schulwesen der Fall ist, der Gemeinsame Europäische Referenzrahmen und dessen Beschreibungssysteme herangezogen werden (vgl. SCHNEIDER/NORTH 2000: 29). BAURMANN geht davon aus, dass eine mittlere Anzahl von Kriterien berücksichtigt werden sollte, die in

Überkategorien zusammengefasst werden, wobei eine Differenzierung und Gewichtung einzelner Kriterien je nach Beurteilungsziel vorzunehmen ist (vgl. REZAT 2014: 101).

Beispielsweise können bei Checklisten Kann-Deskriptoren in einzelne Unterpunkte, die darin enthalten sind, aufgespalten werden, um differenzierte und konkretere Bewertungen vornehmen zu können (vgl. EUROPARAT 2001: 175). Da bei Checklisten jedoch prinzipiell das Hauptaugenmerk darauf gelegt wird, ob ein Kriterium erfüllt oder nicht erfüllt ist, besteht keinerlei Bezug auf eine durch Skalierung aufgefächerte Angemessenheit des Erfüllungsgrades (vgl. BUDDE/RIEGLER/WIPRÄCHTIGER-GEPPERT ²2012: 168). Wird der Gradunterschied beziehungsweise die Intensität sowie Ausprägung eines Merkmals als wichtiges Charakteristikum herangezogen, so findet eine Ordinalskalierung Anwendung (vgl. BENISCHEK 2006 192).

„Le barème est la table de notation élaborée par un évaluateur préalablement à l'évaluation de productions orales ou écrites. Cette table comprend une liste de critères (barème critérié) auxquels l'évaluateur attribue une certaine valeur, un certain nombre de points (barème pondéré) en fonction de ses objectifs d'évaluation. La fonction du barème est d'obtenir plus de fidélité (adéquation) dans les évaluations établies par différents évaluateurs.“ (CUQ 2003: 32).

Als bedeutendes und momentan wichtigstes sowie gängiges Beurteilungsinstrument, in dem Kompetenzenüberprüfungen nach Kriterien und Fähigkeitsdimensionen oder Niveaustufen aufgegliedert präsentiert werden, ist der Beurteilungsraster zu nennen. Diese aus dem angelsächsischen Raum stammenden *rubrics* (vgl. KELLER 2011: 143) hielten in den letzten Jahren immer mehr Einzug in österreichische Schulen und ihr Einsatz ist mittlerweile systematisch verankert (vgl. WINTER 2011: 198), nicht zuletzt durch den zur Beurteilung der standardisierten Reife- und Diplomprüfung entwickelten Raster (vgl. <https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>).

„Une grille est un tableau servant, en fonction de critères sélectionnés, à répertorier, catégoriser, classer des données recueillies sur le terrain. [...] Les grilles d'évaluation : ce sont des tableaux qui comprennent des entrées se référant aux : *compétences de communication de l'apprenant (respect des codes socioculturels attendus, registres de langue); *compétences discursives et textuelles (cohérence et cohésion); *compétences linguistiques (lexique, phonétique et morphosyntaxe). Un barème de notation est fixé pour chacune de ces trois entrées, en fonction de l'objectif que vise le test ou l'examen. Selon l'approche pédagogique et selon le niveau des apprenants, une de ces trois entrées peut être privilégiée.“ (CUQ 2003: 199 f).

4.3.1 Arten von Beurteilungsrastern

Beurteilungsraster können unterschiedliche Formen annehmen abhängig vom Zweck, zudem sie entwickelt wurden. Erstellt werden sie entweder von Einzelpersonen oder ExpertInnengruppen, die versuchen zu definieren, was in einem Lerngebiet unter Berücksichtigung bestimmter Kriterien zu beurteilen ist, und zu beschreiben, was eine gute Leistung darstellt. Unabhängig davon, wer an der Entwicklung eines Beurteilungsrasters beteiligt war, lassen sich prinzipiell drei unterschiedliche Typen hinsichtlich der Intensität der Qualitätsmerkmale unterscheiden. Zunächst sind Primary-Trait-Skalen zu nennen, welche sich auf einen einzigen, klar definierten Aufgabenbereich beziehen. Dabei wird eine Einschätzung hinsichtlich primärer Textmerkmale getätigt, beispielsweise die narrative Struktur eines Textes oder semantische Felder. Aufgrund dieser detailgetreuen Überprüfung einzelner Charakteristika in einem Text ist dieser Typus vergleichsweise kompliziert und aufwändig zu gestalten und findet lediglich in bedeutenden Bildungstests, vor allem in den USA, Anwendung (vgl. KELLER 2011: 144).

Abgesehen von der Betrachtung eines spezifischen Merkmals können Raster die allgemeine Fähigkeit untersuchen, ob und wie sehr ein Text gelungen ist, wobei bezüglich der Auffächerungen in mehrheitliche Dimensionen eine Unterscheidung in holistische und analytische Skalen zu treffen ist (vgl. KELLER 2011: 144 ff). „Ganzheitliche (holistische) Beurteilung führt zu einem globalen, synthetischen Urteil. Verschiedene Aspekte werden vom Beurteilenden intuitiv gewichtet. Analytische Beurteilung betrachtet verschiedene Aspekte jeweils separat.“ (Europarat 2001: 185). Holistische Beurteilungen basieren auf einem allgemeinen Gesamteindruck einer Performanz, die sie dann einer Leistungsstufe zuordnet. Jene Niveaustufen sind mit kurzen, nicht allzu detaillierten Beschreibungen gekennzeichnet, die die wichtigsten Qualitätsansprüche beinhalten und einer Stufe graduiert zuordnen. Dabei wird die Leistung als Gesamtes gesehen und nach bestimmten fachlichen Schlüsselkriterien bewertet²⁴ (vgl. KELLER 2011: 145).

„In the case of writing assessment [...] the reader will read each writing sample quickly for overall impression and place it in one of the piles [...]. Holistic scoring can be guided by scoring rubrics that clarify what each level of quality is like [...] These descriptions of each level do not provide for analysis of the product or performance but simply list the criteria to keep in mind when making overall judgement.“ (WAUGH/GRONLUND¹⁰2013: 164)

²⁴ Ein Beispiel für eine holistische Beurteilungsskala für schriftliche Performanzen ist aufzufinden in: WAUGH/GRONLUND⁽¹⁰2013): 164.

Die zu bewertende Kompetenz wird somit eindimensional ohne feinere Teildimensionen betrachtet, wobei allerdings eine Untergliederung in Teilkompetenzen möglich ist, weshalb mehrere Einzelkriterien zwar zum Einsatz kommen können, diese jedoch nicht aufgefächert berücksichtigt werden. Ein Beispiel hierfür sind die allgemeinen Referenzbeschreibungen des Gemeinsamen Europäischen Referenzrahmens für Sprachen für die einzelnen Niveaustufen (vgl. GROTHJAHN/KLEPPIN 2008: 189). Ohne genauerer Analyse lässt sich jedoch das Fehlen jeglicher diagnostischer Informationen als Kritik äußern, weshalb eine holistische Bewertung bei komplexeren Leistungen als zu wenig dimensional erachtet wird und somit analytische Beurteilungsraster geeigneter wären (vgl. KELLER 2011: 145).

In Abgrenzung zur ganzheitlichen holistischen Bewertung weisen analytische Beurteilungsraster Teildimensionen auf, welche über explizite Beschreibungen verfügen. Der zu messende Kompetenzbereich wird mit verschiedenen voneinander unabhängigen und getrennte Urteilen in seiner Mehrdimensionalität benotet, wodurch sich aufgrund der Betrachtung der einzelnen Beurteilungen abhängig einer gewissen Gewichtung eine endgültige Benotung einer Performanz ergibt. Die Möglichkeit besteht, dass selbst unter Verwendung eines analytischen Rasters ein holistisches Urteil gefällt wird, nämlich dann, wenn Teildimensionen getrennt betrachtet und benotet werden, letztlich aber ein Gesamturteil als gewichtete Summe abgegeben wird, wodurch eine Hierarchie und Gewichtung einzelner Teilkriterien sinnvoll scheint (vgl. GROTHJAHN/KLEPPIN 2008: 190). Diese Mehrdimensionalität analytischer Beurteilungsraster wird formal in einer zweidimensionalen Matrix aufgeschlüsselt, wobei die zu messende Fertigkeit in Teilkompetenzen aufgespalten und weiters in graduierte Niveaustufe mit konkreten Qualitätsbeschreibungen aufgefächert wird. Aufgrund dieser genauen Aufschlüsselung lassen analytische Bewertungsverfahren konkrete diagnostische Informationen über die SchülerInnenleistung zu (vgl. KELLER 2011: 145 f).

4.3.2 Ziele und Funktionen von Beurteilungsrastern

Prinzipiell ist festzuhalten, dass Kriterienraster die Gütekriterien von Testungen hinsichtlich schriftlicher Leistungsbeurteilungen erhöhen wollen und vor allem zum größten Teil zu den Punkten Fairness und Transparenz beitragen. Ziel von einheitlichen Beurteilungsrastern ist es, ein und dasselbe Beurteilungsinstrument für eine Vielzahl an EvaluatorInnen zur Verfügungen zu stellen, sodass in einem großen Kontext nach

denselben Kriterien beurteilt wird und so Leistungen vergleichbarer werden. Den beurteilenden Personen wird durch vorgegebene Raster in gewisser Hinsicht eine Hilfestellung angeboten standardorientiert und kriterienbasiert zu bewerten und zugleich wird die Willkürlichkeit bei Benotungen verringert (vgl. GROTHJAHN/KLEPPIN 2015: 79).

„La grille permet [...] à différents correcteurs/évaluateurs de se référer aux mêmes critères, que la notation se fasse immédiatement après la production ou en différé, des jours, voire des années plus tard. L'évalué doit pouvoir en prendre connaissance avant l'évaluation, de façon à comprendre sur quoi sa performance va être évaluée et donc à pouvoir agir en conséquence.“ (TAGLIANTE 2005: 24)

Beurteilungsraster sind Mittel, die sämtlichen Beteiligten kommunizieren, welche Erwartungen an einen Text gestellt werden. LehrerInnen haben bestimmte Kriterien anhand derer eine Performanz bewertet wird und SchülerInnen wissen im Vorhinein, worauf sie zu achten haben, um eine gelungene Schreibleistung abzuliefern und können im Nachhinein nachvollziehen, wie die Beurteilung zustande kam, weshalb in erster Linie die Transparenz erhöht wird (vgl. KELLER 2011: 146). Studien belegten, dass mit Rastern durchgeführte Beurteilungen nicht nur hilfreich für Lehrpersonen waren, sondern dass auch SchülerInnen förderliche Effekte ausmachen konnten. So werden beispielsweise eine erhöhte Fokussierung auf die eigenen Lernfortschritte, qualitativere Leistungsperformanzen sowie weniger Angst hinsichtlich der Bewertung als positive Auswirkungen genannt. In diesem Zusammenhang wird auch der Terminus *rubric-referenced assessment* verwendet, um die Wirkung des Beurteilungsrasters hervorzuheben (vgl. BROWN/ABEYWICKRAMA 2010: 128). Als bedeutender Ertrag aus diesen lernförderlichen Ergebnissen durch den Einsatz von Rastern ist eine dadurch hervorgehende Motivation seitens der Lernenden zu verzeichnen. Dies geschieht in dem Bewusstsein, welche die nächste erreichbare Stufe in einem Raster ist und woran gearbeitet werden sollte, sodass sie ihre Leistungen in Richtung bessere Beurteilung aktiv trainieren und dahingehend schulen können, weshalb von interner Attribution gesprochen wird (vgl. KELLER 2011: 146 f).

Abgesehen von den personenbezogenen Effekten auf SchülerInnen und LehrerInnen wirken sich Sprachkompetenzskalen allgemein positiv auf Ausrichtung, Organisation und Evaluation des Sprachenlernens aus, denn mit Hilfe eines Referenzsystems werden Lernerfolge für alle Beteiligten weitestgehend informativ beschrieben und transparent offengelegt. (vgl. SCHNEIDER/NORTH 2000: 18 f).

Neben den bislang erwähnten positiven Auswirkungen eines analytischen Rasters als Beurteilungsinstrument, bleibt anzumerken, dass auch hierbei Nachteile zu vermerken sind:

„Indem sie zentrale Beurteilungsbereiche für Leistungen festlegen, können Raster als „Schablone der Wahrnehmung“ wirken und den Blick für individuelle, kreative oder unerwartete Aspekte von Schülerleistungen einengen. [...] Eine Auswahl an Kriterien kann nicht dieselbe „Beurteilungskraft“ haben wie die gesamte Population von Kriterien, über welche Lehrkräfte latent verfügen und welche für die Bewertung einer komplexen Leistung oft auch nötig sind.“ (KELLER 2011: 148)

Einerseits ist die Gefahr vorhanden nicht berücksichtigte Kriterien aus der Beurteilung gänzlich auszuklammern oder gegenteilig, dass aufgrund der Bewertungserfahrung ein holistisch gewonnener Eindruck an das Schema des Rasters angepasst wird, um ein erwünschtes Resultat zu erhalten, wodurch die kriteriale Bezugsnorm sowie die Objektivität, die solche Raster gewähren sollen, untergraben (vgl. KELLER 2011: 148).

Wichtig für einen funktionalen, zweckdienlichen Umgang mit Rastern ist die Qualitätssicherung und deren didaktische Anwendung. Aus diesem Grund ist es absolut notwendig, dass eine Orientierung der Beurteilungsraster an übergeordneten Bildungszielen, wie Bildungsstandards, auszumachen ist. Weiters müssen ausreichend Kriterien vorhanden und in genügend Niveaustufen unterteilt sein. Darüber hinaus ist eine präzise Formulierung der möglichst positiven Kann-Beschreibungen unerlässlich, da aufgrund unklarer und nicht prägnanter Sprache Fehlinterpretationen zustande kommen, worunter einige Gütekriterien leiden (KELLER 2011: 147 f). Eben jene Problematik der Fehlinterpretationen und der Interpretationsobjektivität soll im folgenden Unterkapitel erläutert werden.

4.3.3 Problematik der Interpretationsobjektivität

Wie vorangehend beschrieben wurde, ist es unerlässlich, dass Deskriptoren in klarer und präziser Sprache formuliert werden, um keinen allzu großen Interpretationsraum zu bieten. Allerdings ist hinsichtlich einer Graduierung und eines Ausprägungsgrades eben hier ein Dilemma zu verorten, da sich Abstufungen in der Bewertung produktiver Fremdsprachekompetenz nicht als objektiv präzise definieren lassen. Auch wenn der Versuch vorliegt, möglichst systematische Formulierungen in Deskriptoren zu verpacken, wirkt dies wie eine „Ersetz“-Funktion in der Beschreibung, da nur die Graduierung verändert wird. Solche Feinheiten bezüglich der Unterschiede bei Formulierungen lassen

einen erhöhten Interpretationsraum zu und machen die Bedeutung und ihr Verständnis oftmals von äußeren Faktoren abhängig. (vgl. SCHNEIDER/NORTH 2000: 21 ff)

„Die Sachverhalte, die einer Ordinalskala zugänglich sind, gleichen einander hinsichtlich des betrachteten Merkmals, sie unterscheiden sich jedoch voneinander bezüglich der Merkmalsausprägung. Es bleibt allerdings undefiniert, wie stark die Differenzen zwischen den einzelnen Objekten sind.“ (BENISCHEK 2006: 192)

Ein unterschiedlicher Gebrauch von Beurteilungsinstrument, wie auch von analytischen Rastern, lässt sich konstatieren (vgl. GASSNER/MEWALD/SIGOTT 2008: 15), da die Deskriptoren nicht durchwegs gleich aufgefasst werden (vgl. ALDERSON/CLAPHAM/WALL 1995: 107 ff).

„Wie in den einführenden Bemerkungen erwähnt, sei unter sprachlicher Graduierung die Charakterisierung von Objekten und Prozessen bzw. Handlungen nach dem Grad eines zugrunde liegenden Merkmals verstanden. Diesem Prozeß [sic] liegt ein Vergleich zugrunde, dessen Resultat die Feststellung von Gleichheit bzw. Ungleichheit ist. Zwischen den beiden Polen der Gleichheit und Ungleichheit liegt eine ganze Skala von Abstufungen, die durch verschiedene sprachliche Mittel erfaßt [sic] werden und zu vielfältigen Nuancierungen in der Sprache führen.“ (WINGENDER 2005: 48 f)

Formulierungen, die eine Graduierung auf qualitativer Ebene ausdrücken wollen, sind somit vielfältig realisierbar und nicht quantitativ erfassbar sowie klar abgrenzbar, wodurch eine relativierbare Auffassung dieser entsteht (vgl. WINGENDER 2005: 48 ff). KERBRAT postuliert eine doppelte Subjektivität solcher Graduierungswörter. Zum Einsatz kommt einerseits eine Relativierung der bezeichneten Formulierung an sich und andererseits die Relativierung durch den Betrachter (vgl. RIVARA 1990: 94). Letztere kann im Kontext Schule durch den Einfluss des Wirkungsraumes, der Schule, bedingt werden, beispielsweise könnte eine Lehrkraft durch bestimmte Ansprüche in einer Schule beziehungsweise in einem Lehrerkollegium die Leistungserwartungen erhöhen oder senken. In diesem Fall fällt der Terminus Referenzrahmeneffekt, denn eine Urteilsveränderung ist auf den Rahmen, in dem agiert wird, zurückzuführen. Diesbezüglich können empirisch untersuchte Ergebnisse zu schulartenspezifischen Beurteilungsmaßstäben vorgelegt werden (vgl. MAIER 2015: 96 f).

Diese Problematik bedingt zum Teil, dass ein und dieselbe SchülerInnenperformanz unter Umständen sehr unterschiedliche Beurteilungen abhängig von der evaluierenden Person erhalten. Eine SchülerInnenleistung sollte allerdings kriterienabhängig und nicht

personenabhängig sein, weshalb eine gewisse Verlässlichkeit unterhalb der Beurteilenden gegeben sein sollte²⁵ (vgl. ALDERSON/CLAPHAM/WALL 1995: 128 f).

„*Inter-rater reliability* refers to the degree of similarity between different examiners: can two or more examiners, without influencing one another, give the same marks to the same set of scripts or oral performances? It would not be realistic to expect all examiners to match one another all the time; however, it is essential that each examiner try to match the ‚standard‘ all the time.“ (ALDERSON/CLAPHAM/WALL 1995: 129)

Dass bei subjektiven Bewertungsmethoden keine völlige Übereinstimmung bei evaluatorInnenübergreifenden Beurteilungen stattfinden kann, ist nur realistisch. Es scheint legitim, dass SchülerInnenleistungen auf teilweise unterschiedliche Einschätzungen stoßen können, dennoch gilt es, einen höchstmöglich geringen Unterschied, im Sinne einer standardorientierten, kriterienbasierten Beurteilungskultur, anzustreben. Von großer Bedeutung ist es jedoch, dass Deskriptoren individuell als beständig und durchwegs gleich aufzufassen sind. Ein und dieselbe Person sollte über mehrere unterschiedliche Performanzen hinweg eine Graduierung nicht anders einordnen und interpretieren, sodass im Idealfall jeder Evaluator und jede Evaluatorin für sich selbst die Graduierungen fixiert und eine Performanz in mehreren Durchgängen immer möglichst gleich beurteilt, wodurch eine *intra-rater-reliability* gegeben ist (vgl. ALDERSON/CLAPHAM/WALL 1995: 107 ff).

Um prinzipiell die Gütekriterien in Bezug auf subjektive Beurteilungsverfahren bestmöglich garantieren zu können, bedarf es einer Aufgabenentwicklung durch ExpertInnen, ein Bewusstsein für die Prüfung bei den Getesteten und seitens der EvaluatorInnen eine ausreichende Professionalisierung²⁶ im Umgang mit Beurteilungsrastern und ihren Deskriptoren²⁷ (vgl. ALDERSON/CLAPHAM/WALL 1995: 105 ff).

²⁵ Möglichkeiten eine beurteilendenübergreifende Reliabilität zu gewährleisten, wären beispielsweise Mehrfachbeurteilungen oder Beispielzuordnungen von Leistungen an bestimmte Niveaustufen. Verweis auf: ALDERSON/CLAPHAM/WALL (1995): Kapitel 6: Monitoring examiner reliability. S. 128 - 147.

²⁶ Verweis auf die Wichtigkeit des Trainings Beurteilender fremdsprachlicher Prüfungen: ALDERSON/CLAPHAM/WALL (1995): Kapitel 5: The training of examiners and administrators. S. 105 - 127.

²⁷ In diesem Zusammenhang ist auf die im Zuge einer Dissertation stattgefundene Forschung an der Universität Wien zu analytischen Beurteilungsskalen und deren Deskriptoren hinzuweisen: BERGER (2014).

4.4 Beurteilungsraster des BIFIE auf Niveau B1

Nach erfolgter Abhandlung zur Leistungsbeurteilung schriftlicher Performanzen im Fremdsprachenunterricht und einer Beleuchtung des Bewertungsinstrumentes eines Rasters, soll nachstehend der im österreichischen Französischunterricht Anwendung findende Beurteilungsraster für schriftliche SchülerInnenleistungen auf Sprachniveau B1 unter den bislang dargelegten Gesichtspunkten betrachtet werden.

Der im Annex (auf Seite 113) aufzufindende Beurteilungsraster wurde von ExpertInnen im Bereich der Fremdsprachenevaluation vom Language Testing Center der Universität Klagenfurt in Kooperation mit dem BIFIE entwickelt und erstellt. Zwingend zum Einsatz kommt er bei der Bewertung von Performanzen bei der standardisierten Reife- und Diplomprüfung im Unterrichtsfach Französisch (vgl. <https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>). Eben jene Leistungsfeststellung, die im Schulwesen intern durchgeführt wird, weist einen formellen, punktuell stattfindenden, summativen Bewertungscharakter auf (vgl. DECKE-CORNILL/KÜSTER³2015: 259 ff).

Da es sich bei den SchülerInnenleistungen um individuelle Sprachproduktionen in offenen Aufgabenformaten handelt, sind unterschiedliche, vielfältige Performanzen gegeben, weshalb eine subjektive Einschätzung des Lehrkörpers stattfindet, denn eine persönliche Bewertung des Erfüllungsgrades der Aufgabenstellung in einer vorliegenden Performanz ist gefragt (vgl. ALDERSON/CLAPHAM/WALL 1995: 107).

Der hier behandelte Raster wurde ohne umfassende Begleitliteratur wissenschaftlicher Natur veröffentlicht, weshalb im Folgenden eine deskriptive Arbeit des Verfassers stattfinden wird. Dieser vom BIFIE für Lehrkräfte veröffentlichte Beurteilungsraster versucht nicht, einen Globaleindruck zu initiieren, sondern eine analytische Beurteilung nach unterschiedlichen Gesichtspunkten zu veranlassen. Spezifische Kriterien stellen ein Bezugssystem dar, anhand dessen, an den Standards des GERS orientiert, kriterienbasiert bewertet werden soll. Diesbezüglich unterteilt sich die zweidimensionale Matrix horizontal in vier voneinander unabhängige Kriterien, wobei jedes Kriterium vertikal in elf Niveaustufen graduiert aufgefächert wurde, sodass der Versuch, die Viel-dimensionalität einer schriftlichen Leistung abzudecken, unternommen wurde. Sechs dieser Stufen weisen Beschreibungen auf, die hierarchisch angeordnet sind und die wichtigen Merkmale des übergeordneten Kriteriums abdecken sowie versprachlichen (vgl. <https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>).

Die vier Kriterien des Beurteilungsrasters decken die Erfüllung der Aufgabenstellung, den Aufbau und das Layout, das Spektrum sprachlicher Mittel sowie die Sprachrichtigkeit ab (vgl. [https://www.srdp.at/ downloads/dl/ beurteilungsraster-b1-und-begleittext/](https://www.srdp.at/downloads/dl/ beurteilungsraster-b1-und-begleittext/)).

„Es wird deutlich, dass in den neuen Orientierungsangaben der Schwerpunkt von Sprachrichtigkeit hin zum kommunikativen Erfolg verlagert wurde. Dies ist für die Praxis des Fremdsprachenunterrichts, in der bei der Leistungsbewertung Fehler immer noch in erster Linie als ein Indikator für mangelnde Leistung betrachtet werden, sicherlich eine beträchtliche Umorientierung. [...] Doch ein niedriger Fehlerquotient zeigt nicht unbedingt an, dass etwas „gekonnt“ wird.“ (GROTJAHN/KLEPPIN 2008: 194)

Die Bewertungskriterien des Beurteilungsrasters werden also einem kommunikativen, handlungsorientierten Französischunterricht gerecht und beziehen sich auf eine zielgerichtete, situationsadäquate Kommunikationsfähigkeit, die oberstes Ziel im Fremdsprachenunterricht ist (vgl. Lehrplan HAK 2014: 2 ff).

In der vertikalen Dimension des Beurteilungsrasters sollen nun einige Graduierungen von Deskriptoren exemplarisch herangezogen und näher analysiert werden. Grundsätzlich lässt sich feststellen, dass nur zwei Beschreibungen quantitativen Charakter besitzen, wobei es sich bei einem Deskriptor um keine Graduierung handelt, da die Länge des zu produzierenden Textes nicht abgestuft wird, sondern als erfüllt oder nicht erfüllt anzusehen ist. Einziger Punkt der eine teils quantitative Abstufung zulässt, ist bei der Erfüllung der Aufgabenstellung zu finden und verweist auf die Anzahl der behandelten von den drei aufgetragenen Inhaltspunkten. Sämtliche anderen Deskriptoren beinhalten in ihrer Ausprägung qualitative Graduierungen, die sich zumeist in „Ersetz“-Formulierungen auffinden lassen (vgl. <https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>), wodurch eine subjektive Interpretation der Merkmalsausprägung der den Deskriptoren zugrunde liegenden Fertigkeiten verlangt wird. Die Kriterien werden zwar, wie für einen analytischen Raster gefordert, explizit und genau aufgegliedert, sind allerdings aufgrund der genannten Gradpartikeln nicht gänzlich objektiv wahrnehmbar beschrieben und gekennzeichnet, weshalb die Einschätzung einer SchülerInnenleistung von der Evaluatorin beziehungsweise dem Evaluator in eine Relation, möglicherweise die individuelle Auffassung des Ausprägungsgrades und der Formulierung des Deskriptors oder in ein soziales Bezugssystem, gesetzt werden muss.

Nachstehend sollen einige Formulierungen aus dem Beurteilungsraster präsentiert werden, um den meist qualitativen Charakter, der einer Relation bedarf (vgl. RIVARA 1990 :94), aufzuzeigen. Diesbezüglich wird auf die Niveaustufen im positiven Bereich (Band 6 – 10) verwiesen, da die Ausprägungsstufen hier sprachlich eher als unklar formuliert aufzufassen scheinen und für die positive Beurteilung einer Performanz von Wichtigkeit sind. Weiters wird in den unteren Beurteilungsstufen einige Male mit Negationen gearbeitet, wodurch ein Ausschlusskriterium vorhanden ist und der individuelle Interpretationsraum verringert wird (vgl. [https://www.srdp.at/ downloads/dl/ beurteilungsraster-b1-und-begleittext/](https://www.srdp.at/downloads/dl/ beurteilungsraster-b1-und-begleittext/)).

Im Kriterienbereich Erfüllung der Aufgabenstellung beschreibt der zweite Deskriptor jene Merkmale, welche im Text erfüllt sein sollen, wobei hier die von SCHNEIDER und NORTH (2000) kritisierte „Ersetz“-Formulierung durch die Ausprägungsgrade Anwendung findet.

Erfüllung der Aufgabenstellung: Deskriptor 2	
Band 10	„Formuliert Titel/Betreff/Abschnittsüberschriften/Anrede/Grußzeile <u>treffend</u> “
Band 8	„Formuliert Titel/Betreff/Abschnittsüberschriften/Anrede/Grußzeile <u>gut</u> “
Band 6	„Formuliert Titel/Betreff/Abschnittsüberschriften/Anrede/Grußzeile <u>angemessen</u> “

([https://www.srdp.at/ downloads/dl/beurteilungsraster-b1-und-begleittext/](https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/))

Die hier den unterschiedlichen Niveaustufen zugeordneten und ersetzten Graduierungen *treffend*, *gut*, *angemessen* sind über fünf Abstufungen hinweg eingesetzt worden, wobei laut Bedeutungswörterbuch *treffend* und *angemessen* als synonym gebräuchlich anzusehen sind und im Sinne von „genau richtig“ aufgefasst werden (vgl. <http://www.duden.de/suchen/dudenonline/treffend>), wodurch hier sprachliche keine Abstufung stattfindet, sondern nur andere Formulierungen für eine semantische Größe zu tragen kommen. *Gut* wird als bestimmten Normen entsprechend verstanden und ist ebenso mit oben genannten Worten sinnverwandt (vgl. DUDENREDAKTION ²1985: 315). In der Funktion einer Fokuspartikel, unterstreicht *gut* eine Größe, modifiziert sie in einem semantischen Sinne allerdings nicht. Fokussierer und Graduierer akzentuieren die Größe ihrer Bezugskonstituente und lassen keine scharfe Abgrenzung zu, weshalb sie als ineinander übergehend zu betrachten sind (vgl. GLÜCK ⁴2010: 247), wobei eine weitere exakte linguistische Definition für die vorliegende Arbeit wenig notwendig scheint.

Erfüllung der Aufgabenstellung: Deskriptor 8	
Band 10	„Hält die vorgegebene Wortanzahl ein“
Band 8	„Hält die vorgegebene Wortanzahl ein“
Band 6	„Hält die vorgegebene Wortanzahl ein“

(<https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>)

Hierbei bleibt die Formulierung ident, denn eine Abstufung ist, wie im Begleittext des Beurteilungsrasters auch erörtert wird, nicht möglich (vgl. <https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>), wobei die Frage offen bleibt, mit welcher Punkteanzahl eine solche SchülerInnenleistung zu beurteilen ist. Augenscheinlich liegt dies in der Verantwortung der Lehrperson.

Im Kriterienbereich Aufbau und Layout verweist der zweite Deskriptor auf die Kohärenz des Textes, wobei der Ausprägungsgrad mit dem Gradpartikel *sehr* und dem Adverb *einigermaßen* angegeben wird.

Aufbau und Layout: Deskriptor 2	
Band 10	„Präsentiert die inhaltlichen Punkte <u>sehr klar</u> und <u>systematisch</u> “
Band 8	„Präsentiert die inhaltlichen Punkte <u>klar</u> und <u>systematisch</u> “
Band 6	„Präsentiert die inhaltlichen Punkte <u>einigermaßen</u> und <u>systematisch</u> “

(<https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>)

Werden die Abstufungen dieses Deskriptors verglichen, ist eindeutig eine Skalierung mit Bedeutungsunterschieden erkennbar. Allerdings sind die Ausprägungen keinen exakten Referenzwerten zuzuordnen, weshalb die Höchstpunktzahl, Band 10, die Erweiterung von Band 8 mit dem Adverb *sehr* darstellt und ein hohes Maß fordert (vgl. DUDENREDAKTION ²1985: 577), vergleichsweise ein höheres Maß als Band 8 also. In Band 6 findet eine Abstufung mit dem Adverb *einigermaßen* statt, wodurch sich wieder eine relative Verkleinerung zu Band 8 feststellen lässt. Aus diesem Grund bedarf es einer subjektiven Auffassung und Relation der beurteilenden Person, wodurch ein soziales Bezugssystem für Vergleiche als mögliche Referenz dienen könnte und aussagekräftiger scheint als eine rein kriteriale Zuordnung.

Das Spektrum sprachlicher Mittel wird ebenfalls mit qualitativen Adjektiven abgestuft.

Spektrum sprachlicher Mittel: Deskriptor 1	
Band 10	„Hat ein breites Spektrum an lexikalischen und strukturellen Mitteln“
Band 8	„Hat ein <u>ausreichend</u> breites Spektrum an lexikalischen und strukturellen Mitteln, um sich mit hinreichender <u>Genauigkeit</u> ausdrücken zu können“
Band 6	„Hat ein <u>genügend</u> breites Spektrum an lexikalischen und strukturellen Mitteln, um die Aufgabe erfüllen zu können.“

(<https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>)

Interessant hierbei scheint, dass Band 10 keinen erhöhten Anforderungsbedarf von Band 8 darstellt, sondern von ihm ausgehend Abstufungen vorzufinden sind. In den beiden herabgesetzten Deskriptoren werden *ausreichend* und *genügend* herangezogen, um eine Graduierung auszudrücken, wobei beide als synonym anzusehen sind und somit keinen Bedeutungsunterschied erzielen (vgl. DUDENREDAKTION ²1985: 100/292), weshalb Band 6 und Band 8 mit einer Präzisierung versehen wurden, um zu verdeutlichen, wozu das Spektrum angemessen breit ist. Hierbei ist der Begriff *Genauigkeit* von zentraler Bedeutung, wobei dies weniger als exakte Größe gilt, weswegen eine Punktevergabe in der subjektiven Verantwortung der BeurteilerInnen liegt und somit der intrapersonellen Reliabilität eine enorme Bedeutung zukommt, sodass EvaluatorInnen für sich eine durchwegs konstante Auffassung dieser subjektiven Interpretationsräume vereinbaren (vgl. ALDERSON/CLAPHAM/WALL 1995: 107 ff).

Diese exemplarisch ausgewählten, präsentierten und kommentierten Formulierungen von Deskriptoren in ihrer Abstufung, die keinen Anspruch auf eine umfassend exakte, linguistische Analyse erheben, sollten anhand des Beurteilungsrasters des BIFIE veranschaulichen, wie die Problematik der Interpretationsobjektivität und der subjektiven Relation von Beurteilenden in der Praxis zu verstehen ist. Dieses Element soll unter anderem im empirischen Teil bei unterrichtenden Französischlehrpersonen und somit praktizierenden Evaluatorinnen und Evaluatoren erforscht werden, worauf eine Präsentation der Ergebnisse, die mit der bislang geleisteten theoretischen Arbeit in Beziehung gesetzt werden sollen, folgen wird.

5 Zusammenfassung

Wie in der vorangegangenen theoretischen Kontextualisierung dargeboten wurde, handelt es sich bei Assessment, um einen äußerst weitläufigen Bereich, der unterschiedliche Subdisziplinen in sich vereint und miteinander in Beziehung setzt. Nicht nur im alltäglichen Leben ist der Bewertung und Klassifizierung sämtlicher Situationen und Gegebenheiten nach gewissen Kriterien eine tragende und bedeutende Rolle zuzuweisen, auch im schulischen Kontext wird dadurch eine Vielzahl an unerlässlichen Funktionen erfüllt, ohne die das System Schule nicht auskommen kann. Damit eine solch umfassende Menge an zweckorientierten Methoden zielgerichtet zum Einsatz kommen kann, gilt es auf verschiedene, sich unter einander abgrenzende Formen des Assessments zurückzugreifen, um sich sozusagen einem Methodenpluralismus hinsichtlich des Bewertens zu bedienen. Eben jene Methoden folgen bestenfalls Gütekriterien, die darauf achten, dass in einem wissenschaftlichen Verständnis qualitativ hochwertig, sprich verlässlich, gültig und sachlich, lernförderlich sowie fair und transparent beurteilt wird, was im Fremdsprachenunterricht, bei dem eine Fülle an Kompetenzen aktiv sind, eine Herausforderung darstellt.

Diesbezüglich wurde informiert, was das Assessment im fremdsprachlichen Unterricht zu messen und bewerten versucht: die Leistung der Lernenden. Hierbei wurde konkreter Bezug auf den österreichischen Fremdsprachenunterricht im Schulkontext genommen und aufgezeigt, wodurch sich die Leistungsfeststellung und Leistungsbeurteilung unterscheiden, wobei darauf verwiesen wurde, dass die Kompetenzen, die einer Schülerin beziehungsweise einem Schüler in ihrer Gesamtheit zur Verfügung stehen nicht festgestellt werden können, sondern lediglich nur die gezeigt zum Vorschein kommende Realisierung, eine Performanz. Jegliche Bewertung einer Leistung unterliegt der Bezugnahme zu einer bestimmten Norm, da sie stets in einem Verhältnis zu einem Bezugssystem gesehen wird. Welche Maßstäbe herangezogen werden, obliegt der/den beurteilenden Person/en: „Was als gute Leistung zählt, [...] liegt in unserem Schulsystem weitgehend in der Hand der Lehrer/innen.“ (Stern 2010: 25)

Eine Vorbeugung gegen beurteilende Willkür repräsentieren Standardisierungsprozesse. Bildungsstandards, die Ziele formulieren, welche an einem gewissen Zeitpunkt in der Schullaufbahn zu erreichen sind, vereinheitlichen einerseits die Lehrziele, die erfüllt werden sollten, sowie andererseits im selben Schritt die Maßstäbe, die zur Bewertung

heranzuziehen sind, denn durch die Implementierung von Standards werden zeitgleich Kriterien als Bezugssystem gesetzt, die einheitlich zu berücksichtigen sind.

In Bezug auf die schriftliche Produktion bedurfte es einer Eingrenzung der möglichen Realisierungen im Sinne eines Textsortenkataloges, wobei erforderliche Ansprüche an jede Textsorte definiert und beschrieben wurden, sodass Konsens über die Qualitätsmerkmale einer gelungenen SchülerInnenperformanz herrscht. Bei der Überprüfung im Zuge der standardisierten Reife- und Diplomprüfung ist es notwendig, bereits erwähnte Gütekriterien von Testungen zu erfüllen, damit auch hier in höchstmöglichen Maße fair, transparent, verlässlich, valide und sachlich geprüft werden kann, weshalb zentral erstellte standardisierte Aufgabenstellungen zum Einsatz kommen. Hinsichtlich der Beurteilung ist ebenfalls ein zentrales, an den Standards orientiertes und auf den daraus resultierenden Kriterien basiertes Beurteilungsinstrument unabdingbar.

In diesem Zusammenhang fand eine Auseinandersetzung mit Rastern statt, wobei deren unterschiedliche Arten, Funktionen und Zielverfolgungen sowie Herausforderungen und Problematiken erläutert wurden. Im Fokus standen, in Abgrenzung zu holistischen, analytische Beurteilungsraster, die versuchen, die Mehrdimensionalität von schriftlichen Leistungsperformanzen durch die Unterteilung in verschiedene Kriterien auf einer horizontalen Ebene sowie die Auffächerung dieser in unterschiedliche Niveaustufen auf vertikaler Ebene in einer zweidimensionalen Matrix einzufangen. Dadurch sind allen Beteiligten die Kriterien, anhand welcher bewertet wird, bekannt, wodurch die Transparenz der Leistungsbeurteilung um ein Vielfaches erhöht ist. Qualitätsmerkmal eines den Gütekriterien gerecht werdenden analytischen Rasters ist eine präzise und klare Formulierung der Deskriptoren, um mögliche Interpretationsräume möglichst gering zu halten, da der Wunsch nach hoher Objektivität trotz offener Aufgabenformate verfolgt wird. Doch eine Bewertung von Schreibleistungen unterliegt trotz allem einer gewissen subjektiven Auffassung der Evaluierenden.

In Vorbereitung auf die empirische Forschung sind diesbezüglich einige Deskriptoren des Beurteilungsrasters des BIFIE für das Sprachniveau B1, soweit es im Rahmen dieser Diplomarbeit möglich schien, exemplarisch herangezogen und auf ihre Qualitätsmerkmale sowie Formulierungen hinsichtlich Interpretationssubjektivität durch Beurteilende untersucht worden. Wie aufgezeigt wurde, liegt die tragende Verantwortung bei den Lehrerinnen und Lehrern, weshalb diese in den Mittelpunkt der empirischen Forschung rücken und ihre Auffassungen und Einschätzungen analysiert werden sollen.

III. Empirische Untersuchung

In diesem Teil der Arbeit soll nach einer bereits erfolgten Auseinandersetzung mit den theoretischen Gegebenheiten des Untersuchungsgegenstandes eine empirische Annäherung an die Materie stattfinden. Diesbezüglich wird nachfolgend das Forschungsdesign mit einer Beschreibung und Begründung der Methodenwahl erfolgen, bevor in einem weiteren Abschnitt die erlangten Resultate aufgearbeitet, präsentiert sowie analysiert und mit der vorangehend dargebotenen Theorie in Verbindung gesetzt werden.

6 Forschungsdesign

Interessant für die Forschung in diesem Gebiet sind wie erläutert die aktiv Beurteilenden und deren Einstellungen zur Leistungsbeurteilung sowie Handhabung und Interpretation des zu verwendenden Rasters, weshalb auch sie in den Fokus der Empirie rücken. Damit eben jene individuellen Attitüden erfasst werden können, bedarf es als Ausgangsbasis einer Bewertung einer schriftlichen SchülerInnenperformanz mit Hilfe des Beurteilungsrasters. Darauf aufbauend sollen Instrumentarien der qualitativen Sozialforschung zum Einsatz kommen und die herangezogenen Lehrpersonen auf der Grundlage eines Leitfadeninterviews zur Leistungsbeurteilung schriftlicher Performanzen allgemein sowie zum Beurteilungsraster befragt werden. Diese Ergebnisse sollen unter Anwendung der qualitativen Inhaltsanalyse geordnet, erklärt und resümiert werden, wobei folgend die erhobenen Ergebnisse präsentiert und diskutiert werden.

6.1 Forschungsfragen

Die den Ausgang für diese Diplomarbeit bildende Frage „*Wie lassen sich die Einstellungen Französischlehrender dem Beurteilungsraster des BIFIE auf Sprachniveau B1 gegenüber beschreiben?*“ soll dank der Erhebung von Daten in diesem Teil der Arbeit präsentiert werden. Aus dieser forschungsleitenden Fragestellungen ergaben sich zwei weitere Unterfragen, die sich aus der forschungsleitenden Problemstellungen ableiten. Einerseits scheint unter Berücksichtigung der heranzuziehenden Maßstäbe bei der Leistungsbeurteilung interessant, zu beantworten: „*Wie verständlich und klar erscheinen Fremdsprachen-lehrenden die Formulierungen der Deskriptoren?*“ sowie andererseits aus einer praktischen Perspektive: „*Wie findet der Beurteilungsraster Anwendung bei der Leistungsbewertung schriftlicher Performanzen im Französischunterricht?*“.

6.2 Methodenwahl

Wie bereits dargelegt scheint im Zuge des vorliegenden Forschungsinteresses die Wahl einer qualitativen Methode sinnvoll. Hierbei stehen im Unterschied zu quantitativen Ansätzen keine Hypothesen und Operationalisierungen im Vordergrund, denn die Theorie bildet den Endpunkt, der durch die Empirie entwickelt werden soll. Zielsetzung ist also, aus den erhobenen Daten neue Erkenntnisse zu ziehen und daraus Hypothesen oder Theorien abzuleiten (vgl. FLICK 2009: 24 ff). Bei den in diesem Rahmen erlangten Daten handelt es sich um verbale Daten, welche keine quantifizierten Messwerte darstellen, sondern beschreibendes Material repräsentieren, aus denen ein Induktionsschluss erfolgt (vgl. DÖRING/BORTZ 2006: 297 ff). Auch wenn der Standardisierung bei qualitativen Forschungen keine große Bedeutung zukommt, kann je nach Forschungsdesign eine Unterteilung in straffe und lockere Vorgehensweisen vorgenommen werden. Ersteres zeichnet sich durch begrenzte Fragestellungen und bewusst festgelegte Auswahlverfahren aus (vgl. FLICK 2009: 206). Im Folgenden wird erläutert, wie die Daten zu erheben sind und auf welchen Methoden die empirische Forschung basiert, wodurch erkenntlich wird, dass ein straff gehaltenes Forschungsdesign im qualitativen Sinne vorliegt.

6.2.1 Leistungsbeurteilung einer Performanz

Grundlage und Einstieg in die empirische Forschung bildet die Leistungsbeurteilung einer schriftlichen SchülerInnenperformanz (siehe Anhang Seite 111). Diese authentische Schreibleistung entstand ausgehend von einer standardisierten Aufgabenstellung (aufzufinden im Annex auf Seite 110) des BIFIE. Die Auswahl der Aufgabe ergab sich einerseits durch die Textsorte, die in sämtlichen österreichischen Schultypen auf Sprachniveau B1 zum Einsatz kommt sowie ihren expliziten Ansprüchen hinsichtlich Aufbau und Layout, damit bei der Befragung der Beurteilenden auf extern vorgegebene Qualitätsmerkmale in Abgrenzung zu den eigens einzuschätzenden sprachlichen Qualitätsmerkmalen verwiesen werden kann. Andererseits ist zu berücksichtigen, dass es sich um eine standardisierte Aufgabenstellung des BIFIE handelt, wodurch davon ausgegangen werden kann, dass sie höchstmöglich den Gütekriterien einer angemessenen Aufgabenstellung des Sprachniveaus B1 laut Referenzrahmen darstellt.

Die authentische Schülerleistung entstammt der Feder eines Schülers, der laut Schulcurriculum Sprachniveau B1 erreicht haben sollte und wurde aus Personenschutzgründen anonymisiert. Darüber hinaus wurde die Schreibleistung in getippter Form

zur Beurteilung gegeben, da formale Faktoren²⁸ wie das Schreibbild bereits die Leistungsbeurteilung beeinflussen können und dies aber verhindert werden soll. Hinweise auf das Geschlecht des Lernenden konnten durch Verwendung von geschlechtsanzeigenden Adjektiven nicht verschleiert werden, weshalb allerdings der Name des Schülers durch *Jean Dupont*²⁹ ersetzt wurde, denn auch dies könnte Einfluss auf die Lehrpersonen nehmen (vgl. REZAT 2014: 100).

Diese herangezogene Performanz wurde aus mehreren Gründen ausgewählt. Hauptgrund war, dass im Sinne einer Handlungsorientierung die Aufgabenstellung erfüllt und das kommunikative Ziel erreicht wurde, allerdings in keiner überdurchschnittlichen Art und Weise. Dies war von Bedeutung, denn eine überdurchschnittlich leistungsstarke oder leistungsschwache Performanz wäre im Sinne des Forschungsinteresses wenig aufschlussreich. Eine durchschnittliche Leistung lässt hierbei allerdings genau den Interpretationsraum der Deskriptoren Anwendung finden, der im Zuge dieser Diplomarbeit und empirischen Forschung unter anderem näher beleuchtet werden soll. Weiters sind häufige sowie teils als gravierend zu bezeichnende Fehler hinsichtlich der Sprachrichtigkeit aufzufinden, die jedoch die Handlungsorientierung nicht schmälern, woraus sich eine spannende Grundlage zur Untersuchung ergeben könnte. Deshalb wurde diese spezielle Schreibleistung im Sinne einer bewussten Auswahl, planvoll aufgrund der genannten Überlegungen, wie die dargelegten Kriterien untermauern, hinsichtlich dieses bestimmten Forschungszweckes herangezogen (vgl. KROMREY 2009: 265 f).

Diese durchgeführte Leistungsbeurteilung stellt einen Einstieg in die Untersuchung dar und erlaubt im Zuge der darauf aufbauenden Interviews, einen konkreten Bezug zu einer erfolgten Bewertung mittels Beurteilungsrasters zu nehmen und die vorliegende Einschätzung und Benotung mit der Lehrperson zu besprechen und zu interpretieren. „Bei qualitativen Methoden meint Interpretation vor allem die Analyse von Äußerungen in Interviews oder von bestimmten Ereignissen oder Handlungen [...].“ (FLICK 2009: 73)

²⁸ Der sogenannte Halo-Effekt, oder Ausstrahlungseffekt, bezeichnet unter anderem eine Auswirkung von persönlichen Eigenschaften von Lernenden auf die Leistungsbeurteilung (vgl. REZAT 2014: 100).

²⁹ Jean Dupont gilt als Musternamen des französischen Durchschnittsbürger und ist das Äquivalent zu Max Mustermann (vgl. ROLOFF 2014: 56 f) und bietet somit höchstmögliche Neutralität

6.2.2 Interviews

Als Datenerhebungsmethode im Zuge dieser Forschung soll ein problemzentriertes Leitfadenterview stattfinden, denn „Ziel ist es, die individuelle Sicht des Interviewpartners auf das Thema zu erhalten, wozu ein Dialog zwischen Interviewer und Interviewten mit den Fragen initiiert werden soll.“ (FLICK 2009: 114)

Zum Einsatz kommt ein problemzentriertes Interview, da die Erfahrungen und Wahrnehmungen sowie eine Reflexion der Interviewten hinsichtlich einer bestimmten Problemstellung und eines gewissen Sachverhalts vorrangig sind. Erhoben werden diese Daten mit Hilfe eines Leitfadens von offenen, aber dennoch straffen, Fragen, wodurch keine geschlossenen Dimensionen verwendet werden, wie bei quantitativen Methoden. Innerhalb der qualitativen Forschung weist das problemzentrierte Interview einen weitgehend hohen Grad an Strukturierung auf, abgrenzend von beispielsweise anderen Befragungsarten, und ist deshalb den halbstandardisierten Interviews zuzurechnen. Bevor die offenen Fragen des Leitfadens das Gespräch strukturieren und lenken, werden die wichtigsten demographischen Daten mittels Kurzfragebogen erhoben, um biographische und soziale Hintergrunddaten zu erlangen (vgl. WITZEL/REITER 2012: 5 ff).

Der Leitfaden ist im Zuge dieser Methode kein direktes Methodeninstrument, sondern vielmehr eine Gedächtnisstütze oder ein Instrument zur Ausgestaltung des Gesprächs, wodurch eine breite Abdeckung des behandelten Themengebiets sowie das Aufkommen neuer Aspekte und Sichtweisen möglich ist, ohne jedoch gänzlich den fokussierten Themenbereich zu vernachlässigen (vgl. NOHL²2008: 21).

„Zum einen ‚hakt‘ [...] der Interviewer sozusagen im Gedächtnis die im Laufe des Interviews beantworteten Forschungsfragen ab, kontrolliert also durch die innere Vergegenwärtigung des Leitfadens die Breite und Tiefe seines Vorgehens. Zum anderen kann er sich aus den thematischen Feldern, etwa bei stockendem Gespräch [...], inhaltliche Anregungen holen [...]. Damit lassen sich auch Themenfelder in Ergänzung zu der Logik des Erzählstranges seitens des Interviewten abtasten, in der Hoffnung [...] fruchtbare Themen zu finden bzw. deren Relevanz aus der Sicht der Untersuchten festzustellen und durch Nachfragen zu überprüfen.“ (WITZEL 1982: 90)

Um einen flüssigen, fruchtbaren Dialog zu schaffen, ist es notwendig eine angenehme Gesprächssituation zu schaffen, sodass eine erfolgreiche, persönliche, mündliche Befragung zustande kommen kann, denn Interviews stellen ein formalisiertes Verfahren dar, wodurch eine Künstlichkeit der Befragungssituation entsteht, in der eine

asymmetrische soziale Interaktion vorherrscht (vgl. KROMREY ¹²2009: 338 f). Aus diesem Grund wurde ein grober Ablauf des Interviews konzipiert, der zu einer harmonischen Kommunikationsstimmung beitragen soll, der wie der Leitfaden auch im Annex (ab Seite 114) aufzufinden ist.

Die Fragen des Leitfadens, wodurch eine Beantwortung der forschungsleitenden Problemstellung anvisiert wird, sind in fünf Überbereiche gegliedert. Im ersten Teil sollen mittels erwähnten Kurzfragebogen die wichtigsten allgemeinen Daten zur Person erhoben werden, woran die weiteren forschungsinteressanten Bereiche anschließen und die Einstellung zur standardisierten Leistungsbeurteilung sowie bezüglich des Beurteilungsrasters erforscht werden, bevor konkret eine Konnexion zur bewerteten Performanz hergestellt wird. Die großteils offen formulierten Meinungsfragen zielen darauf ab, die persönlichen Haltungen der Interviewten einzufangen, wobei auch einige geschlossene Entscheidungsfragen ad hoc herangezogen wurden, um bestimmte Positionen klar definieren zu können.

6.2.3 Auswertungsmethode

Die vorliegenden verbalen Daten sollen basierend auf einer qualitativen Inhaltsanalyse untersucht werden. In einem ersten Schritt sind zunächst schriftliche Transkriptionen der mündlichen Interviews anzufertigen, welche den von DRESING/PEHL präsentierten Regeln eines einfach transkribierten Interviews folgen. Dabei wird kein besonderes Augenmerk auf para- und nonverbale Sprachsignale gelegt, da der Hauptfokus, hinsichtlich der für die Auswertung heranzuziehenden Inhaltsanalyse, auf den Inhalt des Redebeitrags zu lenken ist, weshalb diese Transkriptionsform ausgewählt wurde, bei der die Sprache aus Lesbarkeitsgründen ein wenig geglättet wird (vgl. DRESING/PEHL ⁵2013: 17 ff).

Die durch die qualitative Erhebungsmethode gewonnenen Texte weisen gewisse Unschärfen auf, entstanden durch das Offenheitsprinzip in den erfolgten Gesprächen. Dieses unscharfe, schwer interpretierbare Material gilt es zu berücksichtigen und im Sinne einer qualitativen Inhaltsanalyse systematisch auszuwerten, sodass sie intersubjektiv nachvollziehbar sind (vgl. GLÄSER/LAUDEL ⁴2010: 43 ff).

„Die **qualitative Datenanalyse** [...] wertet [...] verbales/textuelles [...] Datenmaterial im Hinblick auf das Forschungsproblem einer empirischen Studie interpretierend (hermeneutisch) aus. Sie folgt dabei dem explorativen (gegenstandserkundenden, hypothesen- und theoriebildenden) Erkenntnisinteresse qualitativer Forschung und ist stark **induktiv** (datengesteuert) ausgerichtet.“ (DÖRING/BORTZ ⁵2016: 599)

Bei dieser Auswertungsform werden in einem Aufarbeitungsprozess den Texten Informationen einem systematischen Verfahren folgend entnommen, wobei die Daten nach relevanten Informationen durchsucht werden. Diese sind in einem nächsten Schritt unterschiedlichen Kategorien zuzuordnen und innerhalb dieser weiterzuverarbeiten, wodurch sie umgewandelt, resümiert, mit weiteren Daten synthetisiert sowie in Verbindung gesetzt werden. Aus diesem Grund ist eine Abgrenzung zu anderen qualitativen Verfahren vorzunehmen, bei denen der Ursprungstext an sich die Analysebasis darstellt und keine Informationen extrahiert und unabhängig ihres Herkunftstextes untersucht werden (vgl. GLÄSER/LAUDEL ⁴2010: 46). Die Zuweisung zu Kategorien ergibt sich sowohl aus deduktiven als auch induktiven Merkmalen, denn eine Gewinnung dieser ist direkt aus dem Text heraus zu vollziehen sowie es auch zu einer vorab erfolgten Definierung von Kategorien kommt. Eine gewisse Strukturierung ist im Vorhinein bereits entstanden, welche jedoch durch die erhaltenen Daten aus einem *bottom-up*-Prozess, also aus einer Induktion des Textes, anzupassen sind. Aufgearbeitete Informationen sollen durch Anwendung von Interpretationsgrundformen veranschaulicht dargestellt und präsentiert werden. Darunter fallen die Zusammenfassung, die Explikation sowie die Strukturierung. Bei ersterem ist es notwendig, die Informationen auf das Wesentliche paraphrastisch zu reduzieren und die Kategorienbildung bei Bedarf iterativ zu bearbeiten und anzupassen. Die Explikation zielt auf die Klärung und Erhellung unverständlicher sowie diskrepanter Textdaten ab, die im Zuge einer Strukturierung geordnet zusammengetragen werden. Im Prozess einer Inhaltsanalyse sind je nach Forschungsgebiet und -zweck die nötigen Interpretationsformen auszuwählen, wobei sie sich nicht gegenseitig ausschließen und in Kombination angewandt werden können (vgl. MAYRING ¹¹2010: 48 ff), wie es im Falle dieser Empirie zur Anwendung kommt.

6.2.4 Lehrpersonen

Im Zuge der empirischen Forschung wurden drei Französischlehrende ausgewählt, um das zu behandelnde Thema aus ihrer Perspektive aufzuarbeiten und zu diskutieren. Die Anzahl scheint ausreichend unter Berücksichtigung einer qualitativen inhaltsbezogenen Studie. In Betracht der Auswahlkriterien wurde auf eine sehr wohl bewusste und absichtsvolle Auslese geachtet, da eine allgemeine Diversifizierung der Befragten angestrebt wurde, damit ein möglichst breites Spektrum an Eindrücken, Einstellungen und Haltungen eingefangen werden kann (vgl. DÖRING/BORTZ ⁵2016: 302). Eine bewusste Auswahl ist allerdings nicht auf bestimmte Personen zurückzuführen, sondern

lediglich auf ihre Hintergrunddaten, wie Schultyp, Unterrichtserfahrung und ähnliches, um grundsätzlich eine möglichst weite Bandbreite abzudecken und eventuell Unterschiede hinsichtlich einzelner Merkmale, wie der Einfluss der Unterrichtserfahrung zur Einstellung dem Beurteilungsraster gegenüber beispielsweise, ausmachen zu können. Eine Repräsentativität, sodass Rückschlüsse auf eine Grundgesamtheit an Lehrpersonen getätigt werden kann, ist nicht gegeben (vgl. BROSIUS/KOSCHEL/HAAS ⁴2008: 73), was allerdings auch nicht Ziel ist, sondern rein die Darbietung und Analyse der Einstellung einzelner Lehrkörper. Jene kooperationsbereiten und Einblick gewährenden Französischlehrende werden im Rahmen der Arbeit weitgehend anonymisiert, was nicht gänzlich möglich scheint ohne auf demographische Hintergrunddaten sowie Kontextinformationen, wie Schultyp beispielsweise, zu verzichten, wovon abgesehen wird (vgl. DÖRING/BORTZ ⁵2016: 284). Folgend ist eine kurze Charakterisierung der an der Forschung teilgenommenen Lehrpersonen aufzufindenden:

Lehrperson 1 (LP1)

Lehrperson 1 ist ein männlicher Lehrkörper, der Französisch sowie ein nicht-sprachliches Unterrichtsfach an einer BHS unterrichtet und unter fünf Jahren Berufserfahrung aufweist. Diese Lehrperson hat noch keine spezifischen Fortbildungen zur standardisierten Leistungsbeurteilung mit Hilfe des Beurteilungsrasters absolviert, hat allerdings eine kurze Einführung im Zuge der fachdidaktischen Ausbildung im Lehramtsstudium erhalten.

Lehrperson 2 (LP2)

Lehrperson 2 ist ein männlicher Lehrkörper, der Französisch sowie ein weiteres sprachliches Unterrichtsfach an einer BHS unterrichtet und circa fünf Jahre Berufserfahrung aufweist. Diese Lehrperson hat bereits mehrere spezifische Fortbildungen zur standardisierten Leistungsbeurteilung absolviert.

Lehrperson 3 (LP3)

Lehrperson 3 ist eine weibliche Lehrperson, die Französisch sowie ein nicht-sprachliches Unterrichtsfach an einer AHS unterrichtet und über 35 Jahre Berufserfahrung aufweist. Diese Lehrperson hat eine spezifische Fortbildung zur standardisierten Leistungsbeurteilung mit Hilfe des Beurteilungsrasters absolviert, empfand es allerdings als grundlegende Umstellung der bisherigen Bewertungspraxis.

7 Präsentation und Analyse der Ergebnisse

In diesem letzten Kapitel der vorliegenden Diplomarbeit werden die im Zuge der Forschung erhaltenen Daten ausgewertet und präsentiert. Zunächst wird in einem ersten Unterkapitel dargelegt, welche Ergebnisse bei der Beurteilung der zu bewertenden Schülerperformanz erhalten wurden, bevor in einem folgenden Kapitel auf die im problemzentrierten Interview erhobenen verbalen Daten eingegangen wird, die im Sinne einer qualitativen Inhaltsanalyse aufbereitet und diskutiert sowie mit den im Theorieteil dargelegten Sachverhalten in Beziehung gesetzt werden.

7.1 Beurteilungen der Performanz

Wie in der theoretischen Kontextualisierung bereits zu erkennen war, ist nicht nur das Zustandekommen einer SchülerInnenperformanz an vielfache Faktoren geknüpft, sondern in gleichem Maße auch die Bewertung einer solchen, sodass, wie die Forschung bereits aufzeigen konnte, Beurteilungen von Sprachproduktionen bei offenen Aufgabenformaten lehrpersonenabhängig, aufgrund subjektiver Bewertungsverfahren, variieren und sich unterscheiden. Eben jene der Subjektivität verschuldeten Divergenzen, wollen standardisierte Beurteilungsinstrumente versuchen, vorzubeugen.

In der durchgeführten Forschung, die qualitativ ausgelegt ist, weshalb die nachstehenden Ergebnisse keinesfalls als repräsentative Daten zur Hypothesenbildung anzusehen sind, sondern vielmehr als Basis für die folgende Inhaltsanalyse dienen, wurde ein und dieselbe Schülerperformanz von drei verschiedenen Französischlehrpersonen mit Hilfe des Beurteilungsrasters bewertet.

Im Folgenden werden die erlangten Ergebnisse hinsichtlich der Beurteilung aufgeschlüsselt dargeboten, verbalisiert und analysiert, bevor die Einstellungen der Französischlehrenden präsentiert werden, worauf der Hauptfokus liegt. Eben deshalb sollen hinsichtlich der Beurteilung lediglich die Divergenzen in Bezug auf die Punktevergabe und Beurteilung, hierbei insbesondere in den einzelnen der vier Kategorien sowie der Gesamtbewertung, beleuchtet und erläutert werden. Dies wird mit Hilfe der nachfolgenden Tabelle veranschaulicht.

Anbei werden die Ergebnisse der Beurteilung der Performanz von den unterschiedlichen Lehrpersonen präsentiert, aufgeteilt auf die einzelnen Kategorien, wobei auch die Divergenzen angezeigt werden:

Punkteverteilung der einzelnen Lehrpersonen

	LP1	LP2	LP3	Divergenz
Erfüllung der Aufgabestellung	9	7	8	20 %
Aufbau und Layout	10	9	5	50 %
Spektrum sprachlicher Mittel	7	6	5	20 %
Sprachrichtigkeit	6	5	4	20 %
Punkte insgesamt	32	27	22	10
Prozent der Gesamtleistung	80 %	67,5 %	55 %	25 %

Bei drei Französischlehrenden, welche dieselbe Performanz mittels Beurteilungsrasters bewerteten, kam es zu einem Spektrum der Gesamtpunkte von maximal 32 und mindestens 22 Punkten, was bedeutet, dass eine Divergenz von zehn Punktwerten auszumachen ist. Dieser Unterschied beträgt im Verhältnis zur höchstmöglich erreichbaren Anzahl von 40 Punkten 25 Prozent, weshalb die Divergenz ein Viertel der Gesamtleistung repräsentiert.

In den Kategorien „Erfüllung der Aufgabenstellung“, „Spektrum sprachlicher Mittel“ sowie „Sprachrichtigkeit“ betragen die Unterschiede zwei Punktwerte, wohingegen in der Kategorie „Aufbau und Layout“ ein Unterschied von fünf Punkten, was die Hälfte der in der Kategorie zu erreichenden Punkte darstellt, aufzufinden ist. Interessant erscheint dies deshalb, da dieser Bereich an bestimmten Charakteristika gemessen werden kann, die eine spezifische Textsorte aufzuweisen hat, wie sie im Textsortenpapier erläutert werden.

Auf diese Besonderheit soll im nachstehenden Kapitel, in dem Einstellungen und Haltungen der Lehrpersonen hinsichtlich des Beurteilungsraster präsentiert werden, näher eingegangen werden, wobei der Versuch unternommen wird, mögliche Gründe diesbezüglich zu erarbeiten.

7.2 Einstellungen der Lehrpersonen dem Beurteilungsraster gegenüber

In diesem Unterkapitel des empirischen Teils sollen die in den Interviews erhobenen Einstellungen der Französischlehrenden präsentiert, diskutiert und miteinander in Beziehung gesetzt werden. Diese Aufarbeitung wird, wie bereits erläutert, im Zuge einer qualitativen Inhaltsanalyse vollzogen, wobei die Inhalte zur besseren Veranschaulichung in Kategorien unterteilt und dort resümiert, einander gegenübergestellt und analysiert werden.

7.2.1 Beurteilung schriftlicher Performanzen

Die Beurteilung von SchülerInnen-texten stellt im Berufsleben einer Fremdsprachenlehrperson ein grundlegendes Aufgabenfeld dar, weshalb dem eine große Bedeutung zuzumessen ist. Lehrperson 1 äußert sich folgend dazu: „Also ich empfinde das Korrigieren von Schülertexten als eine ganz normale, angenehme Sache im Lehrberuf.“ (LP1, Interview 1)

Berücksichtigung in dieser Arbeit fanden Bewertungen mittels Beurteilungsraster, welcher im Zuge der Standardisierungsprozesse und in Hinsicht auf die standardisierte Reife- und Diplomprüfung aufgekommen ist. Dadurch hat sich in der Leistungsbeurteilung schriftlicher Performanzen einiges geändert. Ausgegangen von den Aufgabenformaten und Arbeitsaufträgen sowie -anweisungen bis hin zur Korrektur und Bewertung. Diese Erneuerungen schränken aus Sicht von Lehrperson 3 die Schülerinnen und Schüler im Prozess des Schreibens und in der Produkterstellung ein:

„Ich finde, früher durfte ein Schüler noch freier schreiben und da darf er sich dann auch entfalten und selbstständig Gedanken machen und jetzt bin ich in so ein striktes Format gepresst – ich finde das so arg! (.) Der Mensch wird genormt! Ich muss ja meine drei Punkte abarbeiten und darf nicht weniger, das verstehe ich mit zehn Prozent, aber MEHR? Ich schränke dadurch den Schüler sowas von ein. Und da bin ich total dagegen. [...] Alles schnell abarbeiten. So ein amerikanisches Schulsystem.“ (LP3, Interview 3)

Aus dieser Kritik geht hervor, dass die Formate als Restriktion aufgefasst werden und eine Normierung in einem negativen Sinne zu verorten ist, wodurch die Freiheit und Kreativität im Fremdspracheunterricht in Bezug auf die Schreibproduktion beschränkt werden. Diese Lehrperson bezieht sich dabei auf das amerikanische Schulsystem und

drückt dabei Ablehnung aus, da ihrer Ansicht nach die Aufgaben maschinell abgearbeitet werden und wenige eigene Gedankengänge seitens der SchülerInnen gefragt sind (vgl. Interview 3).

Das Assessment von SchülerInnenleistungen findet, wie in der theoretischen Kontextualisierung aufgezeigt wurde, auf vielen Ebenen und in unterschiedlichen Formen Einzug in den Unterricht und deckt dabei alle Kompetenzbereiche des Fremdsprachenunterrichts ab. In Bezug auf Schreibleistungen sehen die Lehrpersonen dies als Herausforderung an, vergleichsweise zu den rezeptiven Fertigkeiten, da die Beurteilung von Sprachproduktion verhältnismäßig schwieriger objektivierbar ist und vielfältige Ansprüche zu erfüllen sind, sodass die Korrektur und Bewertung einer Performanz dieser gerecht wird: „Ja, es ist natürlich eine Herausforderung“ (LP2, Interview 2).

7.2.2 Beurteilungsraster

Der Beurteilungsraster, dem sich in den Gesprächen intensiv zugewandt wurde, stellt eine vereinheitlichte Grundlage bei der Beurteilung schriftlicher SchülerInnenperformanzen dar, die im Zuge der standardisierten Reife- und Diplomprüfung im österreichischen Schulsystem zur Anwendung kommen muss aber davor auch schon herangezogen werden soll. Er ist somit als Beurteilungsinstrument für alle Französischlehrpersonen auf Sprachniveau B1 anzusehen, weshalb es von besonders großem Interesse ist, zu erfahren, wie die Lehrpersonen im Fremdsprachenunterricht diesen aufnehmen und bewerten sowie in ihrem Unterricht anwenden.

Prinzipiell sind alle der befragten Lehrenden äußerst positiv auf Beurteilungshilfen im Allgemeinen zu sprechen und begrüßen eine einheitliche Basis und Richtlinie sowie Anleitung: „Ich finde es gut, dass es eine Basis gibt, auf die man sich beziehen kann“ (LP1, Interview 1).

Weiters wird es als äußerst positiv aufgenommen, dass ein solch allgemeiner Kriterienkatalog vorliegt, der schulübergreifend Transparenz und höhere Fairness gewährleistet, wie Lehrperson 2 äußert: „Ich finde es jetzt schon grundsätzlich fairer“ (LP2, Interview 2). Dies schätzen alle der befragten Lehrpersonen (vgl. Interview 1-3).

Kritik wird allerdings geübt hinsichtlich der Form und konkreten Umsetzung bei diesem Beurteilungsinstrument: „Eine Anleitung prinzipiell finde ich schon sehr gut und wünschenswert aber in dieser Form nicht wirklich“ (LP3, Interview 3).

„Jedes Mal, wenn ich es verwende (..) finde ich es schon sehr aufwendig, was die genauen Unterpunkte angeht und einem abverlangt, weil ich muss mir das schon jedes Mal noch durchlesen, weil das alles nicht so genau im Kopf bleibt [...] zu umfassend, um es alles im Kopf zu haben. [...] Also man muss das schon immer wieder durchlesen und ich finde, das ist dann doch schon (..) sehr (.) komplex und umfangreich.“ (LP2, Interview 2)

Dabei wird einerseits beanstandet, dass die Anleitung zu umfangreich und dadurch wenig praktisch und effektiv zu gebrauchen ist, sowie andererseits, dass dadurch die Anwendung eines sehr hohen zeitlichen Aufwandes bedarf:

„Mit den neuen Beurteilungsrastern und Korrekturblättern ist es jetzt aufwendiger geworden [...]. Es nimmt jetzt viel mehr Zeit in Anspruch. [...] Der Schüler hat 50 Minuten an dem Text gearbeitet, ich verstehe nicht, warum ich zum Bewerten länger brauchen sollte als der Schüler zur Produktion [...]. Denn alle diese einzelnen Punkte und alle diese Bänder durchzugehen, das geht nicht, das dauert einfach zu lange. [...] Das ärgert ein wenig. Der Aufwand ist einfach zu groß (..) Der Stellenwert ist zu hoch, der Stellenwert wie man mit Beurteilungen umgeht [...]. Lehrer sind hauptsächlich da, um zu Unterrichten eigentlich und weniger zum Raten. (..) Also grundsätzlich begrüße ich so eine Basis aber ich würde es vereinfachen.“ (LP2, Interview 2)

7.2.2.1 Kategorien

Die Form des Rasters wurde, wie bereits aufgezeigt, kritisiert, wobei die Aufteilung in vier Kategorien durchgehend positiv wahrgenommen und als nachvollziehbar sowie logisch empfunden wurde. Die Berücksichtigung der Kategorien „Erfüllung der Aufgabenstellung“, „Aufbau und Layout“, „Spektrum sprachlicher Mittel“ sowie „Sprachrichtigkeit“ scheint allen befragten Lehrpersonen zuzusagen, wobei eine klare Trennung dieser untereinander und auch eine Gleichwertigkeit aller vier genannten nicht gänzlich nachvollzogen werden kann und in Hinsicht darauf kritisch Stellung bezogen wird (vgl. Interview 1-3).

In Bezug auf die Bedeutung und den Stellenwert der einzelnen Kategorien empfinden die Interviewten jene, welche die Sprachkompetenzen abbilden und umfassen, als grundlegend wichtig im Fremdsprachenunterricht, wobei die Aufgabenerfüllung nicht minder bedeutend eingeschätzt wird, da sie als Voraussetzung für eine zufriedenstellende Schreibleistung anzusehen ist. „Aufbau und Layout“ wird insgesamt von allen Personen als die am wenigsten wichtige Kategorie angesehen und genannt (vgl. Interview 1-3).

„Die Kategorien „Spektrum sprachlicher Mittel“ und „Sprachrichtigkeit“ würde ich nun mal als die wichtigsten bezeichnen [...], weil ja das den Kern einer Sprachkompetenz am genauesten trifft, wobei die anderen auch wichtig sind. Also sie sind nicht wegzudenken aber sie sind nun mal nicht so wichtig wie die beiden anderen, die ich gerade genannt habe. [...] Wenn die „Erfüllung der Aufgabenstellung“ nicht passt, dann kann das sprachlich der beste Text der Welt sein [...], deswegen kann diese Kategorie als Voraussetzung angesehen werden.“
(LP1, Interview 1)

Außerdem ließ sich hinsichtlich der Kategorien ausmachen, dass die Lehrpersonen diese nicht gänzlich unabhängig voneinander wahrnehmen und heranziehen. Dabei werden vor allem die beiden Kategorien „Spektrum sprachlicher Mittel“ und „Sprachrichtigkeit“ als zusammengehörig angesehen:

„Zum Beispiel zwischen „Spektrum sprachlicher Mittel“ und „Sprachrichtigkeit“, da neige ich persönlich dazu, dass ich das vermische teilweise, beziehungsweise auch die „Aufgabenerfüllung“ und das „Aufbau und Layout“ (..), insbesondere dass das Layout jetzt einen so hohen Stellenwert bekommt, [...] denke ich manchmal, dass das nicht ganz in Relation steht.“ (LP2, Interview 2)

Lehrperson 1 versteht die vier Kategorien aufgeteilt in zwei Teile, die jeweils zwei Punkte in sich verbinden:

„Ja, ich finde, also das sind so zwei Pärchen [Anmerkung: die Kategorien]. Das eine „Erfüllung der Aufgabenstellung“ und „Aufbau und Layout“, das ist quasi das Drumherum, das passen muss und der harte Kern sind eben „Spektrum sprachlicher Mittel“ und „Sprachrichtigkeit“, der dann im Zweifelsfall, naja, wichtiger ist, sage ich.“ (LP1, Interview 1)

Weiters wird in Bezug auf die vier Kategorien des Beurteilungsrasters davon ausgegangen, dass im Bereich der ersten Gruppierung, der „Erfüllung der Aufgabenstellung“ und des „Aufbau und Layout“ SchülerInnen schneller Punktwerte

sammeln können, da hier keine Sprachkompetenzen gefragt sind, sondern Textsortenwissen antrainiert werden kann, wodurch bereits eine schlechtere Beurteilung vorgebeugt werden kann (vgl. Interview 1-3):

„Das System hat hier einen Bonus für die Lernenden, da alle Kategorien gleichwertig sind und die beiden Kategorien der „Aufgabenstellung“ und „Aufbau und Layout“ sind, wenn ich das so formulieren darf, die geschenkten Punkte, denn da kann man nicht viel falsch machen und die sind relativ leicht zu holen, diese Punkte und damit kann ich natürlich dann im Kern, nämlich bei den „sprachlichen Mitteln“ und bei der „Sprachrichtigkeit“ wieder einiges kompensieren und dann schaut die Gesamtnote auch wieder ein bisschen besser aus.“ (LP1, Interview 1)

Wie obenstehend dargelegt wurde, werden die vier Kategorien des Rasters sehr positiv von den Lehrenden angenommen, wobei die Gleichwertigkeit dieser bei allen Befragten nicht gültig und nachvollziehbar scheint. Hierbei wird vor allem „Aufbau und Layout“ des Textes als zu bedeutend eingestuft (vgl. Interview 1-3). Möglicherweise ist dies auch als Teilerklärung für die hohe Divergenz in jener Kategorie zu verstehen, da Lehrpersonen eher gewillt sind, hier weniger Punkte zu vergeben, da sie den Stellenwert als zu hoch einschätzen, worauf in Unterkapitel 7.2.4 dieser Diplomarbeit weiter eingegangen wird.

In einem nächsten Schritt will auf die Auffächerung der vier Kategorien in elf Niveaustufen eingegangen werden, wobei einzelne Deskriptoren bei sechs dieser Niveaus ausformuliert wurden und als sachliche Bezugsnorm heranzuziehen sind.

7.2.2.2 Deskriptoren

Aus den Interviews ging hervor, dass einerseits die Deskriptoren, als zu umfassend empfunden werden, sodass eine gewisse Unübersichtlichkeit entsteht, und andererseits die Graduierungen der einzelnen Stufen in Hinsicht auf ihre Formulierungen als wenig präzise und aussagekräftig aufgefasst werden (vgl. Interview 1-3). Auf diesen Sachverhalt wurde in Kapitel 4.4 dieser Arbeit im Zuge der deskriptiven Untersuchung des Beurteilungsrasters hingewiesen, wobei dies auf die qualitativen Graduierungen der verschiedenen Ausprägungsstufen, die stets einer Relativierung bedürfen, zurückzuführen ist. Dies ist ein Charakteristikum, welches bei offenen Aufgabenformaten und subjektiven Assessmentmethoden mitschwingt (vgl. BROWN/ABEYWICKRAMA ²2010: 351 ff).

Diese unpräzise empfundenen Formulierungen werden, wie in den Interviews herauszufinden war, in Bezug auf den hier behandelten Raster von den Lehrpersonen bedauernd wahrgenommen, wobei Lehrperson 3 dies folgendermaßen veranschaulicht ausdrückt: „Schade sind halt diese Wischivaschi-Sätze.“ (LP3, Interview 3)

Teilweise waren in den Gesprächen bei Ansprache der Formulierungen einzelner Deskriptoren selbige Wortlaute bei den verschiedenen Befragten zu finden, wobei hier umgangssprachliche Äußerungen in abwertender Form für die Beschreibung eines unpräzisen Charakters gewählt wurden wie „schwammig“ und „Wischivaschi“ (vgl. <http://www.duden.de/rechtschreibung/Wischivaschi>, <http://www.duden.de/rechtschreibung/schwammig#Bedeutung3>).

„Das ist auch so eine Wischivaschi-Formulierung, wo ich mir auch schwer getan habe (.), ob das jetzt mein Empfinden ist oder wie das aufzufassen ist.“ (LP2, Interview 2)

Lehrperson 2 gibt auch Einblick, dass diese Empfindung keinesfalls nur von ihm geteilt wird, sondern dass auch von Kolleginnen und Kollegen in Gesprächen zu dieser Angelegenheit selbige Meinung geteilt wird:

„Wenn ich das richtig verstanden habe, empfinden Sie die Formulierungen doch als schwammig und eher ungenau und wenig präzise. - Ja genau. Und das hört man auch von Kollegen, wenn man mit denen spricht. Der Eindruck wird vermittelt als wäre das schön abgegrenzt aber das ist es nicht (.) weil das immer nur Ausprägungen sind und keine klaren Unterschiede zu erkennen sind.“ (I, LP2, Interview 2)

Zusammenhängend mit dieser Problematik von unpräzise wahrgenommen Formulierungen unterschiedlicher Deskriptoren ist der Begriff der Interpretationssubjektivität zu betrachten, wie in der theoretischen Kontextualisierung in Unterkapitel 4.3.3 bereits aufgezeigt wurde.

„Ja, das ist ein bisschen schwammig. Das ist ein wenig abhängig von der Interpretationsweise eines jeden Lehrers, der das liest.“ (LP1, Interview 1)

7.2.2.3 Interpretationssubjektivität

Dass Deskriptoren nicht durchwegs gleich verstanden werden ist wenig überraschend (vgl. ALDERSON/CLAPHAM/WALL 1995: 107 ff) und bedingt eine ungleiche und andersartige Verwendung des Beurteilungsinstruments (vgl. GASSNER/MEWALD/SIGOTT 2008: 15).

„Ich finde es gut, dass es eine Basis gibt, auf die man sich beziehen kann und dass diese Basis für eigene, sag ich mal, subjektivere Empfindungen Spielraum lässt, aber (..) ich finde es gut, dass es immer eine [...] Basis gibt.“ (LP1, Interview 1)

Wie Lehrperson 1 aufzeigt, ist nicht nur die einheitliche Basis als positiv anzusehen, sondern ebenfalls der Spielraum für subjektive Empfindungen, da dadurch den Beurteilenden ein gewisser Freiraum bleibt, um ihr Ermessen als Evaluatoren und Evaluatorinnen auch persönlich einfließen zu lassen, gegenteilig zu objektiven Bewertungsmethoden.

„Das ist ja auch die Illusion, die alle haben, dass es immer gleich ist aber natürlich ist das abhängig und nicht immer vollkommen objektiv. Was soll man machen? Man muss das Beste herausholen und sich dabei natürlich seinem Umfeld anpassen.“ (LP3, Interview 3)

In dieser Äußerung wird das Umfeld, in dem ein Lehrkörper als Beurteilender agiert miteinbezogen, wobei Lehrperson 3 darauf verweist, dass beispielsweise die Klasse, die KollegInnen beziehungsweise die Schule sehr wohl auf die Auffassung von SchülerInnenleistungen einwirken, wodurch sich schulartenspezifische Ergebnisse in Hinsicht auf Beurteilungsmaßstäbe ergeben (vgl. MAIER 2015: 96 f).

„Das suggeriert eh der Raster, meiner Ansicht nach. Er lässt so viel Spielraum offen, dass man das natürlich in unterschiedlichen Bereichen einsetzen kann und anpasst.“ (LP2, Interview 2)

Da eine objektive Beurteilung bei schriftlichen Performanzen nicht garantiert werden kann, wird verstärkt eine kriterienabhängige und weniger beurteilendenabhängige Bewertung anvisiert, wodurch eine gewisse Verlässlichkeit innerhalb der beurteilenden LehrerInnen gegeben sein sollte (vgl. ALDERSON/CLAPHAM/WALL 1995: 128 f).

7.2.2.4 Intrapersonelle Bezugsnorm

Um einen willkürlichen Umgang mit dem Beurteilungsraster vorzubeugen, ist es notwendig, dass Lehrende für sich selbst eine beständige und durchwegs gleich aufzufassende Norm setzen, die sie mit den Formulierungen der Deskriptoren in Verbindung setzen, sodass sie eine intrapersonelle konstante und fortwährende Bezugsnorm definieren (vgl. ALDERSON/CLAPHAM/WALL 1995: 107 ff).

In den Interviews konnte herausgefunden werden, dass alle Befragten davon ausgehen, für sich selbst eine beständige Auffassung der Deskriptoren festgesetzt zu haben, sodass sie nicht situationsabhängig bei unterschiedlichen Performanzen die Niveaustufen andersartig interpretieren und auffassen. Sie sind demnach davon überzeugt, eine intrapersonelle Bezugsnorm aufzuweisen und durchgehend auf eine gleiche Art und Weise SchülerInnen-Texte bewerten, wobei das Entstehen einer solch intrapersonellen Norm laut Lehrperson 3: „Wahrscheinlich passiert [...], ohne dass man sich dessen bewusst ist. Ja also eine gewisse Richtlinie schon.“ (LP 3, Interview 3).

Äußerst interessant hierbei ist, dass alle Lehrpersonen auf die Punkte der Bände der aufgefächerten Niveaustufen verwiesen haben, weshalb festzuhalten ist, dass die an der Empirie teilgenommenen Lehrpersonen sich teilweise eher an der Höhe der Punktwerte orientieren und ihre Bewertung nach eigenem Empfinden daran festmachen als an den sprachlichen Formulierungen der Deskriptoren in den unterschiedlichen Bändern an sich.

„Ja, also ich, ich richte mich da auch sehr stark nach der Punktzahl, je nachdem in welchem Band ich bin und versuche die Leistung darin einzuordnen nach meinem Maßstab, wie es zum Beispiel bei sechs Punkten für einen solchen Rahmen (.) ahm, ja, angemessen ist, eben noch eine positive Leistung. So versuche ich das zu interpretieren. Ich versuche das auch eben so zu interpretieren, wie die Autoren das gemacht haben und die schreiben da, ok, ja, in der Zeile für sechs Punkte das und das, dann werden die das so gemeint haben, dass in diesem Rahmen eben eine noch positive Leistung verstanden wird.“ (LP3, Interview 3)

Demnach ist zu konstatieren, dass im Prozess der Beurteilung die SchülerInnenleistung beim Rezipieren bereits voreingestuft und mit der zehnstufigen Punkteskala in Bezug gesetzt wird. Diese Punktwerte sind wiederum mit der Notenskala im österreichischen Schulsystem verknüpft, wobei eine Abstufung der Punktwerte gemäß den fünf Noten vorgenommen wird.

„Ja, also sobald es mir schwammig vorkommt, neige ich dazu, dass ich mich auf die Punkte beziehe. (LP2) - Ok (.) und die Punkte setzen Sie dann mit dem Notensystem in Verbindung und gleichen es daran an? (I) - Ja genau, schon. Also ab 6 ist es positiv und dann arbeite ich mich nach oben. In dem Schulsystem ist das schon fest verankert und da kann man dann auch nicht anders. Das ist also nicht aus dem Kopf zu kriegen (LP2)“ (Interview 2)

Dadurch wird verdeutlicht, dass die ausformulierten Bezugsbezeichnungen auf wenig Anklang bei Lehrpersonen stoßen, da sie in ihrer Anzahl zu umfangreich sowie zu unpräzise und wenig aussagekräftig in ihrer Formulierung sind, wodurch bei der Bewertung eine Orientierung an Punktwerten bevorzugt wird. Prinzipiell wird unbewusst mit Hilfe der kognitiven Kompetenz zur Informationsreduktion Wahrgenommenes durch vordiagnostische Klassifikationen zugeordnet (vgl. KLEBER 1992: 78 f).

Dabei ist der Bezug zum Notensystem des österreichischen Schulwesens einleuchtend, denn letztendlich sind Leistungen der Schülerinnen und Schüler am Ende eines Unterrichtsjahres in dieses reduzierte System einzuordnen, weshalb sich diese Klassifizierung bereits bei der Beurteilung einzelner Texte bemerkbar macht und eine möglicherweise unbewusste Orientierung der Lehrpersonen daran stattfindet.

„Ich schau eh immer: „Was würdest du denn da jetzt geben?“ (.), ja, dann gebe ich die Punkte [Anmerkung: die Punkte der Deskriptoren des Rasters im Korrekturblatt] ein und ich komme meistens auf das, was ich eh von vornherein denke. Ja, also das ist so.“ (LP3, Interview 3)

Somit kann davon ausgegangen werden, dass Beurteilende häufig einen ersten Eindruck gewinnen und mit steigender Erfahrung dazu tendieren, eine Performanz einer Note zuzuordnen, wobei im Bewertungsprozess diese Vordiagnose unzureichend ausgeblendet werden kann, weshalb sich dieser Eindruck in der Bewertung, selbst mittels Beurteilungsraster, niederschlägt, was auf die subjektive Auffassung qualitativer Ausprägungsgrade zurückgeführt werden kann. Die Ausprägungsgrade eröffnen zwar einen gewissen Freiraum, wie aufgezeigt wurde, drücken dennoch bestimmte Punkte aus, die erfüllt sein müssen, um den ein oder anderen Punktwert zu erreichen, weshalb er als transparente Argumentationsstütze für die Bewertung angesehen werden kann:

„Der Raster ist eine Argumentationshilfe, das kann manchmal sehr gut sein. Eine transparente Anleitung, auf die kann man sich auch berufen und dafür bin ich auch dankbar.“ (LP3, Interview 3)

7.2.2.5 Anwendung des Rasters

Geschildert wurde bereits, dass Beurteilungsinstrumente einen unterschiedlichen Gebrauch bei Lehrenden finden, wobei im Rahmen der empirischen Forschung auch erhoben werden wollte, wie der Beurteilungsraster Anwendung im Fremdsprachenunterricht findet. Zwingend eingesetzt werden muss er bei der standardisierten Reife- und Diplomprüfung, sollte allerdings in vorherigen Bewertungssituationen Gebrauch finden.

In offiziellen Situationen wie der Reifeprüfung und bei Schularbeiten ziehen die Befragten ausnahmslos den Beurteilungsraster für die Bewertung heran. Bei weiteren SchülerInnen-texten schieden sich die Meinungen. Lehrperson 1, als Junglehrer mit wenig Erfahrung, verwendet den Raster bei allen Performanzen mit der Begründung, dass diese bereits Vorbereitungen auf Überprüfungen in einem offiziellen Rahmen sind und die Bedingungen somit dieselben sein sollten:

„Also ich würde bei bedeutenden Texten den Raster natürlich wirklich ausnahmslos nehmen und bei Hausübungen würde ich das auch. Und wenn das mal so Texte sind, die ich zwischen Tür und Angel kriege, ja, also so By-the-way-Texte, dann korrigiere ich es nur nach diesen vier Kategorien, wo ich sage, da schaue ich schnell drüber. Ja, da wären es dann nur mehr die Kategorien aber sonst bei Hausübungen, die ja auch Vorbereitungen sein sollen, für eine Schularbeit zum Beispiel, da würde ich es dann dementsprechend dann auch gleich ordentlich mit dem Raster korrigieren.“ (LP1, Interview 1)

Demgegenüber empfinden Lehrperson 2 und 3 die Anwendung als zu zeitintensiv, weshalb es ihnen nicht sinnvoll scheint, den Raster bei sämtlichen Bewertungen heranzuziehen, sondern ebenfalls nur bei Texten, die explizit als Vorbereitung auf eine offizielle Überprüfung anzusehen sind, wie Lehrperson 3 folgend erläutert:

„Nein, also nicht immer. Um Gottes Willen, also da würde ich verrückt werden, wenn ich ihn bei allen Hausübungen verwenden würde. (lacht) Also da verlasse ich mich schon auf mein Bauchgefühl oder meinen Kopf beim Verbessern und auf meine Erfahrung. Bei der Schularbeit und nur bei der Schularbeit (..) außer ich weiß, jemand will in Französisch maturieren, dann schon, denn da muss das trainiert werden, weil die Matura wird natürlich auch so beurteilt. Das finde ich so auch vernünftig. Warum sollte ich die anderen auch damit quälen? Und das ist meine Zeit, die da viel intensiver beansprucht wird. Ich meine, die jungen Leute, die arbeiten so aber ich nicht. (LP3, Interview 3)

Aus den hier gegenübergestellten Zitaten können zwei Positionen ausgemacht werden: JunglehrerInnen, die eher dazu neigen, den Raster häufiger anzuwenden, und erfahrenere Lehrpersonen, die sich auf ihre Erfahrung verlassen und den Beurteilungsraster als zeitintensive Erneuerung betrachten. Dies könnte auf die Ausbildung zurückzuführen sein, wobei JunglehrerInnen die Beurteilungspraxis von Beginn an mit Raster kennengelernt haben, während Personen mit viel Expertise auf altbewährte Methoden und ihre Erfahrung zurückgreifen (vgl. Interview 1-3).

Dies kann auch an der Korrekturpraxis erkannt werden, wenn das Vier-Farben-Korrigieren betrachtet wird, dem jüngere Lehrperson, wie in den Interviews erhoben wurde, positiv gegenüberstehen:

„Aber ich finde das Korrigieren und das Analysieren von Schülertexten interessant und vor allem auch mit diesen vier Farben, das finde ich gut, dem stehe ich positiv gegenüber. Aber der Aufwand, das ist einfach ein Wahnsinn.“ (LP2, Interview 2)

Lehrperson 3, welche bereits viele Unterrichtsjahre tätig ist, erachtet dieses System hingegen als wenig praktisch:

„Ja also (..), wissen Sie, ich bin nicht die Wahnsinnige, die mit vier Farben korrigiert, also ich lasse mich da wirklich nicht ‚papierln‘.“ (LP3, Interview 3)

7.2.2.6 Schulungen

Fortbildungen und Schulungen stellen, wie in der Theorie in Kapitel 4.3.3 erarbeitet, einen wichtigen Faktor dar, der zu einem professionellen Umgang mit Beurteilungsinstrumenten beiträgt. Diesbezüglich wurden auch unterschiedliche Meinungen der Lehrenden geäußert. Einerseits ist der Standpunkt angesprochen worden, dass der Raster einerseits selbsterklärend sei und keine weiteren Fortbildungen nötig seien und andererseits, dass Fortbildungen bezüglich der Bewertungskultur förderlich sein können (vgl. Interview 1-3).

„Wir hatten – zum Glück – eine Schulung hier bei uns im Haus. Ich hatte teilweise schon sehr ungute Erfahrungen mit Seminaren aber ja das eine hatten wir und da wurde einem schon eingetrichtert wie OBJEKTIV das nicht ist aber ja, das ist natürlich nicht ganz so, wie behauptet wird. Aber eine Fortbildung hilft sicher. Wichtig ist auch, dass wir eine gute Französisch-Fachgruppe sind, die sich sehr unterstützt, und das ist besonders wichtig, finde ich.“ (LP3, Interview 3)

In oben angeführten Zitat wurde auf die Wichtigkeit der schulinternen Fachgruppe verwiesen, welche neben Schulungen für eine klassenübergreifende Beurteilungspraxis von Bedeutung ist. Weiters wurde die Objektivität bei Beurteilungsrastern angesprochen, wobei dies in Zusammenhang mit den Bezugsnormen im folgenden Unterkapitel dieses inhaltsanalytischen Teiles der Arbeit näher beleuchtet werden soll.

7.2.3 Bezugsnormen

Jegliche Bewertung einer Leistung ist normbezogen und bedarf aus diesem Grunde einer Relativierung zu einem Bezugssystem. Im Rahmen der hier betrachteten Leistungsbeurteilung von schriftlichen Texten mittels Beurteilungsraster setzt eben dieser Raster die Norm, auf die sich zu beziehen ist. Hierbei handelt es sich um einen sachlichen Maßstab, wobei die Sachverhalte, die es zu erfüllen gilt und anhand welcher die Ausprägung der Leistung zu messen ist, als kriteriale Norm angesehen anzusehen ist.

Lehrperson 1 bezieht sich bei Bewertungen an den Maßstab der Minimalstandards, die laut Bildungsstandards auf Sprachniveau B1 zu erreichen sind und sich in den Aufgabenstellungen und Deskriptoren niederschlagen:

„Da halte ich es generell so, dass jeder ein gewisses Mindestmaß anstreben und erreichen muss und alles was mehr ist, ist umso besser, aber kein Bezugsmaßstab.“

(LP1, Interview 1)

Neben den sachlichen Kriterien sind bei Person, die zunächst im Unterricht lehren und in weiterer Folge SchülerInnenleistungen beurteilen, weitere Bezugssysteme auszumachen:

„Ja, die Bezugsnorm ist, wie es bei jedem Lehrer sein wird, dass man die persönliche Komponente und den Bezug zum Schüler leider nicht aus dem Kopf kriegt und ausblenden kann. (.) Wenn man das also korrigiert, denkt man trotzdem an den einen Schüler, ich sage es ganz ehrlich. Und da gibt es die Schüler und Schülerinnen, die die Schlimmen sind und (.) neigen dazu, sich wenig vorzubereiten auf diverse Sachen. Und das kriegt man dann einfach nicht aus dem Kopf beim Korrigieren und da hilft mir der Raster dann schon. Ja.“ (LP2, Interview 2)

Abgesehen von der schon diskutierten Interpretationssubjektivität gilt es auch die individuelle und soziale Bezugsnorm zu berücksichtigen, welchen im schulischen Unterrichtsgeschehen ein hoher Stellenwert beizumessen ist. Während der Befragung der Französischlehrenden sind die soziale und individuelle Bezugsnorm thematisiert worden, wobei es festzuhalten gilt, dass Lehrperson 2 beispielsweise den Raster als Hilfe ansieht,

nicht zu sehr innerhalb eines sozialen Bezugssystems zu bewerten, verdeutlicht dennoch, dass der Bezug zu den SchülerInnen nicht auszublenden ist (vgl. Interview 2).

Die soziale Bezugsnorm als Vergleichsperspektive der Schülerinnen und Schüler im Klassenverband, welche im österreichischen Schulsystem laut Forschungen am häufigsten Anwendung findet (vgl. SCHMIDINGER 2012: 14), wird von den Lehrpersonen in den Interviews als Realität bezeichnet, wobei keine klare positive oder negative Position dazu bezogen wird. In Hinblick auf die individuelle Bezugsnorm ist hierbei allerdings eine positive Stellungnahme zu bemerken, denn Fortschritte sind laut der Befragten auf jeden Fall zu würdigen (vgl. Interview 1-3) und hierzu ist der Spielraum, den die Kriterien des Rasters bieten, zu nutzen:

„Ja natürlich. Das [Anmerkung: individuelle Lernfortschritte] MUSS ich doch auch berücksichtigen. Wenn der eine sich so plagt und dann hängt er einfach nur wegen dem und ich sehe, der macht und macht und macht (..) ja, was mache ich denn dann? Ich jongliere mit den Punkten, damit ich beweisen kann, dass es eine Vier ist [...]. Ja, also das sage ich jetzt so, weil es eh anonym ist. Wenn ich das auf einem Seminar sage, dann heißt es gleich: „Nein, nein, nein! Also, das geht so nicht!“ (..) Aber das wird man doch nicht so machen. (.) Seien wir ehrlich, ich bin nicht die Einzige [...]. Wir sind ja keine Maschinen.“ (LP3, Interview 3)

Wie dem hier präsentierten Zitat zu entnehmen ist, will versucht werden, den Raster als objektives Beurteilungsinstrument zu bezeichnen, wobei die beurteilenden Lehrerinnen und Lehrer es für äußerst wichtig halten die individuelle Bezugsnorm miteinfließen zu lassen und es gar für unmöglich halten, die soziale Bezugsnorm nicht zu berücksichtigen, auch wenn dies unbewusst stattfindet. Bei offenen Aufgabenformaten und dadurch subjektiven Bewertungsmethoden ist somit keinesfalls von Objektivität zu sprechen, wie in den Gesprächen beanstandet wurde (vgl. Interview 1-3).

„Ich kann WIRKLICH NICHT sagen, dass die Beurteilung mit dem Raster objektiv ist. Man will ja nicht betrügen, aber wenn man sieht das ein Schüler wirklich schwach ist und der gibt einfach wirklich alles und nur sein Bestes [...] ja, dann will ich den doch nicht entmutigen und auf der Strecke lassen und wenn der [...] einfach nur versucht einen Vierer zu kriegen, wo ich sage, dass kann ich auch verantworten, dann gebe ich ihm den, [...] wo bleibt denn da auch die Menschlichkeit den Schülern gegenüber.“ (LP3, Interview 3)

Somit kann in Hinsicht auf die Maßstäbe, welche bei der Bewertung schriftlicher SchülerInnenperformanzen im Fremdsprachenunterricht bei den befragten Lehrenden herangezogen werden, alle drei der klassischen Bezugsnormen, wie sachliche, soziale und individuelle Maßstäbe, aufzufinden und vertreten sind (vgl. Interview 1-3).

7.2.4 Performanz

Wie in Unterkapitel 7.1 bei der Analyse der Beurteilungen der Performanz aufgezeigt wurde, ergab sich eine überraschend hohe Divergenz im Bereich des „Aufbau und Layout“, wobei nach den bislang präsentierten Daten, die im Zuge der Interviews erhoben wurden, festgehalten werden kann, dass eben jene Kategorie als die unbedeutendste bei den Lehrpersonen aufgefasst wird und dass dazu geneigt wird, mit den Punktwerten zu „jonglieren“, sodass meist jene Note erreicht wird, von der in einer ersten Einschätzung ausgegangen wird. Aus diesen Gründen könnte vermutet werden, dass die Lehrpersonen, welche die Bewertungen mit der höchsten Divergenz aufweisen, in unterschiedlichen Referenzrahmen agieren und somit mit ungleichen Maßstäben und Erwartungen an ein und dieselbe Schülerleistung herangehen, wodurch selbst bei Anwendung des Rasters eben jener Freiraum genutzt wird, um die Note, die im eigenen Empfinden der Evaluatorin beziehungsweise des Evaluators als adäquat angesehen wird, zu erreichen, wobei Änderungen hierfür möglicherweise in der am wenigsten bedeutenden angesehenen Kategorie zu verorten sind (vgl. Unterkapitel 7.1).

Auf die Frage wie die Punktwertzuordnung zustande kam, wurde prinzipiell die Formulierung der Deskriptoren vorgelesen und in einem nächsten Schritt auf die Punkteanzahl verwiesen und nach dem eigenen Ermessen begründet, inwiefern dieser Punkt in der Performanz gelungen oder misslungen sei (vgl. Interview 1-3).

7.2.5 Handlungsorientierung

Die Handlungsorientierung des Schülers, gemessen an dieser speziellen Performanz, sollte in den Interviews ebenfalls thematisiert werden, wobei nach Überlegungen, alle Lehrenden zustimmten, dass der Schüler auf Sprachniveau B1 sprachlich handlungsfähig ist und kommunizieren kann (vgl. Interview 1-3).

„Ahm ... ja, da würde ich es als das Bedeutendste im Französischunterricht einschätzen, dass die Schüler später im Berufsleben kommunizieren können. Das ist das Wichtigste. Aber in anderen Bereichen, wie in diplomatischen Diensten, ist das

anders zu beurteilen, da dann die Richtigkeit genauso von Bedeutung ist. Aber nicht in der Schule, ja, da die Handlungsorientierung. Bei den Schülern ist es wirklich das wichtigste, dass die Kommunikation funktioniert.“ (LP1, Interview 1)

Die erbrachte Leistung, auf Sprachniveau B1 handlungsfähig, wurde dennoch von zwei Beurteilenden als nur ausreichend eingestuft und eingeschätzt, da im Fremdsprachenunterricht die Sprachrichtigkeit nach wie vor einen hohen Stellenwert genießt, wie Lehrperson 2 folgendermaßen begründet:

„Ähm (.) Also die Sprachrichtigkeit ist natürlich schon zu bewerten aber unterm Strich, gerade bei uns an der BHS, ist es so, dass der Schüler die Sprache auch im beruflichen Kontext verwenden sollte, ja. Und in der beruflichen Praxis geht das nicht. Das ist inakzeptabel mit vielen Fehlern, das reicht nicht. (..) Da ist die sprachliche Richtigkeit schon wichtig.“ (LP2, Interview 2)

Demnach ist nach wie vor eine Diskrepanz zwischen einem handlungsorientierten Ansatz und dem Stellenwert der Sprachrichtigkeit zu verorten, wodurch sich bei Bewertungen ein gewisses Dilemma bei den Lehrpersonen ergibt.

7.2.6 Anregungen und Wünsche

In den Gesprächen wurden seitens der befragten Französischlehrenden einige Anregungen geäußert, wie die Beurteilungspraxis mit Hilfe des Beurteilungsrasters auf Sprachniveau B1 ihrer Ansicht nach verbessert werden könnte. An erster Stelle wurde mehrfach eine Vereinfachung und Präzisierung des vorliegenden Rasters genannt:

„Ich muss ehrlich sagen, mich macht das ja auch verrückt diese Umschreibungen, also (.) ich merke mir das ja auch nicht alles und so viel. [...] Ich hätte da vielleicht lieber nur ganz kurze und präzise Anweisungen [...]. Ja, (..) also prinzipiell, wenn man den Raster überarbeiten könnte und noch präziser machen würde, das wäre schon schön, denke ich jetzt (.) für mich. Eine Anleitung prinzipiell finde ich schon sehr gut und wünschenswert aber in dieser Form nicht wirklich.“ (LP3, Interview 3)

Somit besteht eine Forderung nach Präzisierung und Überarbeitung des bisherigen Beurteilungsrasters. Eine weitere Idee bezüglich einer andersartigen Beurteilungspraxis wird von Lehrperson 2 geäußert, wobei die Anwendung externer EvaluatorenInnen ein Wunsch wäre, wodurch die transparente Basis ohne soziale und individuelle Bezugsmaßstäbe zur Anwendung käme:

„Vielleicht noch als wichtiger Punkt, was für eine objektivere Leistungsbeurteilung gut wäre, ist, wenn der Lehrer den Schüler nicht selbst beurteilt, sondern eine externe Lehrperson, da kann man das persönliche wirklich weglassen, das stört mich schon.“
(LP2, Interview 2)

8 Zusammenfassung

In diesem zweiten Teil wurde sich dem Themengebiet empirisch angenähert, wobei es zunächst galt, das Forschungsdesign zu definieren und zu beschreiben, um zu verdeutlichen, wie die in weiterer Folge präsentierten Daten erhoben und verarbeitet wurden. Im Rahmen der Forschung fanden drei problemzentrierte Interviews statt, bei denen auf eine ausreichende Diversifizierung der Personen geachtet wurde. In den Ergebnissen konnten Unterschiede ausgemacht werden, wobei größtenteils Konsens unter den Lehrenden herrschte und selbige Meinungen und gar Wortlaute geäußert wurden.

Festgehalten werden kann, dass zwei Gruppen von Lehrenden auszumachen sind: Jene Personen, die bereits viel Erfahrung aufweisen und das Bewerten von Schülertexten über lange Zeit ohne Bewertungshilfen ausübten, und eine jüngere Generation, die bereits von Beginn an die Beurteilungspraxis mittels Beurteilungsraster kennenlernte, wobei hier bereits unterschiedliche Stellungen bezogen werden. Letztere Gruppe steht der Anwendung des Rasters prinzipiell positiver gegenüber als erstere Gruppe, die sich an die Änderungen adaptieren musste. Das Instrument einer Bewertungshilfe an sich wird allerdings bei allen Befragten als äußerst wünschenswert begrüßt, wobei die vorliegende Form kritisiert wird, da die Formulierungen als zu wenig präzise und aussagekräftig wahrgenommen werden. Daraus resultiert, dass alle befragten Lehrkörper ihre Bewertungen teilweise an den Punktwerten der Niveaustufen festmachten, wobei hier eine Orientierung an der Notenskala des österreichischen Schulwesens stattfindet.

Der Beurteilungsraster wird als Leitfaden und transparente Richtlinie aufgefasst und dient als Argumentationshilfe, warum eine bestimmte Note zustande gekommen ist. Hierbei wurde betont, dass der subjektive Interpretationsraum, der einen gewissen Spielraum bei der Bewertung offen lässt, als eindeutig vorhanden zu betrachten ist und positiv angenommen wird. Dadurch ist es den Lehrpersonen möglich, andere Bezugsnormen als die sachlichen Maßstäbe der Deskriptoren miteinbeziehen zu können und so individuelle Lernfortschritte seitens der SchülerInnen honoriert werden können. Weiters wurde angemerkt, dass eben jene Form des Rasters seine Anwendung in allen möglich Referenzrahmen zuließe und für die jeweiligen Maßstäbe angewandt werden kann.

Aufgezeigt wurde ebenso eine Diskrepanz zwischen dem Stellenwert der Schreibrichtigkeit und der Handlungsorientierung, wobei im Fremdsprachenunterricht die Handlungsorientierung zunächst als bedeutender angesehen wird, in der Bewertung vergleichsweise zur Sprachrichtigkeit aber kaum Berücksichtigung findet.

IV. Conclusio

Im Fokus dieser Diplomarbeit lag die Annäherung an die Leistungsbeurteilung schriftlicher Performanzen im Französischunterricht mittels Beurteilungsraster des BIFIE auf Sprachniveau B1, wobei insbesondere diese Beurteilungshilfe sowie ihre Anwendung im schulischen Fremdsprachenunterricht in den Mittelpunkt gesetzt wurde.

Ziel dieser vorliegenden Arbeit war es einerseits den Kontext theoretisch zu erarbeiten, indem zunächst der weitreichende und für diese Thematik grundlegende Bereich des Assessments und die damit in Verbindung stehenden Subdisziplinen dargeboten wurden. Bereits in dieser ersten Annäherung konnten die Komplexität und der Umfang dieses Arbeitsfeldes veranschaulicht werden, da vielfältige Formen sowie Funktionen in Hinsicht auf Bewertungen im Kontext Schule vorzufinden sind. In einem allgemeinen Verständnis wird Bewertung oft als eine simple Tätigkeit wahrgenommen, bei der zwischen *richtig* und *falsch* zu unterscheiden ist, der vielfältigen Formen und Möglichkeiten sind sich aber in der Allgemeinheit nur wenige bewusst. Mindestens ebenso, wenn nicht gar komplexer, verhält es sich mit der Begrifflichkeit der Leistung, wobei in diesem Rahmen auf Leistungen von Schülerinnen und Schülern in Bezug auf den Fremdsprachenunterricht Bezug genommen wurde.

Um eben jene Leistungen einheitlich messen und bewerten zu können, sind im Zuge der Implementierung von Standards, also einheitlichen Lehr- und Lernzielen, an Standards orientierte Bewertungshilfen entwickelt worden, denen bestimmte Kriterien, die es zu erfüllen gilt, zugrunde liegen. In diesem Sinne ist von standardorientierter und kriterienbasierter Leistungsbeurteilung zu sprechen. Das hier im Mittelpunkt der Forschung stehende Bewertungsinstrument stellt eine solche Hilfe dar, wodurch eine einheitliche, schulübergreifende Leistungsbeurteilung im österreichischen Fremdsprachenunterricht erreicht werden will.

Wie in der theoretischen Ausarbeitung bereits aufgezeigt wurde, ist jegliche Bewertung, die nicht objektiv ist, an eine Relativierung geknüpft. Bei der Beurteilung von schriftlichen Performanzen handelt es sich um subjektive Bewertungsverfahren, da das Urteil einer Lehrperson gefragt ist. Hierbei sind unterschiedliche Bezugssysteme zu verorten, die zur Anwendung kommen können. Unter anderem ein kriterialer Maßstab, wobei sich Beurteilende an einer sachlichen Bezugsnorm, gewissen Kriterien, die zu erfüllen sind, orientieren, wie es im Falle des Beurteilungsrasters plausibel scheinen mag.

Mit Hilfe einer auf der theoretischen Kontextualisierung aufbauenden empirischen Untersuchung wollte der forschungsleitenden Frage „*Wie lassen sich die Einstellungen Französischlehrender dem Beurteilungsraster des BIFIE auf Sprachniveau B1 gegenüber beschreiben?*“ nachgegangen werden. Dabei galt es, im Zuge der problemzentrierten Interviews einerseits die allgemeine Haltung zu erheben und andererseits im Spezifischeren zu erfahren wie klar diese kriteriale Bezugsnorm den Französischlehrenden erscheint, wobei folgende Subfrage definiert wurde: „*Wie verständlich und klar erscheinen Fremdsprachen-lehrenden die Formulierungen der Deskriptoren?*“ Ein weiteres Ziel war es, eine praktische Perspektive einzunehmen und zu erfahren inwiefern diese Bewertungshilfe tatsächlich in das Assessment von SchülerInnenleistungen in der Unterrichtsrealität integriert wird, wobei eine zweite Subfrage in der Forschung Berücksichtigung fand: „*Wie findet der Beurteilungsraster Anwendung bei der Leistungsbewertung schriftlicher Performanzen im Französischunterricht?*“.

Im Rahmen der empirischen Forschung konnte festgestellt werden, dass die Einstellung der Lehrpersonen in Hinblick auf einheitliche Beurteilungsinstrumente prinzipiell sehr positiv ist, da sie eine transparente Richtlinie und Anleitung durchgehend als wünschenswert ansehen. Störend empfunden wurde allerdings großteils eine Objektivität, die solchen kriterienbasierten Beurteilungen zugesprochen wird, wobei dieses Postulat keinerlei Gültigkeit besitzt. Die Subjektivität bei der Bewertung von SchülerInnentexten spielt eine zentrale Rolle, welche die Lehrenden jedoch ebenfalls positiv verstehen. Durch teilweise unpräzise und wenig aussagekräftige Formulierungen bei der Beschreibung einzelner Deskriptoren in unterschiedlichen Niveaustufen, sind subjektive Interpretationsräume aufzufinden, wobei den Beurteilenden hier Freiraum geboten wird, nach eigenem Ermessen und Empfinden die Leistung einzustufen. Diesbezüglich gaben die Befragten an, dass individuelle Lernfortschritte honoriert werden müssen und auch Einfluss auf die Bewertung mittels Raster haben. Dadurch ist zu konstatieren, dass trotz national einheitlicher Beurteilungsverfahren, dennoch Freiraum für Individualisierung in der Leistungsbeurteilung bleibt und bei der Verwendung des Beurteilungsrasters keine rein kriteriale Bezugsnorm zu verorten ist, da sämtliche Bezugssysteme, welche Personen, die gleichzeitig Lehrende und Beurteilende sind, nicht gänzlich ausblenden können, stets in unbewusster Weise Einfluss auf die Handlung des Bewertens nehmen.

Nicht nur eine transparente Anleitung und Richtlinie stößt auf zusagende Akzeptanz, sondern ebenso die Aufteilung des Rasters in die vier Kategorien, deren Berücksichtigung als sinnhaft und logisch nachvollzogen wird, während die Gleichwertigkeit dieser auf Ablehnung trifft, da konsenshaft die Meinung vertreten wird, das Textformat sei weniger bedeutend im Vergleich zu sprachlichen Kompetenzen, die den Kern des Fremdsprachenunterrichts repräsentieren.

Die Formulierungen der Deskriptoren wurden kritisch bemängelt, da einerseits die Anzahl der Beschreibungen als zu umfassend angesehen wird, wohingegen die Ausdifferenzierung aufgrund der unpräzisen Formulierungen schwer fällt. Diesbezüglich wird der Beurteilungsraster heftig beanstandet, denn eine Anwendung ist aus diesem Grund wenig effizient und wird vielmehr als sinnfreier Zeitaufwand aufgefasst der nicht in ausgeglichener Relation steht. Hierbei konnte ausgemacht werden, dass Lehrperson 1 und 2, welche sich bereits von Anbeginn mit der Beurteilungspraxis mittels Raster konfrontiert sahen, eher dazu neigen, den Raster häufiger zu verwenden, auch wenn der Zeitaufwand kritisiert wird, wohingegen Lehrperson 3, die bereits mehr Erfahrung aufweist und für welche der Raster eine Neuerung in der Bewertungskultur darstellt, eine weitaus höhere Abneigung zum Ausdruck brachte.

Aus den Interviews ging interessanterweise hervor, dass sich die befragten Lehrenden weniger an den unpräzise empfundenen Formulierungen orientieren, sondern eher an den Punktwerten der Niveaustufen, welche an die Notenskala des österreichischen Schulsystems angeglichen werden. Somit ist grundsätzlich eine subjektive Einschätzung von vordiagnostischem Charakter und die Einordnung in dieses System ausschlaggebend bei der Beurteilung einer Performanz und weniger der Bezug auf die Deskriptoren, die im Beurteilungsraster aufgefächert werden. Wodurch, wie in der Theorie aufgezeigt wurde, zu bestätigen ist, dass es weitgehend in der Hand der LehrerInnen liegt, was als gute Leistung zählt.

In Hinsicht auf die Handlungsorientierung im Fremdsprachenunterricht wird diese als teilweise wichtigstes Gut eingeschätzt, wohingegen der Sprachrichtigkeit ebenfalls ein hoher Stellenwert beigemessen wird. Interessant erscheint, dass im Prozess der Bewertung, wie aus den Interviews hervorging, lediglich die Sprachrichtigkeit und weniger die Handlungsorientierung beachtet wurde, da diese nicht klar im Bewertungsinstrument deklariert ist.

Fazit ist also, dass der Raster unter den Befragten eher bei den beiden Lehrenden mit wenig Erfahrung zur Anwendung kommt, die ihn auch für die Bewertung von Hausübung heranziehen, wohingegen die erfahrene Lehrperson dazu neigt, nur in offiziellen Situationen das Beurteilungsinstrument zu berücksichtigen. Der Zeitaufwand, den der Raster bei einer Korrektur abverlangt, wird von allen Interviewten beanstandet und diesbezüglich ist große Unzufriedenheit zu verorten. In Bezug auf die Kritik der unpräzisen Formulierungen herrschte Einstimmigkeit, weshalb einige Forderungen beziehungsweise Wünsche zu einer überarbeiteten, vereinfachten Form mit mehr Präzision geäußert wurden, sodass der Beurteilungsraster ohne negative Ressentiments Verwendung finden kann.

Im Zuge dieser Diplomarbeit schien es nicht möglich weitere interessante Punkte zu erforschen und zu behandeln, ohne den Rahmen zu übertreten, wobei für weitere Forschungen in diesem Gebiet vielfältige Anknüpfungspunkte auszumachen wären. Ursprünglich wollte die empirische Annäherung im Rahmen einer quantitativen sowie qualitativen Forschung durchgeführt werden, wobei große Kooperationsbereitschaft bei Lehrenden, deren Zeit begrenzt ist, notwendig wäre. Hierfür ist bereits ein operationalisierter Fragebogen erarbeitet worden und liegt dem Verfasser dieser Arbeit vor. Dieser ist somit für eine aufbauende quantitative Forschung verwendbar. Aus Mangel an kooperationsbereiten Lehrpersonen in ausreichendem Ausmaß, wurde dieses Vorhaben letztendlich lediglich mittels qualitativer Methoden umgesetzt. Im Rahmen eines Methodenmix aus quantitativen und qualitativen Instrumentarien der Sozialforschung könnten weitgehendere Einblicke in dieses Forschungsfeld erlangt werden.

Darüber hinaus wäre eine Auseinandersetzung und Befragung einer weiteren betroffenen Zielgruppe von großem Interesse: die Lernenden. Hierbei könnten Einstellungen und Haltungen der Schülerinnen und Schüler in Hinsicht auf den Beurteilungsraster erhoben werden und jenen der Lehrenden gegenübergestellt oder mit diesen in Verbindung gesetzt werden. Eine weitere Möglichkeit besteht in einer umfassenden linguistischen Untersuchung, wie es in der vorliegenden Arbeit exemplarisch in Unterkapitel 4.4 vorgenommen wurde, indem der Beurteilungsraster hinsichtlich seiner qualitativen Ausprägungen der einzelnen Niveaustufen und deren Abgrenzung untereinander analysiert wird. Außerdem konnte in den Interviews nach wie vor ein Dilemma zwischen Sprachrichtigkeit und handlungsorientierter Kommunikationsfähigkeit verzeichnet

werden, weshalb auch dies als spannende Perspektive im Bereich der Leistungsbeurteilung anzuführen ist.

V. Résumé en langue française

Ce travail est rédigé en vue d'analyser l'évaluation des textes écrits à l'aide de la grille d'évaluation dans le contexte des cours de français dans les écoles autrichiennes pour le niveau de langue B1. Dans ce travail je souhaite examiner, d'un côté, l'évaluation dans le cadre des cours de langues étrangères et plus précisément la grille du BIFIE sous un angle théorique et d'un autre côté la position des enseignant(e)s à propos de cette grille.

J'ai décidé d'approfondir ce sujet en raison de l'importance de l'évaluation dans le contexte scolaire et dans le souhait d'accomplir cette évaluation d'une manière juste. Ces thèmes ont suscité mon intérêt et demandent une étude plus détaillée. Évidemment, le travail s'appuie sur l'élaboration du vaste sujet de l'évaluation scolaire et par conséquent, il contient des éléments et interprétations qui ont déjà été publiés dans des ouvrages scientifiques.

Dans ce mémoire, il est donc question d'examiner la position des enseignant(e)s de français quant à ce domaine, en se penchant tout particulièrement sur la grille du BIFIE. Ce travail se divise en deux grandes parties afin de donner d'une part un aperçu théorique et d'autre part une analyse empirique.

La première partie traitera tout d'abord de la base du sujet de l'évaluation et dans une seconde partie de l'évaluation dans le cadre scolaire. Pour cette raison, il semble indispensable de prendre en compte la performance d'un/e élève et des facteurs qui l'influencent. La performance est non seulement le résultat de facteurs externes, mais aussi de l'évaluation d'un/e professeur/e, c'est pourquoi il est essentiel d'aborder les valeurs de références jouant un rôle central dans ce contexte. Dans le cadre de la standardisation du système scolaire en Autriche, le but était de créer une évaluation commune en définissant pour cela une grille d'évaluation déterminée par rapport aux standards minimums acquis en fin d'étude et s'appuyant sur des critères concrets et unitaires.

La deuxième partie sera consacrée à l'étude empirique pour pouvoir explorer les avis des enseignant(e)s sur ce domaine de travail et particulièrement sur la grille d'évaluation au niveau de langue B1 ainsi que la manière dont elle est utilisée pour l'évaluation. Pour mieux comprendre les résultats de la recherche, ils seront élaborés thématiquement et divisés en différentes catégories afin de les illustrer réorganisés et résumés.

La partie théorique

L'évaluation

L'évaluation est un besoin fondamental pour un être humain afin d'atteindre une optimisation. Ce processus d'évaluation est nécessaire car cela permet d'agir dans les situations complexes de notre vie. Grâce à notre compétence cognitive, il nous est possible de classer les informations d'une manière inconsciente, ce qui nous rend capable de trouver des réactions adéquates à chaque situation. Évaluer signifie donc déterminer la valeur de quelque chose en relation avec une certaine référence. De ce fait, l'évaluation mesure et caractérise une situation ou performance.

Une grande importance est accordée à l'évaluation dans le contexte scolaire, réalisée par l'utilisation de méthodes et outils très variés, qui peuvent être divisés en trois domaines : le système d'éducation, l'enseignant(e) et l'élève. L'attention est mise sur la plus grande partie de ce travail, c'est-à-dire la catégorie des professeur(e)s. Les formes de l'évaluation sont très différentes et dépendent de la fonction envisagée. Parmi ces formes, la distinction entre l'évaluation formative et l'évaluation sommative est la plus connue. La première vise à améliorer l'apprentissage en cours tandis que l'évaluation sommative a pour fonction de vérifier, après une formation, si l'acquisition des compétences visées a été faite et dans quelles proportions. Une autre distinction se fait au niveau des références où une performance peut être évaluée en relation avec d'autres performances ou en relation avec des critères prédéfinis. Si le jugement d'un/e enseignant/e est demandé, l'évaluation est subjective tandis que l'objectivité n'est garantie que si la performance est bonne ou mauvaise sans avoir besoin d'une estimation d'un/e expert/e, comme c'est exigé dans le cadre d'une performance de langue productive, donc un texte par exemple.

Comme il s'agit ainsi en l'occurrence de l'évaluation d'un texte écrit à l'aide d'une grille, cette évaluation est sommative, car l'évaluateur ou l'évaluatrice évalue la maîtrise des compétences pour prendre une décision de classement sur la base de la performance d'un/e élève. Ainsi, l'évaluation est subjective puisqu'un jugement est porté par l'enseignant(e), c'est-à-dire non objectif. Cette décision, prise en cours du processus de l'évaluation, doit se référer à certains critères définis par la grille d'évaluation.

La performance scolaire en langue étrangère

En ce qui concerne l'évaluation des performances scolaires dans le cadre du cours de langue étrangère, il est indispensable de mentionner la complexité de l'apprentissage d'une langue étrangère, qui exige plein de compétences différentes. Les compétences d'un utilisateur ou bien d'une utilisatrice d'une langue étrangère comprennent aussi bien des compétences générales que des compétences communicatives langagières.

A ce sujet il faut nommer le savoir ou bien la connaissance du monde, les aptitudes pratiques et le savoir-faire ainsi que le savoir-être et le savoir-apprendre qui comporte la connaissance de la langue et de la communication. Ce qui est vu au premier plan, en général, ce sont les compétences communicatives langagières se composant des compétences linguistiques (lexicale, grammaticale, sémantique, phonologique, orthographique et orthogénique), des compétences sociolinguistiques et des compétences pragmatiques.

Mais l'évaluation directe des compétences semble difficilement réalisable. Il faut donc différencier les termes « compétence » et « performance ». Par compétence nous entendons les capacités d'un/e élève à exercer une certaine activité. Cela représente le total des connaissances et des aptitudes dont une personne est capable. Au contraire, la performance est opposée à la compétence, car elle représente le comportement émergé, qui peut être observé et mesuré puis par la suite évalué.

La base d'une évaluation est ainsi la performance d'un/e élève. Ces réalisations individuelles, par exemple une activité sportive, un dessin ou bien un texte oral ou écrit, dépendent des facteurs externes : l'effort, le talent et les ressources.

Chaque évaluation est mise en relation avec une certaine référence. Dans le contexte scolaire trois références principales sont prises en compte : Premièrement, il faut nommer la référence d'un critère. Dans ce cas les performances sont évaluées au niveau de certains critères concrets qu'il faut atteindre. Deuxièmement, l'évolution positive d'un/e élève est prise en tant que référence. Cela veut dire que le niveau atteint n'est pas le plus important tant qu'on remarque une progression, puisqu'il s'agit d'un indicateur d'investissement et d'efforts. L'individu lui-même est alors pris comme référence. Troisièmement, il est possible de se référer à la moyenne des performances d'un groupe, donc à une référence sociale. A l'école ce critère est souvent représenté par les élèves d'une classe. A cette occasion une certaine performance est mise en relation avec d'autres

performances, celle des collègues, indépendante d'une progression individuelle ou de critères concrets. En réalité, ces trois références illustrées sont appliquées dans tous le processus de l'évaluation scolaire.

La standardisation de l'évaluation scolaire

La troisième partie du contexte théorique se penchera sur la standardisation du système de l'éducation et mettra l'accent en particulier sur l'évaluation commune. Cette standardisation s'est produite à deux niveaux : elle touche les objectifs pédagogiques et, partant de là, les critères d'évaluation. Dans le cadre du baccalauréat central, les performances des élèves sont évaluées à partir des mêmes tâches, des mêmes exercices, représentant les standards éducatifs, c'est-à-dire le minimum à atteindre. Comme les objectifs pédagogiques sont standardisés, il est nécessaire de faire de même en ce qui concerne l'évaluation des performances de ces objectifs.

Quant aux tâches des compétences réceptives, l'évaluation se fait d'une manière objective alors que les exercices de production de langue posent problème :

Les productions, écrites ou orales, sont des performances individuelles qu'il faut évaluer à l'aide d'un jugement d'évaluateur ou d'évaluatrice, il s'agit donc d'une évaluation subjective. Pour la faire le plus objectivement possible, on élabore une grille d'évaluation bien construite. Grâce à cette grille du BIFIE, les textes écrits seront évalués en vue des mêmes critères :

„La grille permet [...] à différents correcteurs/évaluateurs de se référer aux mêmes critères, que la notation se fasse immédiatement après la production ou en différé, des jours, voire des années plus tard. L'évalué doit pouvoir en prendre connaissance avant l'évaluation, de façon à comprendre sur quoi sa performance va être évaluée et donc à pouvoir agir en conséquence.“ (TAGLIANTE 2005: 24)

La grille d'évaluation du BIFIE

Après avoir abordé le domaine fondamental de l'évaluation scolaire et avoir révélé la différence entre compétence et performance, nous allons traiter de la standardisation dans le système éducatif autrichien en tenant compte du cours de français. L'utilisation d'une grille doit donc être introduite dans ce contexte. Pour explorer encore plus cet outil, l'accent sera mis sur la grille d'évaluation qui est appliquée dans le cadre du cours de français au niveau B1.

À l'aide d'une analyse descriptive de cette grille, il sera possible de montrer les avantages et les inconvénients qu'elle comporte. De plus, j'aborderai une question difficile, à savoir s'il est justifié de parler de cette grille comme d'un outil objectif. Nous examinerons cela de plus près dans l'étude empirique.

La grille d'évaluation du BIFIE (cf. annexe) est composée de quatre catégories ayant toutes la même valeur, entre elles : l'accomplissement de la tâche, la structure, la variété des moyens linguistiques et la précision linguistique. Chacune de ces classifications comporte différents critères à prendre en compte dans chacune de ces catégories. La grille est alors divisée en quatre parties à l'horizontale, et en onze niveaux à la verticale, ces derniers représentant les points à gagner. Dans chaque catégorie, les critères sont décrits par un énoncé illustrant ce qui est nécessaire pour atteindre ce niveau.

L'enseignant doit choisir quel niveau de critère correspond le mieux à la performance. La moyenne des points dans une catégorie est calculée et la note finale est composée des points des quatre catégories.

Le but principal d'une grille est d'obtenir, en fait, plus de fiabilité concernant une évaluation subjective. En analysant linguistiquement les descripteurs de la grille du BIFIE, il a été révélé que certains soient en effet assez imprécis a été révélé. C'est pourquoi on peut constater que les énoncés sont interprétés de manière très différente par les enseignant(e)s et qu'une subjectivité reste dans l'évaluation avec l'utilisation d'une grille. A cause des descriptions vagues de différents niveaux, il est intéressant de savoir, si les évaluateurs ou bien évaluatrices se réfèrent uniquement aux critères définis par la grille ou également à d'autres références possibles.

Dans ce cadre, il faut poser les questions suivantes :

« Avons-nous la même image d'une performance adéquate ? Que signifie 'un bon travail' pour nous ? Quelle est la différence entre 'adéquat', 'bon' et 'pertinent' concernant un critère de compétence communicative langagière ?

Comme l'évaluation d'une langue est un processus complexe, cette subjectivité liée à la grille d'évaluation sera examinée dans la partie empirique à travers des interviews avec des expert(e)s dans ce domaine, et des professeur(e)s de français qui doivent utiliser cette grille pour évaluer les élèves.

L'étude empirique

L'étude empirique permet de combiner des recherches approfondies présentées dans la partie théorique à l'aide d'une étude détaillée afin d'intégrer la recherche et la pratique dans le système scolaire. La grille sera examinée et analysée d'un point de vue théorique qui, par la suite, sera comparée avec les positions des enseignant(e)s par rapport à elle, c'est-à-dire nous verrons son utilisation de manière concrète.

Les questions de recherche

Les questions de recherche définies dans ce travail regroupent ce qu'il est prévu d'étudier dans le cadre du mémoire. L'objet de recherche est abordé dans la partie théorique alors que l'étude empirique vise à examiner les problèmes présentés dans la partie préalable et à illustrer la pratique d'évaluation des enseignant(e)s de français.

Concernant la question de recherche centrale, on cherche à explorer quelles positions sont prises par les professeur(e)s envers la grille d'évaluation du BIFIE. C'est la question la plus large que nous puissions poser par rapport à cette étude :

« Comment pouvons-nous décrire la position des enseignant(e)s de français vis-à-vis de la grille d'évaluation du BIFE au niveau de langue B1 ? »

En explorant le sujet de cette question centrale, deux sous-questions surviennent afin de pouvoir centrer la recherche et illustrer le sujet dans toute sa complexité :

« A quel point, les descripteurs, semblent-ils précis et compréhensibles en référence à un certain niveau ? »

« Comment cette grille d'évaluation est-elle utilisée en évaluant des textes écrits au cours de français ? »

Ces sous-questions aident à spécifier la question de recherche centrale et à guider l'étude en direction de l'objectif de recherche. La méthodologie s'est alignée sur l'objectif et sur les questions de recherche.

La méthodologie

Afin de pouvoir répondre aux objectifs initiaux définis dans les questions de recherche et d'aborder le sujet de ce travail, c'est-à-dire une recherche qualitative, il est adéquat de se baser sur des évaluations d'un texte concret suivies par des interviews avec des professeur(e)s de français, pour explorer les informations prévues quant au phénomène central. Le but est alors d'obtenir un aperçu détaillé des avis des enseignant(e)s concernant cet outil d'évaluation qu'est la grille BIFIE. Les participants de l'étude, les évaluateurs et évaluatrices interviewé(e)s, ont été choisis au niveau de l'âge, de l'expérience, du milieu de travail entre autres, pour garantir une certaine diversification parmi les individus.

L'évaluation d'une performance

Un texte concret (cf. annexe) sera évalué à l'aide de la grille d'évaluation, se trouvant au centre du mémoire, par différent(e)s enseignant(e)s. Il est intéressant de comparer les résultats de manière à savoir s'ils sont très variés ou plutôt semblables. Le choix du texte dépend du niveau concernant les différentes catégories. Il ne s'agit pas d'une performance exceptionnellement bonne ou mauvaise, c'est pourquoi la subjectivité peut être bien examinée.

L'interview

A l'aide des interviews, je souhaite approfondir la matière et explorer la position des enseignant(e)s, ainsi que la signification de la grille pour eux. Avec cette stratégie d'investigation, des données intéressantes pour l'étude ont été recueillies. Par ailleurs, cette recherche qualitative soutient la théorie et aide à mettre en lumière de nouveaux aperçus sur ce qui est déjà connu du contexte théorique. Il s'agit donc également d'une collecte de preuves pour montrer de nouveaux faits recueillis dans le cadre des interviews.

Les interviews menées seront analysées par la suite en référence à Philipp MAYRING et son modèle de l'analyse de contenu. Cela veut dire que les données rassemblées seront regroupées en catégories, résumées et discutées compte tenu des informations de la partie théorique de ce travail.

Les résultats

L'évaluation d'une performance

Dans la première étape, les résultats de l'évaluation d'une performance de production de langue sont analysés. Trois enseignants ont évalué une seule et unique performance. Le total des points est sur 40 et l'étude montre que les résultats vont de 22, 27 et 32 points. Cela veut dire qu'il y a une différence de 10 points, ou bien de 25 %, entre l'évaluation de la même performance à l'aide de la grille d'évaluation.

Comme démontré dans la partie théorique, ce résultat n'est pas surprenant car plusieurs recherches sur ce sujet d'évaluation de performance individuelle de langue ont montré qu'une différence de résultat découle d'une évaluation subjective. Ceci est dû aux différentes attentes de l'évaluateur ou de l'évaluatrice. A partir de ce résultat, on peut constater que la grille représente un instrument transparent mais qu'elle ne garantit pas une évaluation objective.

Les interviews : la position des enseignant(e)s

L'analyse des données précise que toutes les personnes interviewées ont une certaine directive unitaire pour garantir l'objectivité d'une évaluation en prenant en compte les mêmes catégories : „Ich finde es gut, dass es eine Basis gibt, auf die man sich beziehen kann“ (LP1, Interview 1). Une telle base pour se guider semble utile, car l'évaluation des performances individuelles de production de langue est vue comme un défi en raison des différentes performances possibles : „Ja, es ist natürlich eine Herausforderung“ (LP2, Interview 2). En outre, une grille commune garantit une transparence des critères de l'évaluation pour tous, y compris pour les professeur(e)s, les élèves et les parents.

La grille d'évaluation du BIFIE est par contre fortement critiquée à cause de différents défauts. D'une part l'outil semble trop vaste et imprécis en raison de plusieurs descripteurs divisés en six seuils de niveau, ce qui rend son application moyennement pratique et qui exige un travail intensif.

„Mit den neuen Beurteilungsrastern und Korrekturblättern ist es jetzt aufwendiger geworden [...]. Es nimmt jetzt viel mehr Zeit in Anspruch [...]. Denn alle diese einzelnen Punkte und alle diese Bänder durchzugehen, das geht nicht, das dauert einfach zu lange. [...] Das ärgert ein wenig. Der Aufwand ist einfach zu groß [...]. Also grundsätzlich begrüße ich so eine Basis aber ich würde es vereinfachen.“
(LP2, Interview 2)

La division en quatre catégories paraît logique d'après les interviewés mais, ils ne sont pas tout à fait d'accord sur le fait que chaque catégorie ait la même valeur. En ce qui concerne la catégorie de la structure, tous les enseignants interviewés approuvent que ce soit la catégorie la moins importante en comparaison avec l'accomplissement de la tâche, la variété des moyens linguistiques et la précision linguistique.

„Die Kategorien „Spektrum sprachlicher Mittel“ und „Sprachrichtigkeit“ würde ich nun mal als die wichtigsten bezeichnen [...], weil ja das den Kern einer Sprachkompetenz am genauesten trifft, wobei die anderen auch wichtig sind. Also sie sind nicht wegzudenken aber sie sind nun mal nicht so wichtig wie die beiden anderen, die ich gerade genannt habe. [...] Wenn die „Erfüllung der Aufgabenstellung“ nicht passt, dann kann das sprachlich der beste Text der Welt sein [...], deswegen kann diese Kategorie als Voraussetzung angesehen werden.“
(LP1, Interview 1)

Les compétences communicatives langagières sont estimées comme les plus importantes pour une production écrite en cours de français. L'accomplissement de la tâche représente surtout une condition indispensable afin que le texte soit réussi.

Quant aux descripteurs des différents niveaux séparés qui se ressemblent et qui ne sont pas différenciés très précisément comme la partie théorique et l'analyse descriptive ont montré, l'étude empirique pouvait confirmer ce fait et a donné la réponse à l'une des sous-questions de recherche. Alors, les descripteurs sont soumis à une conception subjective et c'est pourquoi les critères présentés dans la grille sont interprétés de manière différente selon chaque personne :

„Schade sind halt diese Wischiwaschi-Sätze.“ (LP3, Interview 3)

L'interprétation de l'évaluateur ou l'évaluatrice est donc cruciale pour l'évaluation d'une performance :

„Ja, das ist ein bisschen schwammig. Das ist ein wenig abhängig von der Interpretationsweise eines jeden Lehrers, der das liest.“ (LP1, Interview 1)

Ce fait n'est guère surprenant car la théorie a déjà démontré que l'évaluation d'un texte varie en fonction de la personnalité et de la conception de l'évaluateur ou de l'évaluatrice.

Ce qui était vraiment intéressant en interrogeant les enseignant(e)s, est le fait que chacun d'entre eux se réfère à la valeur des points, mis en relation avec le système de notation aux écoles autrichiennes, à cause de cette imprécision des descriptions des critères.

„Ja, also ich, ich richte mich da auch sehr stark nach der Punktzahl, je nachdem in welchem Band ich bin und versuche die Leistung darin einzuordnen nach meinem Maßstab, wie es zum Beispiel bei sechs Punkten für einen solchen Rahmen (.) ahm, ja, angemessen ist, eben noch eine positive Leistung. So versuche ich das zu interpretieren.“ (LP3, Interview 3)

Il en résulte que les professeur(e)s estiment le niveau et la note de la performance à partir de leur expérience et de leur conception personnelle. Les niveaux de la grille sont plus ou moins vus comme soutien argumentatif pour leur jugement.

„Der Raster ist eine Argumentationshilfe, das kann manchmal sehr gut sein. Eine transparente Anleitung, auf die kann man sich auch berufen und dafür bin ich auch dankbar.“ (LP3, Interview 3)

Concernant l'utilisation de la grille, il était déjà mentionné qu'elle est considérée comme ayant des problèmes pratiques selon deux personnes interviewées. Deux positions différentes peuvent être vues à ce sujet. Prenons la personne 1, un jeune enseignant qui a appris l'évaluation des textes dès le début à l'aide de la grille. Celui-ci l'utilise presque tout le temps tandis que la personne 3, une enseignante avec beaucoup d'expérience, ne l'utilise que dans des situations officielles quand il faut l'appliquer impérativement.

Quant aux références, il était déjà clair que les enseignants ne se réfèrent pas uniquement aux critères définis dans la grille mais aussi à d'autres références possibles, par exemple la progression individuelle d'un/e élève ou une référence sociale, la classe ou bien l'école, donc le contexte dans lequel l'évaluation est faite.

Conclusion

Pour conclure, on constate que la position des enseignant(e)s vis-à-vis la grille d'évaluation du BIFIE est assez critiquable en raison des défauts énumérés préalablement. Elle est donc utilisée de manière très diverses et variées parmi les professeur(e)s de français en Autriche.

Ce que nous avons remarqué dans le cadre de cette étude, c'est que cet outil d'évaluation ne garantit pas d'évaluation commune dans un contexte interscolaire. Les catégories et les critères sont les mêmes pour tous les élèves et il s'agit d'un instrument transparent mais l'évaluation est surtout dépendante de la conception de l'évaluateur ou bien de l'évaluatrice. Chaque évaluation est toujours liée au contexte dans laquelle elle sera faite.

Dans les interviews avec les professeur(e)s le souhait d'une nouvelle édition remaniée de la grille a été manifesté : elle devrait être plus simple dans l'application et plus précise dans la formulation :

„Ich muss ehrlich sagen, mich macht das ja auch verrückt diese Umschreibungen, also (.) ich merke mir das ja auch nicht alles und so viel. [...] Ich hätte da vielleicht lieber nur ganz kurze und präzise Anweisungen [...]. Ja, (..) also prinzipiell, wenn man den Raster überarbeiten könnte und noch präziser machen würde, das wäre schon schön, denke ich jetzt (.) für mich. Eine Anleitung prinzipiell finde ich schon sehr gut und wünschenswert aber in dieser Form nicht wirklich.“ (LP3, Interview 3)

Ce sujet de l'évaluation est omniprésent dans le quotidien scolaire pour les élèves ainsi que pour les enseignant(e)s et c'est pourquoi des recherches complémentaires seraient intéressantes, comme par exemple la position des élèves à propos de cette grille d'évaluation.

VI. Bibliographie

VI.1 Sekundärliteratur

- ADAMZIK, Kirsten (2005): Textsorten im Fremdsprachenunterricht – Theorie und Praxis. In: ADAMZIK, Kirsten/KRAUSE, Wolf-Dieter (Hrsg.): Text-Arbeiten. Textsorten im fremd- und muttersprachlichen Unterricht an Schule und Hochschule. S. 205 - 237. Tübingen: Narr.
- ALDERSON, Charles/CLAPHAM, Caroline/WALL, Dianne (1995): Language test construction and evaluation. Cambridge: Cambridge University Press.
- BENISCHEK, Isabella (2006): Leistungsbeurteilung im österreichischen Schulsystem. Wien [u.a.]: LIT.
- BERGER, Armin (2014): Validating analytic rating scales. A multi-method approach to scaling descriptors for assessing academic speaking. Wien: Dissertation.
- BIRKEL, Peter (2003): Aufsatzbeurteilung – ein altes Problem neu untersucht. In: Didaktik Deutsch, 15. S. 46 - 63. Baltmannsweiler: Schneider.
- BLOOM, Benjamin (Hrsg.)/ENGELHART, Max/FURST, Edward/HILL, Walker/KRATHWOHL, David (²1973): Taxonomie von Lernzielen im kognitiven Bereich. Weinheim [u.a.]: Beltz.
- BÖRNER, Wolfgang (³2001): Schreiben im Fremdsprachenunterricht. In: JUNG, Udo (Hrsg.): Praktische Handreichung für Fremdsprachenlerer. Band 2. S. 287 - 293. Frankfurt am Main [u.a.]: Peter Lang.
- BREUGNOT, Jacqueline (2000): Fremdsprachen. In: REICH, Hans/HOLZBRECHER, Alfred/ROTH, Hans-Joachim (Hrsg.): Fachdidaktik interkulturell. Ein Handbuch. Opladen: Leske + Budrich.
- BROOKHART, Susan M. (2010): Wie sag ich's meinem Schüler? So kommt Ihr Feedback wirklich an. Weinheim [u.a.]: Beltz.
- BROSIUS, Hans-Bernd/KOSCHEL, Friederike/HAAS, Alexander (Hrsg.) (⁴2008): Methoden der empirischen Kommunikationsforschung. Eine Einführung. Wiesbaden: VS.
- BROWN, Douglas H./ABEYWICKRAMA, Priyanvada (²2010): Language assessment. Principles and classroom practices. White Plains, NY: Pearson Longman.

- BRUSCH, Wilfried (2009): Didaktik des Englischen. Ein Kerncurriculum in zwölf Vorlesungen. In: BÖRNER, Otfried/EDELHOFF, Christoph (Hrsg.): Unterrichts Perspektiven Fremdsprachen. Braunschweig: Schroedel.
- BUDDE, Monika/RIEGLER, Susanne/WIPRÄCHTIGER-GEPPERT, Maja (2012): Sprachdidaktik. Berlin: Akademie Verlag
- BUHREN, Claus G. (2015): Feedback - Definitionen und Differenzierungen. In: BUHREN, Claus G. (Hrsg.): Handbuch. Feedback in der Schule. S. 11 - 30. Weinheim [u.a.]: Beltz.
- BURWITZ-MELZER, Eva (2010): Sprachenportfolios. In: HALLET, Wolfgang/KÖNIGS, Frank (Hrsg.): Handbuch Fremdsprachendidaktik. S. 232 - 236. Seelze: Friedrich.
- BUSCHKÜHLE, Carl-Peter/DUNCKER, Ludwig/OSWALT, Vadim (2009): Bildung zwischen Standardisierung und Heterogenität: ein interdisziplinärer Diskurs.
- CASPARI, Daniela/GROTJAHN, Rüdiger/KLEPPIN, Karin (2010): Testaufgaben und Lernaufgaben. In: PORSCH, Raphaela/TESCH, Bernd/KÖLLER, Olaf (Hrsg.): Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I. S. 46 - 68. Münster [u.a.]: Waxmann.
- CASPARI, Daniela (2008): Zur Situation des Französischunterrichts. In: TESCH, Bernd/LEUPOLD, Eynar/KÖLLER, Olaf (Hrsg.): Bildungsstandards Französisch: konkret. Sekundarstufe I: Grundlagen, Aufgabenbeispiele und Unterrichts Anregungen. Berlin: Cornelsen Scriptor.
- CEBS (Hrsg.) (2013): Wegweiser: Die kompetenzorientierte mündliche Reife- und Diplomprüfung. Lebende Fremdsprachen. Wien: CEBS.
- CHAPELLE, Carol (2012): Conceptions of validity. In: FULCHER, Glenn/DAVIDSON, Fred (Hrsg.): The Routledge Handbook of Language Testing. S. 21 - 33. London [u.a.]: Routledge.
- CHUDASKE, Jana (2012): Sprache, Migration und schulfachliche Leistung. Einfluss sprachlicher Kompetenz auf Lese-, Rechtschreib- und Mathematikleistungen. Wiesbaden: Springer VS.
- CUQ, Jean-Pierre/GRUCA, Isabelle (2005): Cours de didactique du français langue étrangère et seconde. Grenoble: PUG.

- CUQ, Jean-Pierre (2003): Dictionnaire de didactique du français. Langue étrangère et seconde. Paris: CLE International.
- DECKE-CORNILL, Helene/KÜSTER, Lutz (³2015): Fremdsprachendidaktik. Tübingen: Narr.
- DÖRING, Nicola/BORTZ, Jürgen (⁵2016): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Heidelberg [u.a.]: Springer.
- DÖRING, Nicola/BORTZ, Jürgen (⁴2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg [u.a.]: Springer.
- DRESING, Thorsten/PEHL, Thorsten (⁵2013): Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende. Marburg: Dresing und Pehl.
- DRIESCHNER, Elmar (2011): Bildungsstandards und Lerndiagnostik. In: SACHER, Werner/WINTER, Felix (Hrsg.): Diagnose und Beurteilung von Schülerleistungen. S. 109 - 122. Baltmannsweiler: Schneider.
- DUDENREDAKTION (Hrsg.) (⁹2007a): Duden. Das Fremdwörterbuch. Band 5. Mannheim [u.a.]: Dudenverlag.
- DUDENREDAKTION (Hrsg.) (⁶2007b): Duden. Deutsches Universalwörterbuch. Mannheim [u.a.]: Dudenverlag.
- DUDENREDAKTION (Hrsg.) (²1985): Duden. Das Bedeutungswörterbuch. Band 10. Mannheim [u.a.]: Dudenverlag.
- EDMONDSON, Willis/HOUSE, Juliane (³2006): Einführung in die Sprachlehrforschung. Tübingen [u.a.]: Francke.
- EUROPARAT (Hrsg.) (2001): Der Gemeinsame Europäische Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin: Langenscheidt.
- EYNAR, Leupold (2010): Französisch lehren und lernen. Das Grundlagenbuch. Seelze: Kallmeyer.
- FÄCKE, Christiane (2010): Fachdidaktik Französisch. Eine Einführung. Tübingen: Narr.
- FELDER, Ekkehard (2003): Sprache als Medium und Gegenstand des Unterrichts. In: BREDEL, Ursula/GÜNTHER, Hartmut/KLOTZ, Peter/OSSNER, Jakob/SIEBERT-OTT, Gesa (Hrsg.): Didaktik der deutschen Sprache - ein Handbuch. Teilband 1. S. 42 - 51. Paderborn: Schöningh und Fink.

- FENGLER, Jörg (³2004): Feedback geben. Strategien und Übungen. Weinheim [u.a.]: Beltz.
- FLICK, Uwe (2009): Sozialforschung. Methoden und Anwendungen. Ein Überblick für die BA-Studiengänge. Reinbeck: Rowohlt.
- GASSNER, Otmar/MEWALD, Claudia/SIGOTT, Guenther (2008): Testing Writing Specifications for the E8-Standards Writing Tests. LTC Technical Report 4. Klagenfurt: LTC.
- GLÄSER, Jochen/LAUDEL, Grit (⁴2010): Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. Wiesbaden: VS.
- GLÜCK, Helmut (Hrsg.) (⁴2010): Metzler Lexikon: Sprache. Stuttgart [u.a.]: Metzler.
- GROTJAHN, Rüdiger/KLEPPIN, Karin (2015): Prüfen, Testen, Evaluieren. München: Klett-Langenscheidt.
- GROTJAHN, Rüdiger (2010): Sprachtests: Formen und Funktionen. In: HALLET, Wolfgang/KÖNIGS, Frank G. (Hrsg.): Handbuch Fremdsprachendidaktik. S. 211 - 215. Seelze: Friedrich.
- GROTJAHN, Rüdiger (2008): Tests und Testaufgaben: Merkmale und Gütekriterien. In: TESCH, Bernd/LEUPOLD, Eynar/KÖLLER, Olaf (Hrsg.): Bildungsstandards Französisch: konkret. Sekundarstufe I: Grundlagen, Aufgabenbeispiele und Unterrichts Anregungen. S. 149 - 186. Berlin: Cornelsen.
- GROTJAHN, Rüdiger/KLEPPIN, Karin (2008): Bewertung produktiver sprachlicher Leistungen. In: TESCH, Bernd/LEUPOLD, Eynar/KÖLLER, Olaf (Hrsg.): Bildungsstandards Französisch: konkret. Sekundarstufe I: Grundlagen, Aufgabenbeispiele und Unterrichts Anregungen. S. 187 - 204. Berlin: Cornelsen.
- HELLER, Kurt (Hrsg.) (²2001): Hochbegabung im Kindes- und Jugendalter. Göttingen [u.a.]: Hogrefe.
- HILGERS, Denis (2008): Performance Management. Leistungserfassung und Leistungssteuerung in Unternehmen und öffentlichen Verwaltungen. Wiesbaden: Gabler.
- HILLBRANDT, Christian/SINTZEN-KÖNIGSFELD, Wilhelm (2009): Schulentwicklung durch Standardisierung, Anmerkungen zur Entwicklung von Schule durch Bildungsstandards aus historischer und aktueller Sicht. In: HELSPER, Werner/HILLBRANDT, Christian/SCHWARZ, Thomas (Hrsg.): Schule und Bildung

- im Wandel. Anthologie historischer und aktueller Perspektiven. S. 15 - 35. Wiesbaden: Springer VS.
- HUGHES, Arthur (²2003): Testing for language teacher. Cambridge: Cambridge University Press.
- INGENKAMP, Karl-Heinz/LISSMANN, Urban (⁶2008): Lehrbuch der Pädagogischen Diagnostik. Weinheim [u.a.]: Beltz.
- INGENKAMP, Karl-Heinz (Hrsg.) (⁹1995): Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte. Weinheim [u.a.]: Beltz.
- ISAK, Gabriele (2010): Bildungsstandards als Chance für den Fremdsprachenunterricht. In: MUHR, Rudolf/BIFFL, Gudrun (Hrsg.): Sprache-Bildung-Bildungsstandards-Migration. Chancen und Risiken der Neuorientierung des österreichischen Bildungssystems. S. 81 - 90. Wien: Lang.
- JANG, Eunice Eunhee (2012): Diagnostic assessment in language classrooms. In: FULCHER, Glenn/DAVIDSON, Fred (Hrsg.): The Routledge Handbook of Language Testing. S. 120 - 133. London [u.a.]: Routledge.
- JÜRGENS, Eiko (1992): Leistung und Beurteilung in der Schule. Eine Einführung in Leistungs- und Bewertungsfragen aus pädagogischer Sicht. Sankt Augustin: Academia.
- KELLER, Stefan (2011): Beurteilungsraster und Kompetenzmodelle. In: SACHER, Werner/WINTER, Felix (Hrsg.): Diagnose und Beurteilung von Schülerleistungen. S. 143 - 160. Baltmannsweiler: Schneider.
- KLAFKI, Wolfgang (⁵1996): Neue Studien zur Bildungstheorie und Didaktik. Weinheim [u.a.]: Beltz.
- KLEBER, Eduard (1992): Diagnostik in pädagogischen Handlungsfeldern. Einführung in Bewertung, Beurteilung, Diagnose und Evaluation. Weinheim [u.a.]: Juventa.
- KÖNIGSLEHNER, Hans (1996): Leistung durch Motivation. In: Erziehung und Unterricht, 9/1996. S. 661 - 668. Wien: ÖBV.
- KRAUSE, Wolf-Dieter (2005): Pragmatische Linguistik und Fremdsprachenunterricht. In: ADAMZIK, Kirsten/KRAUSE, Wolf-Dieter (Hrsg.): Text-Arbeiten. Textsorten im fremd- und muttersprachlichen Unterricht an Schule und Hochschule. S. 205 - 237. Tübingen: Narr.

- KROMREY, Helmut (¹²2009): Empirische Sozialforschung. Stuttgart: Lucius & Lucius.
- LIENERT, Gustav/RAATZ, Ulrich (⁶1998): Testaufbau und Testanalyse. Weinheim [u.a.]: Beltz.
- LUCYSHYN, Josef (2007): Bildungsstandards in Österreich – Entwicklung und Implementierung: Pilotphase 2 (2004 - 2007). In: Erziehung und Unterricht. Heft 7-8/2007. S. 566 - 587. Wien: ÖBV.
- MAIER, Uwe (2015): Leistungsdiagnostik in Schule und Unterricht. Schülerleistungen messen, bewerten und fördern. Bad Heilbrunn: Klinkhardt.
- MAYRING, Philipp (¹¹2010): Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim [u.a.]: Beltz.
- NEUWEG, Georg Hans (⁵2014): Schulische Leistungsbeurteilung. Rechtliche Grundlagen und pädagogische Hilfestellungen für die Schulpraxis. Linz: Trauner.
- NOHL, Arnd-Michael (²2008): Interview und dokumentarische Methode. Anleitungen für die Forschungspraxis. Wiesbaden: VS.
- ÖSZ (Hrsg.). (2008): Leistungsfeststellung auf der Basis des Gemeinsamen europäischen Referenzrahmens für Sprachen. Graz: ÖSZ.
- OSCARSON, Mats (1997): Self-assessment of foreign and second language proficiency. In: CLAPHAM, Caroline/CORSON, David (Hrsg.): Encyclopedia of language and education. Vol 7: Language testing and assessment. S. 175 - 187. Dordrecht [u.a.]: Kluwer.
- PIEPHO, Hans-Eberhard (1974): Kommunikative Kompetenz als übergeordnetes Lernziel im Englischunterricht. Dornburg-Frickhofen: Frankonius.
- POSPESCHILL, Markus (2010): Testtheorie, Testkonstruktion, Testevaluation. München [u.a.]: Ernst Reinhardt.
- RAABE, Horst (⁴2003): Französisch. In: BAUSCH, Karl-Richard/CHRIST, Herbert/ KRUMM, Hans-Jürgen (Hrsg.): Handbuch Fremdsprachendidaktik. Tübingen [u.a.]: Francke.
- RAITH, Thomas (2011): Kompetenzen für aufgabenorientiertes Fremdsprachenunterrichten. Eine qualitative Untersuchung zur Ausbildung von Fremdsprachenlehrkräften. Tübingen: Narr.
- RENNERT, Peter/SCHMIEDEL, Herbert (Hrsg.): Physik. Mannheim [u.a.]: BI.

- REZAT, Sara (2014): Fachspezifische Leistungsbewertung und -beurteilung. In: GOER, Charis/KÖLLER, Katharina (Hrsg.): Fachdidaktik Deutsch. Grundzüge der Sprach- und Literaturdidaktik. S. 97 - 111. Paderborn: Fink.
- RHEINBERG, Falko (2002): Bezugsnormen und schulische Leistungsbeurteilung. In: WEINERT, Franz (Hrsg.): Leistungsmessungen in Schulen. S. 59 - 71. Weinheim [u.a.]: Beltz.
- RIVARA, René (1990): Le système de la comparaison. Sur la construction du sens dans les langues naturelles. Paris: Minuit.
- ROLOFF, Eckart (2014): Gestatten Erika Mustermann und Otto Normalverbraucher. Aufklärung über vermeintliche Durchschnittsmenschen, die im Alltag am Rande stehen. In: HÖMBERG, Walter/ROLOFF, Eckart (Hrsg.): Jahrbuch für Marginalistik III. S. 47 - 60. Münster: LIT.
- ROST, Jürgen (2004): Lehrbuch Testtheorie – Testkonstruktion. Bern [u.a.]: Hans Huber.
- SACHER, Werner (2011): Durchführung der Leistungsüberprüfung und Leistungsbeurteilung. In: SACHER, Werner/WINTER, Felix (Hrsg.): Diagnose und Beurteilung von Schülerleistungen. S. 27 - 48. Baltmannsweiler: Schneider.
- SALDERN, Matthias von (1999): Schulleistung in Diskussion. Baltmannsweiler: Schneider.
- SALDERN, Matthias von (2011): Schulleistung 2.0. Von der Note zum Kompetenzraster. Norderstedt: BoD.
- SALZMANN, Patrizia (2015): Lernen durch kollegiales Feedback. Die Sicht von Lehrpersonen und Schulleitungen in der Berufsbildung. In: BECKER, Rolf et al. (Hrsg.): Empirische Erziehungswissenschaft. Band 57. Münster [u.a.]: Waxmann.
- SCHNEIDER, Günther/NORTH, Brian (2000): Fremdsprachen können – was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit. Zürich [u.a.]: Rüegger.
- STERN, Thomas (2010): Förderliche Leistungsbewertung. Wien: Özeps.
- STOCKMANN, Reinhard/MEYER, Wolfgang (2010): Evaluation. Eine Einführung. Opladen [u.a.]: Barbara Budrich.
- TAGLIANTE, Christine (2005): L'évaluation et le Cadre européen commun. Paris: CLE.

- TREMEL, Alfred (²2006): Lernen. In: KRÜGER, Heinz-Hermann/GRUNERT, Cathleen (Hrsg.): Wörterbuch Erziehungswissenschaft. Opladen [u.a.]: Barbara Budrich.
- VOSS, Bernd (³2001): Sprache im Unterricht – Unterrichtssprache: Zur Bedeutung der Unterrichtssprache im Fremdsprachenunterricht. In: JUNG, Udo (Hrsg.): Praktische Handreichung für Fremdsprachenlehrer. Band 2. S. 105 - 112. Frankfurt am Main [u.a.]: Peter Lang.
- WALL, Diane (2012): Washback. In: FULCHER, Glenn/DAVIDSON, Fred (Hrsg.): The Routledge Handbook of Language Testing. S. 79 - 92. London [u.a.]: Routledge.
- WAUGH, Keith/GRONLUND, Norman (¹⁰2013): Assessment of Student Achievement. Boston [u.a.]: Pearson.
- WEEDEN, Paul/WINTER, Jan/BROADFOOT, Patricia (2002): Assessment. What's in it for schools? London [u.a.]: Routledge.
- WEINERT, Franz (²2002): Schulleistungen – Leistungen der Schule oder der Schüler? In: WEINERT, Franz (Hrsg.): Leistungsmessungen in Schulen. S. 73 - 86. Weinheim [u.a.]: Beltz.
- WEINERT, Franz (2001): Vergleichende Leistungsmessung in Schulen – Eine umstrittene Selbstverständlichkeit. In: WEINERT, Franz (Hrsg.): Leistungsmessung in Schulen. S. 17 - 31. Weinheim [u.a.]: Beltz.
- WINGENDER, Monika (2005): Wesen und Funktion der Graduierung in der Sprache: In: GÜNTHER, Hans/JACHNOW, Helmut (Hrsg.): Kognition, Sprache und phraseologische/parömiologische Graduierung. S. 42 - 57. Wiesbaden: Harrassowitz.
- WINTER, Felix (2011): Aufgaben und Perspektiven einer reformierten Leistungsbeurteilung. In: SACHER, Werner/WINTER, Felix (Hrsg.): Diagnose und Beurteilung von Schülerleistungen. S. 197 - 216. Baltmannsweiler: Schneider.
- WITZEL, Andreas/REITER, Herwig (2012): The Problem-centred Interview. Principles and Practice. Los Angeles [u.a.]: SAGE.
- WITZEL, Andreas (1982): Verfahren der qualitativen Sozialforschung – Überblick und Alternative. Frankfurt [u.a.]: Campus.
- WOOLFOLK, Anita (¹⁰2008): Pädagogische Psychologie. München [u.a.]: Pearson.
- WOTTAWA, Heinrich/THIERAU, Heike (³2003): Lehrbuch Evaluation. Bern [u.a.]: Hans Huber.

VI.2 Internetquellen

Online-Journale

BARNES, Melissa (2016): The Washback of the TOEFL iBT in Vietnam. In: Australian Journal of Teacher Education. Vol 41, 7. S. 157 – 174.

<http://ro.ecu.edu.au/ajte/vol41/iss7/10/> (20.12.2016)

KARA, Filiz/ÇELIKLER, Dilek (2015): Development of achievement test. Validity and reliability study for achievement test on matter changing. In: Journal of Education and Practice. Vol 6, 24. S. 21 – 26.

<http://www.iiste.org/Journals/index.php/JEP/article/view/25254/25854>

(20.12.2016)

Internetseiten

BUNDESINSTITUT FÜR BILDUNGSFORSCHUNG, INNOVATION UND ENTWICKLUNG (BIFIE):
Bericht zur Entwicklung der standardisierten kompetenzorientierten schriftlichen Reife- und Diplomprüfung (2012):

https://www.parlament.gv.at/PAKT/VHG/XXIV/III/III_00422/imfname_307187.pdf

(20.01.2017)

BUNDESINSTITUT FÜR BILDUNGSFORSCHUNG, INNOVATION UND ENTWICKLUNG (BIFIE):
Beurteilungsraster für Fremdsprachen auf Sprachniveau B1 mit Begleittext (2014):

<https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/>

(20.01.2017)

BUNDESINSTITUT FÜR BILDUNGSFORSCHUNG, INNOVATION UND ENTWICKLUNG (BIFIE):
Bildungsstandards (2011 – 2017):

<https://www.bifie.at/bildungsstandards> (20.01.2017)

BUNDESINSTITUT FÜR BILDUNGSFORSCHUNG, INNOVATION UND ENTWICKLUNG (BIFIE):
Textsortenkatalog (2013):

<https://www.bifie.at/node/1462> (10.01.2017)

BUNDESMINISTERIUM FÜR BILDUNG: standardisierte Reife- und Diplomprüfung (2016):

https://www.bmb.gv.at/schulen/unterricht/ba/reifepruefung_bhs.html (20.01.2017)

BUNDESMINISTERIUM FÜR BILDUNG: überfachliche Kompetenzen (2016):

<https://www.bmb.gv.at/schulen/unterricht/uek/sozial.html> (10.01.2017)

CAMBRIDGE ONLINE WÖRTERBUCH:

<http://dictionary.cambridge.org/dictionary/english/assessment> (10.01.2017)

<http://dictionary.cambridge.org/dictionary/english/measurement> (10.01.2017)

DUDEN ONLINE:

<http://www.duden.de/suchen/dudenonline/treffend> (20.01.2017)

<http://www.duden.de/rechtschreibung/Wischiwaschi> (10.03.2017)

<http://www.duden.de/rechtschreibung/schwammig> (10.03.2017)

LEHRPLAN DER HAK (2014):

https://www.hak.cc/files/syllabus/Lehrplan_HAK_2014.pdf (10.01.2017)

LEISTUNGSBEURTEILUNGSVERORDNUNG (LBVO):

<https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009375> (10.01.2017)

SCHULUNTERRICHTSGESETZ (SCHUG):

<https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009600> (10.01.2017)

VII. Abbildungsverzeichnis

Abbildung 1: Das Assessment und seine Teilbereiche	5
(BROWN/ABEYWICKRAMA ² 2010: 6)	
Abbildung 2: Das Leistungsmodell nach HELLER und HANY: Zusammensetzung von Leistungsexzellenz	22
(SALDERN 2011: 39)	
Abbildung 3: Von der Kompetenz zur bewerteten Leistungsperformanz	27
(SALDERN 2011: 38)	

VIII. ANNEX

VIII.1 Schriftliche Performanz

Standardisierte Aufgabenstellung des BIFIE auf Niveau B1³⁰



Vous avez fait un voyage en France avec l'Agence *Soleil-Voyages*. Vous n'avez pas été content(e) de votre séjour. Vous décidez d'écrire un courriel de réclamation à l'agence pour exprimer votre colère.

Dans votre courriel, vous :

- racontez quel voyage vous avez fait
- expliquez pourquoi vous n'êtes pas content(e)
- précisez ce que vous attendez de l'agence *Soleil-Voyages*

Écrivez environ 200 mots.

³⁰ https://www.srdp.at/fileadmin/user_upload/downloads/Matura_2013-14/09_LFS_Fra/KL14_PT1_HAK_FRA_SR_B1_AU.pdf (20.01.2017)

SchülerInnenperformanz³¹

Bonjour Messieurs et Mesdames !

La semaine dernière, j'ai fait un voyage en France avec votre l'Agence Soleil-Voyages. Malheureusement je ne suis pas content avec le tour j'ai participé la semaine dernière, elle s'appelle « La magnifique France ».

Pourquoi, vous demandez ?

Premièrement le tour en bus était horrible : trois heures sans climatisation et sans une récréation. De plus, les trois heures avec des autres touristes, qui chanté très fort tout le voyage. Bien sûr, ils ont bu beaucoup de vin.

Deuxièmes, quand je suis arrivé devant l'hôtel j'étais choqué car l'hôtel était très sale et très âgé. Les chambres était très petit et la vue sur la mère, que vous décrivez sur l'internet ne pas exister.

Troisièmes, vous décriis la « nuit française » comme une nuit très intéressante avec des spécialités françaises, comme des crêpes, des tartes, et des spécialités méditerranées comme Ratatouille et les salâtes avec poisson et des olives. Vraiment dit, cette « nuit française » était une table avec des aliments de commodité et deux boutais avec vin le moins chère.

J'attende de votre l'Agence Soleil-Voyages de me retourner mon argent immédiatement où de m'offre des autres voyages

Je vous demande d'accréditer mes salutations.

Un client en colère

Jean Dupont

mots: 205

³¹ Diese authentische SchülerInnenperformanz wurde anonymisiert und in digitaler Form zur Beurteilung den Lehrpersonen vorgelegt, um einen etwaigen Halo-Effekt zu verhindern. (siehe Kapitel 6.2.1)

Korrekturblatt

Korrekturblatt			
Erfüllung der Aufgabenstellung	Aufbau und Layout	Spektrum sprachlicher Mittel	Sprachrichtigkeit
1)	1)	1)	1)
2)	2)	2)	2)
3)	3)	3)	3)
4)	4)	4)	4)
5)	5)	5)	
6)	6)		
7)			
8)			
Beurteilung :	Beurteilung :	Beurteilung :	Beurteilung :
Beurteilung insgesamt: /40		Note:	

Beurteilungsraster des BIFIE auf Sprachniveau B1³²

Beurteilungsraster B1		Beurteilungsraster B1	
B1 Erfüllung der Aufgabenstellung*	B1 Aufbau und Layout	B1 Spektrum sprachlicher Mittel	B1 Sprachrichtigkeit
<p>10</p> <p>(1) Hält die vorgegebene Textsorte durchgehend ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile treffend (3) Führt alle inhaltlichen Punkte an und behandelt sie so ausführlich wie für die Aufgabenstellung möglich (4) Führt veranschaulichende Details und Beispiele für alle inhaltlichen Punkte an (5) Erklärt Sachverhalte sehr erfolgreich ASD* (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem klar und deutlich ASD (5b) Erklärt Gründe für Handlungen sehr gut ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) ein</p>	<p>10</p> <p>(1) Sehr klarer Gesamtaufbau (2) Präsentiert die inhaltlichen Punkte sehr klar und systematisch (Kohärenz) (3) Gliedert den Text sehr gut (4) Hält sich durchgehend an das textspezifische Layout (5) Durchgehend guter Einsatz von verschiedenen textgrammatischen Mitteln (Kohäsion)</p>	<p>10</p> <p>(1) Hat ein breites Spektrum an lexikalischen und strukturellen Mitteln (2) Verwendet selten Wiederholungen, die auf begrenzte Ausdrucksmöglichkeiten hinweisen (3) Verwendet meist treffende Formulierungen; drückt sich größtenteils präzise aus (4) Hat keine Formulierenschwierigkeiten auf Grund eines mangelnden Wortschatzes (5) Schreibt durchgehend in einer der Textsorte angemessenen Stil</p>	<p>10</p> <p>(1) Beherrscht den Grundwortschatz sowie häufige Strukturen und Wendungen sehr gut (2) Macht nur solche sprachlichen Fehler, welche die Kommunikation nicht beeinträchtigen (3) Rechtschreibung und Zeichensetzung sind so korrekt, dass sie das Verständnis nicht beeinträchtigen (4) Fast keine Interferenzen aus anderen Sprachen</p>
<p>9</p> <p>(1) Hält die vorgegebene Textsorte fast durchgehend ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile gut (3) Führt alle inhaltlichen Punkte an und behandelt sie ziemlich ausführlich (4) Führt veranschaulichende Details und Beispiele für fast alle inhaltlichen Punkte an (5) Erklärt Sachverhalte erfolgreich ASD (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem gut ASD (5b) Erklärt Gründe für Handlungen gut ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) ein</p>	<p>9</p> <p>(1) Klarer Gesamtaufbau (2) Präsentiert die inhaltlichen Punkte klar und systematisch (Kohärenz) (3) Gliedert den Text durchgehend in passende Absätze (4) Hält sich fast durchgehend an das textspezifische Layout (5) Fast durchgehend guter Einsatz von textgrammatischen Mitteln (Kohäsion)</p>	<p>9</p> <p>(1) Hat ein ausreichend breites Spektrum an lexikalischen und strukturellen Mitteln, um sich mit hinreichender Genauigkeit ausdrücken zu können (2) Verwendet nur gelegentlich Wiederholungen, die auf begrenzte sprachliche Ausdrucksmöglichkeiten hinweisen (3) Verwendet einige treffende Formulierungen (4) Hat kaum Formulierenschwierigkeiten auf Grund eines mangelnden Wortschatzes (5) Schreibt fast durchgehend in einem der Textsorte angemessenen Stil</p>	<p>9</p> <p>(1) Beherrscht den Grundwortschatz sowie häufige Strukturen und Wendungen gut (2) Macht nur solche sprachlichen Fehler, welche die Kommunikation kaum beeinträchtigen (3) Rechtschreibung und Zeichensetzung sind so korrekt, dass sie das Verständnis kaum beeinträchtigen (4) Nur gelegentlich Interferenzen aus anderen Sprachen</p>
<p>8</p> <p>(1) Hält die vorgegebene Textsorte überwiegend ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile annähernd annähernd (3) Führt alle inhaltlichen Punkte an, aber behandelt sie wenig ausführlich ODER führt nur zwei von drei inhaltlichen Punkten an, aber behandelt sie ziemlich ausführlich (4) Führt einige veranschaulichende Details und Beispiele an (5) Erklärt Sachverhalte annähernd erfolgreich ASD (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem annähernd erfolgreich ASD (5b) Erklärt Gründe für Handlungen annähernd erfolgreich ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) ein</p>	<p>8</p> <p>(1) Engmaßiger Gesamtaufbau (2) Präsentiert die inhaltlichen Punkte annähernd systematisch (Kohärenz) (3) Gliedert den Text größtenteils in passende Absätze (4) Hält sich größtenteils an das textspezifische Layout (5) Eingeschränkter, aber vorwiegend passender Einsatz von elementaren textgrammatischen Mitteln (Kohäsion)</p>	<p>8</p> <p>(1) Hat ein genügend breites Spektrum an lexikalischen und strukturellen Mitteln, um die Aufgabe erfüllen zu können (2) Verwendet einige Wiederholungen und/oder Umschreibungen auf Grund eingeschränkter sprachlicher Ausdrucksmöglichkeiten (3) Hat manchmal Formulierenschwierigkeiten auf Grund eines mangelnden Wortschatzes (4) Schreibt überwiegend in einem der Textsorte angemessenen Stil (5) Erkennt Satzbauelemente / einzelne Wörter aus der Aufgabenstellung</p>	<p>8</p> <p>(1) Beherrscht den Grundwortschatz sowie häufige Strukturen und Wendungen ausreichend gut (2) Macht nur beim Formulieren komplexerer Sachverhalte sprachliche Fehler, welche die Kommunikation beeinträchtigen (3) Rechtschreibung und Zeichensetzung sind so korrekt, dass sie das Verständnis nur teilweise beeinträchtigen (4) Teilweise Interferenzen aus anderen Sprachen</p>
<p>7</p> <p>(1) Hält die vorgegebene Textsorte annähernd ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile annähernd (3) Führt nur zwei von drei inhaltlichen Punkten an und behandelt sie wenig ausführlich (4) Führt keine veranschaulichenden Details an (5) Erklärt Sachverhalte nicht erfolgreich ASD (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem nicht erfolgreich ASD (5b) Erklärt Gründe für Handlungen nicht erfolgreich ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) nicht ein</p>	<p>7</p> <p>(1) Wenig angemessener Gesamtaufbau (2) Präsentiert die inhaltlichen Punkte wenig systematisch (Kohärenz) (3) Gliedert den Text nur mangelhaft (4) Hält sich nicht an das textspezifische Layout (5) Sehr eingeschränkter oder zum Teil unpassender Einsatz von textgrammatischen Mitteln (Kohäsion) (6) Häufige Gedankenstränge</p>	<p>7</p> <p>(1) Hat nicht genügend lexikalische und strukturelle Mittel, um die Aufgabe erfüllen zu können (2) Verwendet viele Wiederholungen auf Grund beschränkter sprachlicher Ausdrucksmöglichkeiten (3) Hat häufig Formulierenschwierigkeiten auf Grund eines mangelnden Wortschatzes (4) Schreibt nur ansatzweise in einem der Textsorte angemessenen Stil (5) Erkennt ganze Sätze / Wendungen aus der Aufgabenstellung</p>	<p>7</p> <p>(1) Beherrscht den Grundwortschatz sowie häufige Strukturen und Wendungen nur mangelhaft (2) Macht schon beim Formulieren einfacher Sachverhalte sprachliche Fehler, welche die Kommunikation beeinträchtigen (3) Rechtschreibung und Zeichensetzung sind so mangelhaft, dass sie das Verständnis häufig beeinträchtigen (4) Häufig Interferenzen aus anderen Sprachen</p>
<p>6</p> <p>(1) Hält die vorgegebene Textsorte nicht ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile nicht annähernd / nicht annähernd / unvollständig (3) Führt nur zwei von drei inhaltlichen Punkten an und behandelt sie wenig ausführlich (4) Führt nicht genügend veranschaulichende Details an / Führt manchmal irrelevante inhaltliche Punkte oder Details an (5) Erklärt Sachverhalte nicht erfolgreich ASD (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem nicht erfolgreich ASD (5b) Erklärt Gründe für Handlungen nicht erfolgreich ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) nicht ein</p>	<p>6</p> <p>(1) Strukturiert den Text kaum (2) Präsentiert die inhaltlichen Punkte unsystematisch (Kohärenz) (3) Gliedert den Text nicht zufriedenstellend (4) Hält sich nicht an das textspezifische Layout (5) Unpassender oder fehlender Gebrauch von textgrammatischen Mitteln (Kohäsion)</p>	<p>6</p> <p>(1) Hat kaum genügend lexikalische und strukturelle Mittel, um sich annähernd ausdrücken zu können (2) Übernimmt fast vollständig den Text aus der Aufgabenstellung</p>	<p>6</p> <p>(1) Text ist auf Grund der sprachlichen Fehler überhaupt nicht verständlich (2) Produziert nicht genügend Sprache für eine Beurteilung</p>
<p>5</p> <p>(1) Hält die vorgegebene Textsorte nicht ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile nicht annähernd / nicht annähernd / unvollständig (3) Führt nur zwei von drei inhaltlichen Punkten an und behandelt sie wenig ausführlich (4) Führt keine veranschaulichenden Details an (5) Erklärt Sachverhalte nicht erfolgreich ASD (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem nicht erfolgreich ASD (5b) Erklärt Gründe für Handlungen nicht erfolgreich ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) nicht ein</p>	<p>5</p> <p>(1) Keine Struktur erkennbar (2) Produziert nicht genügend Sprache für eine Beurteilung</p>	<p>5</p> <p>(1) Text ist auf Grund der sprachlichen Fehler überhaupt nicht verständlich (2) Produziert nicht genügend Sprache für eine Beurteilung</p>	<p>5</p> <p>(1) Text ist auf Grund der sprachlichen Fehler überhaupt nicht verständlich (2) Produziert nicht genügend Sprache für eine Beurteilung</p>
<p>4</p> <p>(1) Hält die vorgegebene Textsorte nicht ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile nicht annähernd / nicht annähernd / unvollständig (3) Führt nur zwei von drei inhaltlichen Punkten an und behandelt sie wenig ausführlich (4) Führt keine veranschaulichenden Details an (5) Erklärt Sachverhalte nicht erfolgreich ASD (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem nicht erfolgreich ASD (5b) Erklärt Gründe für Handlungen nicht erfolgreich ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) nicht ein</p>	<p>4</p> <p>(1) Strukturiert den Text kaum (2) Präsentiert die inhaltlichen Punkte unsystematisch (Kohärenz) (3) Gliedert den Text nicht zufriedenstellend (4) Hält sich nicht an das textspezifische Layout (5) Unpassender oder fehlender Gebrauch von textgrammatischen Mitteln (Kohäsion)</p>	<p>4</p> <p>(1) Text ist auf Grund der sprachlichen Fehler kaum verständlich (2) Macht systematisch elementare Fehler bei der Verwendung des Grundwortschatzes sowie häufiger Strukturen und Wendungen (3) Rechtschreibung und Zeichensetzung sind so mangelhaft, dass sie das Verständnis häufig beeinträchtigen (4) Systematische Interferenzen aus anderen Sprachen</p>	<p>4</p> <p>(1) Text ist auf Grund der sprachlichen Fehler kaum verständlich (2) Macht systematisch elementare Fehler bei der Verwendung des Grundwortschatzes sowie häufiger Strukturen und Wendungen (3) Rechtschreibung und Zeichensetzung sind so mangelhaft, dass sie das Verständnis häufig beeinträchtigen (4) Systematische Interferenzen aus anderen Sprachen</p>
<p>3</p> <p>(1) Hält die vorgegebene Textsorte nicht ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile nicht annähernd / nicht annähernd / unvollständig (3) Führt nur zwei von drei inhaltlichen Punkten an und behandelt sie wenig ausführlich (4) Führt keine veranschaulichenden Details an (5) Erklärt Sachverhalte nicht erfolgreich ASD (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem nicht erfolgreich ASD (5b) Erklärt Gründe für Handlungen nicht erfolgreich ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) nicht ein</p>	<p>3</p> <p>(1) Keine Struktur erkennbar (2) Produziert nicht genügend Sprache für eine Beurteilung</p>	<p>3</p> <p>(1) Text ist auf Grund der sprachlichen Fehler kaum verständlich (2) Macht systematisch elementare Fehler bei der Verwendung des Grundwortschatzes sowie häufiger Strukturen und Wendungen (3) Rechtschreibung und Zeichensetzung sind so mangelhaft, dass sie das Verständnis häufig beeinträchtigen (4) Systematische Interferenzen aus anderen Sprachen</p>	<p>3</p> <p>(1) Text ist auf Grund der sprachlichen Fehler kaum verständlich (2) Macht systematisch elementare Fehler bei der Verwendung des Grundwortschatzes sowie häufiger Strukturen und Wendungen (3) Rechtschreibung und Zeichensetzung sind so mangelhaft, dass sie das Verständnis häufig beeinträchtigen (4) Systematische Interferenzen aus anderen Sprachen</p>
<p>2</p> <p>(1) Hält die vorgegebene Textsorte nicht ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile nicht annähernd / nicht annähernd / unvollständig (3) Führt nur zwei von drei inhaltlichen Punkten an und behandelt sie wenig ausführlich (4) Führt keine veranschaulichenden Details an (5) Erklärt Sachverhalte nicht erfolgreich ASD (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem nicht erfolgreich ASD (5b) Erklärt Gründe für Handlungen nicht erfolgreich ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) nicht ein</p>	<p>2</p> <p>(1) Keine Struktur erkennbar (2) Produziert nicht genügend Sprache für eine Beurteilung</p>	<p>2</p> <p>(1) Text ist auf Grund der sprachlichen Fehler überhaupt nicht verständlich (2) Produziert nicht genügend Sprache für eine Beurteilung</p>	<p>2</p> <p>(1) Text ist auf Grund der sprachlichen Fehler überhaupt nicht verständlich (2) Produziert nicht genügend Sprache für eine Beurteilung</p>
<p>1</p> <p>(1) Hält die vorgegebene Textsorte nicht ein (2) Formuliert Titel / Betreff / Abschnittsüberschriften / Anrede / Grußzeile nicht annähernd / nicht annähernd / unvollständig (3) Führt nur zwei von drei inhaltlichen Punkten an und behandelt sie wenig ausführlich (4) Führt keine veranschaulichenden Details an (5) Erklärt Sachverhalte nicht erfolgreich ASD (5a) Kommuniziert die eigene Meinung zu einer Situation oder einem Problem nicht erfolgreich ASD (5b) Erklärt Gründe für Handlungen nicht erfolgreich ASD (6) Hält die vorgegebene Wortanzahl (+/- 10 %) nicht ein</p>	<p>1</p> <p>(1) Keine Struktur erkennbar (2) Produziert nicht genügend Sprache für eine Beurteilung</p>	<p>1</p> <p>(1) Text ist auf Grund der sprachlichen Fehler überhaupt nicht verständlich (2) Produziert nicht genügend Sprache für eine Beurteilung</p>	<p>1</p> <p>(1) Text ist auf Grund der sprachlichen Fehler überhaupt nicht verständlich (2) Produziert nicht genügend Sprache für eine Beurteilung</p>
<p>0</p> <p>(1) Verfehlt die Aufgabenstellung* (2) Produziert nicht genügend Sprache für eine Beurteilung (3) Schreibt unleserlich, die Kommunikation ist daher nicht ergebnisbringend</p>	<p>0</p> <p>(1) Keine Struktur erkennbar (2) Produziert nicht genügend Sprache für eine Beurteilung</p>	<p>0</p> <p>(1) Text ist auf Grund der sprachlichen Fehler überhaupt nicht verständlich (2) Produziert nicht genügend Sprache für eine Beurteilung</p>	<p>0</p> <p>(1) Text ist auf Grund der sprachlichen Fehler überhaupt nicht verständlich (2) Produziert nicht genügend Sprache für eine Beurteilung</p>

³² <https://www.srdp.at/downloads/dl/beurteilungsraster-b1-und-begleittext/> (20.01.2017)

VIII.2 Intervieworganisation

Ablauf der Interviews

Einführung

- Begrüßung und Bedankung für die Teilnahmebereitschaft
- Vorstellung des Interviewers und des Forschungsvorhabens
- Schaffung einer angenehmen Gesprächsatmosphäre
- Information über Ablauf der Befragung und Tonbandaufzeichnung
- Einholung der Einverständniserklärung
- mögliche Fragen seitens der interviewten Person

Problemzentriertes Leitfadeninterview

- Aufnahme starten und Interview beginnen
- an den Leitfragen orientieren
- bei wichtigen Punkten nachfragen
- aktives Zuhören, Offenheit ermöglichen, bei Unklarheiten nachfragen, paraphrasieren

Abschluss

- Kontrolle, ob alle wichtigen Punkte abgehandelt wurden
- für die Kooperation bedanken
- Aufnahme beenden und Interesse an Interviewergebnissen abfragen

Leitfaden für das problemzentrierte Interview

Interviewprotokoll	Nummer:
Ort	
Datum	
Zeit	

Teil 1: Allgemeine demografische Daten

- Geschlecht: m w
- Lehrperson an einer: AHS BHS
- Unterrichtserfahrung: 0 – 5 Jahre 6 – 10 Jahre
 11 – 15 Jahre 16 – 20 Jahre
 21 – 25 Jahre mehr als 25 Jahre
- besuchte Fortbildungen zur standardisierten Leistungsbeurteilung mit Beurteilungsraster:
 keine eine zwei drei oder mehr

Teil 2: Bewertung schriftlicher Performanzen

Wie nehmen Sie die Leistungsbeurteilung von schriftlichen SchülerInnentexten wahr? Stellt es vergleichsweise eine Herausforderung dar oder wie stehen sie der Bewertung von Performanzen prinzipiell gegenüber?

Worin sehen Sie mögliche Begründungen?

Teil 3: Beurteilungsraster

Inwiefern betrachten Sie Bewertungshilfen allgemein als nützlich?

Wie stehen Sie dem Beurteilungsraster des BIFIE auf Sprachniveau B1 konkret gegenüber?

Wie positiv oder negativ stehen Sie dem Beurteilungsraster gegenüber?

Wie hilfreich empfinden Sie diesen Beurteilungsraster?

Zunächst zu den Kategorien: Bewerten Sie alle vier Kategorien als gleich bedeutend? Oder sehen Sie gewisse als wichtiger und andere als weniger wichtig an?

Wie klar abgegrenzt scheinen Ihnen die Formulierungen der einzelnen Ausprägungsgrade der unterschiedlichen Niveaustufen?

Eventuell folgende Beispiele anführen und besprechen:

Erfüllung der Aufgabenstellung: Deskriptor 2	
Band 10	„Formuliert Titel/Betreff/Abschnittsüberschriften/Anrede/Grußzeile <u>treffend</u> “
Band 8	„Formuliert Titel/Betreff/Abschnittsüberschriften/Anrede/Grußzeile <u>gut</u> “
Band 6	„Formuliert Titel/Betreff/Abschnittsüberschriften/Anrede/Grußzeile <u>angemessen</u> “
Aufbau und Layout: Deskriptor 2	
Band 10	„Präsentiert die inhaltlichen Punkte <u>sehr klar</u> und <u>systematisch</u> “
Band 8	„Präsentiert die inhaltlichen Punkte <u>klar</u> und <u>systematisch</u> “
Band 6	„Präsentiert die inhaltlichen Punkte <u>einigermaßen</u> und <u>systematisch</u> “
Spektrum sprachlicher Mittel: Deskriptor 1	
Band 10	„Hat ein breites Spektrum an lexikalischen und strukturellen Mitteln“
Band 8	„Hat ein <u>ausreichend</u> breites Spektrum an lexikalischen und strukturellen Mitteln, um sich mit hinreichender <u>Genauigkeit</u> ausdrücken zu können“
Band 6	„Hat ein <u>genügend</u> breites Spektrum an lexikalischen und strukturellen Mitteln, um die Aufgabe erfüllen zu können.“
Erfüllung der Aufgabenstellung: Deskriptor 8	
Band 10	„Hält die vorgegebene Wortanzahl ein“
Band 8	„Hält die vorgegebene Wortanzahl ein“
Band 6	„Hält die vorgegebene Wortanzahl ein“

Teil 4: subjektiver Interpretationsraum

Sind Sie der Ansicht, dass einige Formulierungen einen subjektiven Interpretationsraum offenlassen?

Wie schätzen Sie die Resultate einer Leistungsbeurteilung einer schriftlichen Performanz durch mehrere Lehrpersonen ein? Erwarten Sie selbige oder ähnliche Ergebnisse?

Teil 5: Bezugsnormen

Haben Sie für sich persönlich eine Norm für einzelne Deskriptoren definiert?

- Orientierung an den Punkten
- Bezug zu Notensystem
- Einbindung der individuellen Bezugsnorm?
- Einbindung der sozialen Bezugsnorm?
- Referenzrahmeneffekt

Inwieweit kommt eine rein sachliche Bezugsnorm hinsichtlich der Beschreibungen zur Anwendung und inwieweit spielen auch individuelle sowie soziale Bezugssysteme eine Rolle?

Meinen Sie, dass Sie eine intra-personelle Reliabilität, also eine immer gleiche Auffassung und Interpretation von Kriterien für Sie selbst haben, bei der Beurteilung von SchülerInnenleistungen, sodass sie nicht situationsabhängig variieren?

Teil 6: Anwendung des Rasters

Ich hätte dann auch noch Fragen zur Anwendung des Raster. Wie Sie ihn anwenden. Wenden Sie den Raster prinzipiell immer beim Korrigieren von Schülertexten an oder nur in bestimmten (offiziellen) Situationen zum Beispiel?

Wie häufig korrigieren Sie Schülertexte hinsichtlich der vier Kriterien des Beurteilungsrasters ohne diesen gänzlich heranzuziehen?

In welchen Situationen korrigieren Sie ausschließlich hinsichtlich der kriterialen Bezugsnorm, ohne die individuelle und/oder die soziale Bezugsnorm zu berücksichtigen, unter Verwendung des Beurteilungsrasters?

Für wie wichtig befinden Sie Fortbildungsmöglichkeiten hinsichtlich des Umganges mit dem Beurteilungsraster, also Rater-Schulungen?

Wie fühlen Sie sich im Umgang mit dem Raster? Fühlen Sie sich ausreichend professionalisiert oder wünschen Sie sich mehr Fortbildungsmöglichkeiten und Schulungen. Welche Bedeutung messen Sie dem bei?

Teil 7: Performanz

Bezug zur korrigierten Performanz ... eventuell:

- Allgemeine Stellungnahme zum Text.
- Welche Bereiche als sehr gut/gut/weniger gut/schlecht bewertet?
- Worin liegen die Begründungen?
- ...

Ausgefülltes Korrekturblatt → *Bezug zur Leistungsbeurteilung der Performanz!*

Teil 8: Handlungsorientierung

Welchen Stellenwert hat Ihrer Ansicht nach die Handlungsorientierung im Fremdsprachenunterricht?

- Im Verhältnis zur Sprachrichtigkeit
- Bezug zur Performanz

Weitere **Zwischenfragen** je nach Gesprächsverlauf ad hoc stellen.

Alle Teile des Interviews anvisieren und ansprechen. Bei gewissen Sachverhalten gezielt nachfragen und Informationen aus dem Theorieteil präsentieren sowie die Einstellungen und Ansichten der Lehrpersonen diesbezüglich befragen.

Vielen Dank für Ihre Kooperation und Teilnahme!

Ich freue mich, Ihre Erfahrungen als einen wertvollen Beitrag für meine Forschung und wissenschaftliche Zwecke heranziehen zu können.

Einwilligungserklärung

zur Erhebung und Auswertung personenbezogener Interviewdaten

Hiermit bestätige ich die Auskunft und Information über das Forschungsinteresse und das damit einhergehende Vorhaben mit den im Interview erhobenen Daten, die mittels Aufnahmegerät aufgezeichnet und in weiterer Folge transkribiert werden dürfen. Personenbezogene Daten werden größtmöglich verschleiert, sodass die Interviewdaten anonym im Sinne einer wissenschaftlichen Auseinandersetzung und Bearbeitung präsentiert werden dürfen.

Unter den oben dargelegten Sachverhalten bin ich einverstanden ein Interview zu führen.

Vorname, Nachname

Ort, Datum

Unterschrift

VIII.3 Transkriptionen

Interviewprotokoll	Interview 1
Ort	Wien
Datum	25.02.2017
Zeit	10:20 – 10:50

Interviewprotokoll	Interview 2
Ort	Wien
Datum	06.03.2017
Zeit	11:30 – 11:55

Interviewprotokoll	Interview 3
Ort	Wien
Datum	08.03.2017
Zeit	08:10 – 08:45

Anbei sind die Transkriptionen der drei im Zuge der Forschung durchgeführten Interviews aufzufinden, welche nach den Regeln des einfachen Transkriptionssystems nach DRESING/PEHL (⁵2013) verschriftlicht wurde.

Interview 1

I Zunächst würde ich als Einstiegsfrage gerne wissen, wie Sie die Leistungsbeurteilung von schriftlichen Performanzen empfinden. Vor allem interessant ist, ob Sie es vergleichsweise zu anderen Kompetenzen als Herausforderung ansehen, diese Leistungen zu bewerten oder nicht.

LP1 Also ich empfinde das Korrigieren von Schülertexten als eine ganz normale, angenehme Sache im Lehrberuf, vor allem mit dem Raster. Er ist sehr transparent und ich finde gut, dass es ihn gibt.

I Also finden Sie Beurteilungsinstrumente prinzipiell hilfreich, um die Bewertung eben anzuleiten?

LP1 Ja! Ich finde es gut, dass es eine Basis gibt, auf die man sich beziehen kann und dass diese Basis für (..) für eigene, sag ich mal, subjektivere Empfindungen Spielraum lässt, aber (...) ich finde es gut, dass es immer eine objektive Basis gibt.

I Mhm, ok. Und wenn wir nun konkret diesen Beurteilungsraster heranziehen, wie empfinden Sie diesen persönlich? Wie ist Ihre Einstellung allgemein dazu?

LP1 Ah (.) sehr gut!

I Ok. Ich würde gerne auf die Kategorien und die Deskriptoren eingehen, da der Raster ja in vier Kategorien unterteilt ist und in diesen Kategorien in verschiedene Niveaustufen aufgefächert ist.

LP1 Ja.

I Zu den Kategorien zuerst: Wie empfinden Sie die Aufteilung der Kategorien? Finden Sie gewisse Kategorien wichtiger als andere oder wie sehen Sie das?

LP1 Ah (..) ja. Also, (..) die, ah (.) die Kategorien „Spektrum sprachlicher Mittel“ und „Sprachrichtigkeit“ würde ich nun mal als die wichtigsten bezeichnen.

I Ok, mit welcher Begründung?

LP1 Ahm (..) weil, ja weil das den Kern einer Sprachkompetenz am genauesten trifft, wobei die anderen auch wichtig sind. Also sie sind nicht wegzudenken aber sie sind nun mal nicht so wichtig wie die beiden anderen, die ich gerade genannt habe. Also ja, wenn „Erfüllung der Aufgabenstellung“ nicht passt, dann kann das sprachlich der beste Text der Welt sein (..) deswegen (...)

I (...) deswegen kann diese Kategorie als Voraussetzung an einen Text angesehen werden? Meinen Sie das?

LP1 Ja genau. Und „Aufbau und Layout“, die ah (...) die würde ich auch zu diesem Bereich der „Erfüllung der Aufgabenstellung“ dazugeben. (...) Ja, ich finde, also das sind so diese zwei Pärchen. Das eine „Erfüllung der Aufgabenstellung“ und „Aufbau und Layout“, das ist so quasi das Drumherum, das passen muss und der harte Kern sind eben „Spektrum sprachlicher Mittel“ und die „Sprachrichtigkeit“, der dann im Zweifelsfall, naja, wichtiger ist, sage ich mal

I Ok, aha. Gut. Und bei der Korrektur und beim Bewerten gibt es da Unterschiede bei den Kategorien. Fällt es Ihnen einfacher, gewisse Kategorien zu bewerten oder empfinden Sie alle als gleich von der Komplexität her beim Bewerten?

LP1 Ahm, also ich habe noch nicht solche Unmengen korrigiert (...) und ahm, (...) ich habe den Eindruck, dass es gleich ist.

I Mhm, ok.

LP1 Aber da fehlt mir eindeutig vielleicht auch noch die Erfahrung.

I Ich verstehe. Und wenn wir uns nun gewisse Deskriptoren genauer anschauen. (...) Ahm, vielleicht bleiben wir auch gleich bei der ersten Kategorie „Erfüllung der Aufgabenstellung“, der zweite Deskriptor lautet in Band zehn ja „Formuliert Titel/Betreff/Abschnittsüberschriften/Anrede/Grußzeile treffend“ und dann in Band acht ist es statt „treffend“ als „gut“ formuliert und bei Niveaustufe sechs als „angemessen“. Wenn wir uns jetzt diese Formulierungen ansehen, inwieweit sind sie für Sie voneinander abgegrenzt?

LP1 Ja, (...) ich schaue gerade am Raster. Das ist eine ist „treffend“ und „gut“ und dann „angemessen“. Ja, das ist ein bisschen schwammig. Das ist ein wenig abhängig von der Interpretationsweise eines jeden Lehrers, der das liest. Also (...) am meisten ist für mich die Formulierung in Band sechs „angemessen“ von „treffend“ noch zu unterscheiden. Aber zwischen „gut“ und „treffend“ ist es schon schwierig, das gebe ich zu.

I Mhm, und wie würden Sie dann vorgehen? Haben Sie für sich persönlich eine Norm oder beziehen Sie es auf die soziale Bezugsnorm, also zum Mittel der Klasse, wie eine Schülerleistung vergleichsweise in der Klasse ist. Oder auf eine individueller Ebene, also ob sich ein Schüler für sich selbst

verbessert hat in Bezug auf einen Deskriptor. Wie nehmen Sie also so einen Deskriptor war und setzten ihn dann in Verbindung?

LP1 Also da würde ich so vorgehen (..) ahm, ja, also ich würde von der Logik des Rasters ausgehen, dass Band 10 einem „sehr gut“ entspricht und dann eben abstufen, sodass acht „gut“ ist und neun eben dazwischen.

I Also für Sie haben die Niveaustufen eine eigene Bezugsnorm, die mit dem Notensystem in Verbindung zu setzen ist? Und je nachdem wie der Schülertext dann aussieht, wird mit dieser Norm, die Sie für sich, wie illustriert, definieren, der Text bewertet.

LP1 Hm (.) ja. Ja. Das ist wieder diese subjektive Geschichte dann, wo ich mir dann denke, das liegt im subjektiven Ermessen

I Die Frage ist dann, inwiefern ist dieses subjektive Ermessen innerhalb einer Lehrperson stabil von Text zu Text und ob sie sich verändert, textabhängig, oder ob sie gleich bleibt, beispielsweise.

LP1 Ja, fragen Sie mich, ob das für mich gleich bleibt?

I Ja, ob Sie das für Sie als gleich bleibend einschätzen würden oder ob Sie sagen, ja, das schwankt auch innerhalb einer Person auch wieder.

LP1 Ja, ahm, ich würde mal davon ausgehen, dass es gleich bleibt bei mir.

I Ok, ja. Sehr interessant. Ahm (..) und wenn wir uns nun die Kategorie „Spektrum sprachlicher Mittel“ ansehen, da wurden bei der Performanz, die Sie bewertet haben, bei Deskriptor eins acht Punkte vergeben. Der Deskriptor wäre, wenn wir auf den Raster schauen: „Hat ein ausreichend breites Spektrum an lexikalischen und strukturellen Mitteln“. Jetzt ist die Frage: Ist für Sie die Unterscheidung zwischen „ausreichend breit“ und zum Beispiel „genügend breit“ weit genug abgegrenzt oder wie fassen Sie das auf?

LP1 Ahm ja, (.) also „ausreichend breit“, da ist natürlich die Frage, wofür ausreichend breit und wenn es bei Band acht steht, dann muss es ausreichend breit für eine solide, gute Note sein. So interpretiere ich das. Und „genügend breit“, ja (..) hm, also, das ist vielleicht nicht unbedingt auf die Note vier bezogen, aber so genügend, dass es noch, ja, ein halbwegs passabler Text ist.

I Ok, wenn ich das jetzt richtig verstanden haben, dann sagen Sie, dass Sie für sich beispielsweise für „ausreichend“ festmachen, dass es eben eine (..) ähm (.) solide Leistung erzielt wird, und eben genügend ist für eine akzeptable positive Leistung.

LP1 Ah, ja!

I Also Sie binden diese Gradpartikel für sich selbst an einen Maßstab, der für Sie gleichbleibend ist.

LP1 Ja, also ich, ich richte mich da auch sehr stark nach der Punktzahl, je nachdem in welchem Band ich bin und versuche die Leistung darin einzuordnen nach meinem Maßstab, wie es zum Beispiel bei sechs Punkten für einen solchen Rahmen (.) ähm, ja, angemessen ist, eben noch eine positive Leistung. So versuche ich das zu interpretieren. Ich versuche das auch eben so zu interpretieren, wie die Autoren das gemacht haben und die schreiben da, ok, ja, in der Zeile für sechs Punkte das und das, dann werden die das so gemeint haben, dass in diesem Rahmen eben eine noch positive Leistung verstanden wird.

I Ja, interessant, ok. Ich hätte dann auch noch Fragen zur Anwendung des Raster. Wie Sie ihn anwenden. Wendest du den

Raster prinzipiell immer heranziehen beim Korrigieren von Schülertexten oder nur in bestimmten Situationen zum Beispiel?

LP1 Nein, immer!

I Immer?

LP1 Ja, also ich finde den super! Ich nehme ihn immer!

I Also auch bei Hausübungen zum Beispiel?

LP1 Ja!

I Mhm. Und dabei auch den gesamten Raster an sich, also die Kategorien mit allen einzelnen Deskriptoren? Oder korrigieren Sie Hausübungen nach den Kategorien des Rasters aber gehen nicht immer ganz genau mit dem Raster aller Deskriptoren durch, um quasi grundlegend die Basis des Rasters anzuwenden?

LP1 Also ich würde bei bedeutenden Texten den Raster natürlich wirklich ausnahmslos nehmen und bei Hausübungen würde ich das auch. Und wenn das mal so Texte sind, die ich zwischen Tür und Angel kriege, ja, also so By-the-way-Texte, dann korrigiere ich es nur nach diesen vier Kategorien, wo ich sage, da schaue ich schnell drüber. Ja, da wären es dann nur mehr die Kategorien

aber sonst bei Hausübungen, die ja auch Vorbereitungen sein sollen, für eine Schularbeit zum Beispiel, da würde ich es dann dementsprechend dann auch gleich ordentlich mit dem Raster korrigieren.

I Ja gut, ok. Und (.) ahm, wenn wir jetzt die Bezugsnormen im Klassenraum beachten, eben die sachliche Bezugsnorm, die an bestimmten Kriterien festgemacht wird, die individuelle Bezugsnorm, also die persönliche Verbesserung eines Schülers, und die soziale, also im Klassenverband. Wenn Sie mit dem Raster Texte bewerten, lassen Sie wirklich nur die sachliche Bezugsnorm an den Kriterien des Rasters miteinfließen oder spielt da die individuelle Bezugsnorm auch eine Rolle, dass Sie sagen, ok, der Schüler hat sich von einem Text zum anderen wirklich verbessert und das wird auch berücksichtigt, oder ist es wirklich rein sachlich? Beispielsweise ist es ein ausreichend breites Spektrum an einem objektiven Kriterium gemessen oder auf einer individuellen Ebene, eben ein ausreichend breites Spektrum für diesen einen Schüler.

LP1 Das ist eine echt gute Frage! Darüber habe ich mir noch nicht so genau den Kopf zerbrochen. Jetzt in diesem Fall würde ich antworten, dass ich (..) auch in Hinblick auf Menschlichkeit

darauf Rücksicht nehmen, wenn ich sehe, dass er sich verbessert hat und ich ihm im Zweifelsfall da eine bessere Note zugestehe, in Hinblick darauf, wenn er sich wirklich verbessert hat, sollte das schon honoriert werden.

I Mhm, ja. Und wenn man die soziale Bezugsnorm nun auch betrachtet? Ahm, würden Sie die Texte auch im Bezug zum Klassenverband irgendwie einstufen oder dem zuordnen oder wird das nicht miteinbezogen?

LP1 Nein, also da schaue ich mir schon jeden einzeln an und die Entwicklung, die er durchgemacht hat, weil ich dann irgendwie nicht will, (..) dass Schüler untereinander ein Konkurrenz-Verhältnis aufbauen und warum sollte Schüler A mit Schüler B verglichen werden. Das halte ich nicht für sehr sinnvoll.

I Ja und ahm, wenn wir jetzt gerade über die Bezugsnormen sprechen, meinen Sie, dass sich die subjektive Auffassung der Deskriptoren bei der Lehrperson an der sozialen Bezugsnorm orientiert? Wenn man also in einer leistungsstarken Klasse unterrichtet, dass man dann den Deskriptor anders interpretiert, als in einer leistungsschwachen Klasse, dass sich das Empfinden also der sozialen Bezugsnorm anpasst.

LP1 Nein, da halte ich es generell so, dass jeder ein gewisses Mindestmaß anstreben und erreichen muss und alles was mehr ist, ist umso besser, aber kein Bezugsmaßstab.

I Also steigt für Sie mit einem erhöhten Leistungsniveau der Klasse nicht automatisch die Leistungserwartung und Auffassung der Leistungsmaßstäbe.

LP1 Nein. Genau das bleibt gleich, orientiert an einem Mindestmaß. Ich halt es so, dass ein bestimmtes Mindestniveau zu erreichen ist und diejenigen, die motivierter sind, die erreichen mehr und das ist auch gut so und wird mit besseren, sehr guten Noten honoriert aber angestrebt wird das Mindestniveau und das verändert sich nicht durch die leistungsstarken Schüler und wird nicht angehoben. Da fährt die Eisenbahn drüber. Das bleibt so.

I Wie fühlen Sie sich im Umgang mit dem Raster? Fühlen Sie sich ausreichen professionalisiert oder wüschten Sie sich mehr Fortbildungsmöglichkeiten und Schulungen. Welche Bedeutung messen Sie dem bei?

LP1 Also ich denke, dass wir in der Lehramtsausbildung in Französisch, ahm, eine sehr ausgeprägte Fachdidaktikausbildung

haben und da ist das auch ausreichend aber nur bezogen auf den Raster könnten mehr Übungen stattfinden, sodass uns die Lehrveranstaltungsleiter mehr Übungstexte zum Bewerten zukommen lassen könnten. Aber extra nur eine Fortbildung zu dem Raster, also ahm das ist, glaub ich, nicht notwendig. Den beim Unterrichten ist das dann eh learning by doing und es ist ja auch wirklich sehr übersichtlich und einfach zu verstehen.

I Gut, wenn ich nun noch auf die Performanz eingehe. Und zwar haben die beiden Kategorien „Erfüllung der Aufgabenstellung“ und „Aufbau und Layout“ sehr gut abgeschnitten und die beiden sprachlichen Kategorien haben weniger gut abgeschnitten und waren nur knapp positiv. Es gibt hier also eine große Ungleichheit in diesen Kategorien. Vorhin wurde angemerkt, dass die sprachlichen Kategorien für Sie als die wichtigeren anzusehen sind Insgesamt kam es dann zu einer Note von einem Zweier. Wie empfinden Sie nun diese Bewertung?

LP1 Ahm (.) das ist so aufzufassen: Das System hat hier einen Bonus für die Lernenden, da alle Kategorien gleichwertig sind und die beiden Kategorien der „Aufgabenstellung“ und „Aufbau und Layout“ sind, wenn ich das so formulieren darf, die geschenkten Punkte, denn da kann man nicht viel falsch machen und die sind

relativ leicht zu holen, diese Punkte und damit kann ich natürlich dann im Kern, nämlich bei den sprachlichen Mitteln und bei der Sprachrichtigkeit wieder einiges kompensieren und dann schaut die Gesamtnote auch wieder ein bisschen besser aus.

I In diesem Text sind sprachlich gesehen eine Reihe an Mängeln erkennbar, allerdings sind sie nicht wirklich kommunikationsstörend und aus einer handlungsorientierten Perspektive konnte der Schüler seine Botschaft kommunizieren und hat somit den Auftrag erfüllt. Wie schätzen Sie die Handlungsorientierung in der Fremdsprache ein. Ist es das oberste Gut des Fremdsprachenunterrichts für Sie, dass man kommunizieren kann und die Sprachrichtigkeit, solange sie nicht kommunikationsstörend ist, nur als nebensächlich anzusehen ist? Wie sehen Sie das?

LP1 Ahm (..) ja, da würde ich es als das Bedeutendste im Französischunterricht einschätzen, dass die Schüler später im Berufsleben kommunizieren können. Das ist das Wichtigste. Aber in anderen Bereichen, wie in diplomatischen Diensten, ist das anders zu beurteilen, da dann die Richtigkeit genauso von Bedeutung ist. Aber nicht in der Schule, ja, da die

Handlungsorientierung. Bei den Schülern ist es wirklich das wichtigste, dass die Kommunikation funktioniert.

I Ok, vielen Dank. Prinzipiell wäre es das von meiner Seite. Wenn es keine weiteren Gedanken gibt, die geäußert werden wollen, dann würde ich hiermit das Interview beenden und mich nochmals herzlich bei Ihnen bedanken.

LP1 Nein, ich glaube, ich habe einen guten Einblick gewährt und habe gerne teilgenommen.

I Gut. Vielen herzlichen Dank für dieses interessante Gespräch!

Interview 2

I Sie wissen ja, dass das grundlegende Thema die Leistungsbeurteilung von schriftlichen Performanzen ist, wie empfinden Sie es, Schülertexte zu bewerten.

LP2 Ja, es ist natürlich eine Herausforderung (.) ahm, mit den neuen Beurteilungsrastern und Korrekturblättern ist es jetzt aufwendiger geworden (.) ich finde es jetzt schon grundsätzlich fairer (..) ja, ahm (.) es braucht jetzt einfach mehr Zeit. Es nimmt jetzt viel mehr Zeit in Anspruch und ich bin aber grundsätzlich (.) finde ich das ok aber vielleicht findet man ein besseres Mittel, also ein vereinfachtes.

I Also Bewertungshilfen allgemein stehen Sie positiv gegenüber?

LP2 Ja, genau, da dadurch alle aufgrund derselben Basis korrigieren.

I Und wenn wir gleich übergehen zu dieser Bewertungshilfe im Speziellen. Wie stehen Sie dem Beurteilungsraster auf Sprachniveau B1 gegenüber?

LP2 Genau, es ist einfach (.) ähm, jedes Mal, wenn ich es verwende (..) finde ich es schon sehr aufwendig, was die genauen

Unterpunkte angeht und einem abverlangt, weil ich muss mir das schon jedes Mal noch durchlesen, weil das alles nicht so genau im Kopf bleibt

I Wenn ich das also richtig verstehe, ist es schon sehr umfassend?

LP2 Genau zu umfassend, um es alles im Kopf zu haben Außer man hat halt ständig mit Klassen auf B1-Niveau zu tun aber ich habe eher A2-Klassen und da verwende ich dann ein anderes Raster. Also man muss das schon immer wieder durchlesen und ich finde, das ist dann doch schon (..) sehr (.) komplex und umfangreich.

I Und wenn wir uns dem Raster und seinen vier Kategorien zuwenden, wie stehen Sie dieser Aufteilung gegenüber und den einzelnen Kategorien an sich?

LP2 Grundsätzlich gut (.) also ich finde das wirklich nicht schlecht. Das sind Parameter, die man grundsätzlich schon als (..) Korrekturleser eines Textes innerlich (.) berücksichtigen würde aber so sieht man das noch einmal transparent vor sich. Das finde ich gut.

I Und wenn wir diese vier Kategorien heranziehen, sind diese ihrem Empfinden zufolge als gleichwertig zu betrachten oder empfinden Sie gewisse Kategorien als wichtiger oder weniger wichtig?

LP2 Ahm (.) Ich tue mir bei dem einen leichter als bei dem anderen. Zum Beispiel zwischen Spektrum sprachlicher Mittel und Sprachrichtigkeit, da neige ich persönlich dazu, dass ich das vermische teilweise, beziehungsweise auch die Aufgabenerfüllung und das Aufbau und Layout (..), insbesondere dass das Layout jetzt einen so hohen Stellenwert bekommt, ist manchmal (..) denke ich manchmal, dass das nicht ganz in Relation steht.

I Also empfinden Sie Aufbau und Layout als weniger bedeutend in Relation gesehen zu den anderen.

LP2 Ja genau, wobei es auch durchaus um wichtige Punkte geht, wenn ich mich jetzt näher damit beschäftige, wie Kohäsion und Kohärenz. Aber das ist nur einer und der letzte Deskriptor. Ja genau, und womit ich einfach auch ein Problem habe, sind so manche sehr globale Bezeichnungen, wie zum Beispiel klarer Gesamtaufbau oder schreibt fast durchgehend in einem der

Textsorte angemessenen Stil, das ist auch so eine Wischiwaschi-Formulierung, wo ich mir auch schwer getan habe (..), ob das jetzt mein Empfinden ist oder wie das aufzufassen ist.

I Wenn ich das richtig verstanden habe, empfinden Sie die Formulierungen doch als schwammig und eher ungenau und wenig präzise.

LP2 Ja genau. Und das hört man auch von Kollegen, wenn man mit denen spricht. Der Eindruck wird vermittelt als wäre das schön abgegrenzt aber das ist es nicht (.) weil das immer nur Ausprägungen sind und keine klaren Unterschiede zu erkennen sind.

I Ok, verstehe. Und wenn wir uns jetzt bestimmte Deskriptoren genauer anschauen. (..) Ahm, vielleicht bleiben wir auch gleich bei der ersten Kategorie „Erfüllung der Aufgabenstellung“, der zweite Deskriptor heißt in Band zehn „Formuliert Titel/Betreff/Abschnittsüberschriften/Anrede/Grußzeile treffend“ und dann in Band acht wird statt „treffend“ die Bezeichnung „gut“ gewählt und bei Niveaustufe sechs ist es „angemessen“. Wenn wir uns jetzt diese Formulierungen ansehen, inwieweit sind sie für Sie voneinander abgegrenzt?

LP2 Ahm (..) da hatte ich jetzt auch das Problem, wie ich mir die Formulierungen angeschaut habe und dann den Text.

I Gut, aber die Formulierung sprachlich ganz allgemein, ohne sich jetzt speziell auf diesen Text zu beziehen.

LP2 Ja , schwammig. Schwammig.

I Aha ok. Gut. Und wenn diese Formulierungen schwammig sind, orientieren Sie sich dann bei der Bewertung auch an der Höhe der Bände und den Punkten oder an einer eigenen Bezugsnorm.

LP2 Also wenn ich den Raster verwende, dann bringe ich das schon eher (.) ja, schon eher mit den Punkten in der Verbindung als jetzt rein mit den Formulierungen. (..) Also ich lese mir alles durch aber kreuze dann an, was ich für mich empfinde.

I Ok, also verstehe ich das schon richtig, dass Sie der Meinung sind, dass diese Formulierungen einen subjektiven Interpretationsraum offen lassen.

LP2 Ja. (.) Ja. (...) Wie soll das auch möglich sein, (.) dass ich jetzt (.) das Korrekte unter Anführungszeichen ankreuze. Was wäre hier das objektiv richtige? Wer sagt mir das? Das gibt es ja

einfach auch nicht.

I Ja, bei offenen Formaten und subjektiven Beurteilungsmethoden gibt es das eine Korrekte nun nicht. Glauben Sie, dass Sie für sich selbst eine intrapersonelle Bezugsnorm (.) ah, festgesetzt haben, sodass sie die unterschiedlichen Niveaustufen immer gleich auffassen trotz schwammiger Formulierungen?

LP2 Ja, also sobald es mir schwammig vorkommt, neige ich dazu, dass ich mich auf die Punkte beziehe.

I Ok (.) und die Punkte setzen Sie dann mit dem Notensystem in Verbindung und gleichen es daran an?

LP2 Ja genau, schon. Also ab 6 ist es positiv und dann arbeite ich mich nach oben. In dem Schulsystem ist das schon fest verankert und da kann man dann auch nicht anders. Das ist also nicht aus dem Kopf zu kriegen

I Aha, ja. Und prinzipiell ist der Raster ja als sachliche Bezugsnorm anzusehen, sodass rein kriterial bewertet werden kann. Lassen Sie persönlich dann auch die individuell Bezugsnorm eines Schülers miteinfließen oder berücksichtigen Sie auch die soziale Bezugsnorm in der Klasse?

- LP2 Ja, die Bezugsnorm ist, wie es bei jedem Lehrer sein wird, dass man die persönliche Komponente und den Bezug zum Schüler leider nicht aus dem Kopf kriegt und ausblenden kann. (.) Wenn man das also korrigiert, denkt man trotzdem an den einen Schüler, ich sage es ganz ehrlich. Und da gibt es die Schüler und Schülerinnen, die die Schlimmen sind und (.) neigen dazu, sich wenig vorzubereiten auf diverse Sachen. Und das kriegt man dann einfach nicht aus dem Kopf beim Korrigieren und da hilft mir der Raster dann schon. Ja.
- I Ja ok. Und empfinden Sie dann in diesem Zusammenhang die Bewertung mit dem Raster als objektiv?
- LP2 Da habe ich auch (..). Also da gibt es ein Problem, dass Schüler Phrasen und Sätze und teilweise ganze Texte auswendig lernen und selbst aber nichts produzieren können. Unterm Strich geht das sehr weit, dass sie unglaublich viel auswendig lernen oder etwas zu verstehen und dann haben sie schon eine super Bewertung, obwohl sie nicht wirklich eine eigene Leistung vollbracht haben. Und dann neige ich dazu, dass ich das im Raster weniger gut ankreuze.
- I Ist dieser subjektive Interpretationsraum dann also positiv zu erachten, sodass sie einen Spielraum frei haben?
- LP2 Ja, genau. Das ist dann positiv, wenn ich entscheiden kann, wie gut das ist, wenn ich weiß die Schülerin hat das nur stupide auswendig gelernt und kann es eigentlich nicht.
- I Wie gehen Sie dann heran, wenn Sie als Vergleich eine auswendig gelernte Performanz ohne Fehler vor sich haben und eine Performanz, die eindeutig selbst produziert wurde mit viel Bemühungen aber deshalb auch sprachliche Mängel aufweist. Wenn man das vergleicht (..)
- LP2 Dann ist mir die Bemühung lieber! (..) Die schneidet aber natürlich schlechter ab.
- I Ok, und glauben Sie, dass die Erwartungen und die Auffassungen von Schülerleistungen davon abhängen, in welchem Rahmen man agiert. Also man spricht vom sogenannten Referenzrahmeneffekt in der Forschung, was bedeutet, dass sich die Erwartungen einer Lehrperson dem Umfeld anpassen, beispielsweise steigen die Erwartungen leistungsstarken Schulen oder umgekehrt.

LP2 Ja, das suggeriert eh der Raster, meiner Ansicht nach. Er lässt so viel Spielraum offen, dass man das natürlich in unterschiedlichen Bereichen einsetzen kann und anpasst. Aber unterm Strich geht es dann darum zu zeigen, dass er ab einer positiven Note ein B1-Niveau erreicht hat oder geht es darum, welche Note er hat und das dann aber so anzuwenden, wie es gedacht ist.

I Mhm, ja, sehr interessant. Da spielt auch die Handlungsorientierung eine Rolle. Konnte der Schüler mit seiner Performanz kommunizieren? Ist seine Botschaft angekommen?

LP2 Ja, das hat grundsätzlich funktioniert. Somit hat er eigentlich trotz sprachlicher Mängel das B1-Niveau, weil er konnte kommunizieren.

I Welchen Stellenwert nimmt die Handlungsorientierung im Fremdsprachenunterricht ein, ihrer Ansicht nach?

LP2 Ähm (.) Also die Sprachrichtigkeit ist natürlich schon zu bewerten aber unterm Strich, gerade bei uns an der BHS, ist es so, dass der Schüler die Sprache auch im beruflichen Kontext verwenden sollte, ja. Und in der beruflichen Praxis geht das

nicht. Das ist inakzeptabel mit vielen Fehlern, das reicht nicht. (..) Da ist die sprachliche Richtigkeit schon wichtig.

I Ja, es kommt natürlich immer auf den Kontext an.

LP2 Ja, das stimmt natürlich. Es kommt immer auf den Kontext an.

I Wie wäre seine sprachliche Handlung als Tourist einzustufen, nicht als Arbeitender im Tourismus?

LP2 Ja, als Tourist schon. Das passt dann, schon.

I Weil wir hier gerade die Textsortencharakteristika haben aus dem Textsortenkatalog: Glauben Sie, dass jene Kategorien wie „Erfüllung der Aufgabenstellung“ und „Aufbau und Layout“ eher einheitlicher zu bewerten sind als die „Sprachrichtigkeit“ und die „sprachlichen Mittel“?

LP2 Ja, schon. Insofern, weil gerade bei der „Sprachrichtigkeit“ die Lehrpersonen unterschiedliche, höhere oder niedrigere, Erwartungen haben. Und „Aufgabenstellung“ wird am einheitlichsten sein, da ich das einfach abhaken kann und schaue, was ist erfüllt und was nicht. Bei den anderen wird es schwierig.

- I Aha, interessant. Wenn wir jetzt auf diesen Text eingehen. Wie war da zunächst der erste Gesamteindruck?
- LP2 Ja, er war eigentlich besser, bevor ich dann genau geschaut habe.
- I Neigen Sie dazu, dass Sie beim ersten Eindruck bereits den Text versuchen einzustufen auf einer Notenskala?
- LP2 Ja, ich neige leider dazu. Ich habe zuerst zu einem Dreier tendiert und beim genaueren Punkte vergeben wurde es dann knapp doch eher ein Genügend. Ich glaube, ich war dann (..) bei der „Sprachrichtigkeit“ doch ein wenig zu streng (.) vielleicht. Aber ich finde das Korrigieren und das Analysieren von Schülertexten interessant und vor allem auch mit diesen vier Farben, das finde ich gut, dem stehe ich positiv gegenüber. Aber der Aufwand, das ist einfach ein Wahnsinn.
- I Aha, ja. Wie wenden Sie den Raster denn prinzipiell an? Nur bei der Matura, also der standardisierten Reife- und Diplomprüfung oder bei Schularbeiten oder auch bei sämtlichen Hausübungen immer den Raster?
- LP2 Also ich vereinfache ihn auf das, worauf es wirklich ankommt (.) darauf minimieren, ja.

- I Also die Kategorien würden Sie beibehalten.
- LP2 Ja, die Kategorien würde ich beibehalten und dann würde aus Zeitgründen sehr prägnant pro Kategorie abarbeiten. Der Schüler hat 50 Minuten an dem Text gearbeitet, ich verstehe nicht, warum ich zum Bewerten länger brauchen sollte als der Schüler zur Produktion, deshalb eher vereinfacht, ja. Anders kann ich mir das aber nicht vorstellen. Denn alle diese einzelnen Punkte und alle diese Bänder durchzugehen, das geht nicht, das dauert einfach zu lange. (.) Es gibt auch einfach wichtigere Dinge, die in Relation zum Aufwand wichtiger sind, wie gute Unterrichtsvorbereitung, motivierende Lernmethoden, didaktisches Material (.) ah, ja. Es kann nicht sein, dass die meiste Arbeit auf solche Dinge dann fällt. Das ärgert ein wenig. Der Aufwand ist einfach zu groß (..) Der Stellenwert ist zu hoch, der Stellenwert wie man mit Beurteilungen umgeht allgemein. Lehrer sind hauptsächlich da, um zu Unterrichten eigentlich und weniger zum Raten. (..) Also grundsätzlich begrüße ich so eine Basis aber ich würde es vereinfachen.
- I Ja, mhm, gut. Also wenn ich das richtig verstehe, finden Sie es schwer, sich in dieser schwammigen Welt zu orientieren.

- LP2 Ja, genau. Sehr gut formuliert. (lacht) Ja.
- I Ok, prinzipiell hätte ich jetzt alle Fragen abgehandelt, wenn es von Ihrer Seite vielleicht noch Eindrücke, Meinungen oder Einblicke gäbe, gerne
- LP2 Vielleicht noch als wichtiger Punkt, was für eine objektivere Leistungsbeurteilung gut wäre, ist, wenn der Lehrer der Schüler nicht selbst beurteilt, sondern eine externe Lehrperson, da kann man das persönliche wirklich weglassen, das stört mich schon.
- I Ok, vielen Dank für die interessanten Einblicke und für das Gespräch.

Interview 3

I Sie wissen ja, dass das grundlegende Thema die Leistungsbeurteilung von schriftlichen SchülerInnentexten ist, wobei Sie als Lehrperson mit langjähriger Berufserfahrung viel Expertise aufweisen.

LP3 Wissen Sie, jetzt nach den neuen Umstellungen gar nicht mehr so. Es hat sich vieles verändert. Also das war für mich schon der schwere Hammer, weil endlich habe ich mir gedacht: „So, jetzt kannst du endlich entspannt unterrichten“ und dann, nach all diesen vielen Jahren, dann kam die Reform, die grundlegende Änderung. Und Französisch das ist schon ganz anders, ja. Es macht teilweise auch Spaß, weil mehr auf Hörtexte Wert gelegt wird als früher aber dafür tritt die Grammatik in den Hintergrund, meiner Meinung nach. Zum Korrigieren sind diese Bereiche natürlich wunderbar. Das geht echt flott und schreiben müssen sie weniger als früher. (.) Viel weniger, ja. Ich finde, früher durfte ein Schüler noch freier schreiben und da darf er sich dann auch entfalten und selbstständig Gedanken machen und jetzt bin ich in so ein striktes Format gepresst – ich finde das so arg! (.) Der Mensch wird genormt! Ich muss ja meine drei Punkte abarbeiten und darf nicht weniger, das verstehe ich mit

zehn Prozent, aber MEHR? Ich schränke dadurch den Schüler sowas von ein. Und da bin ich total dagegen. Das ist unsere Gesellschaft. Das sage ich jetzt als politisch bildende Geschichtslehrerin. Nein. Furchtbar! Und das ist Ausdruck unserer Gesellschaft. Alles schnell abarbeiten. So ein amerikanisches Schulsystem.

I Also Ihnen wird man zu sehr eingeschränkt als SchülerIn und die Aufgabenformate lassen zu wenig Freiraum?

LP3 Ja, ganz genau aber jetzt habe ich mich eben halbwegs adaptiert an diese Neuerungen.

I Mhm, ok. Und wie geht es Ihnen prinzipiell, wenn Sie schriftliche Texte beurteilen?

LP3 Ja also (..), wissen Sie, ich bin nicht die Wahnsinnige, die mit vier Farben korrigiert, also ich lasse mich da wirklich nicht ‚papierln‘, weil ich habe schon die ganze Zeit bemerkt, ich schau eh immer: „Was würdest du denn da jetzt geben?“ (.), ja, dann gebe ich die Punkte ein und ich komme meistens auf das, was ich eh von vornherein denke. Ja, also das ist so. Und wenn mir jetzt einer sagt, dass das sowas von transparent ist (..) naja, doch, also transparent schon, weil es zum Ausdrucken ist und jeder es

hat aber von Gerechtigkeit und Fairness kann nicht gesprochen werden. Ich kann hier schon einfach geben sieben Punkte und da soundso viele, also das ist doch NICHT WAHR. Ja, wenn ich zehn Punkte habe, kann ich immer noch jonglieren.

I Ok, also lässt Ihrer Ansicht nach die Sprache des Beurteilungsrasters Interpretationsraum offen?

LP3 Ja. Also ich finde, diese Diskussionen, die ja irgendwann einmal angefangen haben – früher war das ja nicht so – weil jetzt zeige ich die Liste her und sage, du hast so viele Punkte und alle sind plötzlich so zufrieden. Das ist schon Humbug. Das muss man doch sagen, ja. Wenn ich das auf Seminaren gehört haben (...) jetzt habe ich das Wort: die OBJEKTIVITÄT.

I Ja, bitte führen Sie Ihre Ansicht dazu aus!

LP3 Also ich kann WIRKLICH NICHT sagen, dass die Beurteilung mit dem Raster objektiv ist. Man will ja nicht betrügen, aber wenn man sieht das ein Schüler wirklich schwach und der gibt einfach wirklich alles und nur sein Bestes und irgendetwas bringt er einfach nicht zusammen (.) ja, dann will ich den doch nicht entmutigen und auf der Strecke lassen und wenn der nicht in Französisch maturiert, sondern einfach nur versucht einen

Vierer zu kriegen, wo ich sage, dass kann ich auch verantworten, dann gebe ich ihm den, denn (..) ja, wo bleibt denn da auch die Menschlichkeit den Schülern gegenüber.

I Ok, wenn ich das richtig verstehe, ist für Sie die individuelle Bezugsnorm auch von Wichtigkeit und nicht nur die rein sachliche Bezugsnorm, wie es der Raster ja eigentlich möchte?

LP3 Ja natürlich. Das MUSS ich doch auch berücksichtigen. Wenn der eine sich so plagt und dann hängt er einfach nur wegen dem und ich sehe, der macht und macht und macht (..) ja, was mache ich denn dann? Ich jongliere mit den Punkten, damit ich beweisen kann, dass es eine Vier ist.

I Also sie passen die Ergebnisse im Raster dann schon auch an und lassen dabei andere Bezugssysteme einfließen.

LP3 Ja, also das sage ich jetzt so, weil es eh anonym ist. Wenn ich das auf einem Seminar sage, dann heißt es gleich: „Nein, nein, nein! Also, das geht so nicht!“ (..) Aber das wird man doch nicht so machen. (.) Seien wir ehrlich, ich bin nicht die Einzige.

I Es ist nachvollziehbar, dass die Lehrperson, wenn sie auch beurteilt, nicht gänzlich andere Bezugsnormen ausblenden kann.

LP3 Ja. Das muss ich doch auch (.) schon. (.) Wir sind ja keine Maschinen.

I Ok. Ich verstehe. Es handelt sich beim Beurteilen von Schülertexten ja auch um subjektive Bewertungsverfahren.
Vielleicht gehen wir nochmal auf den Raster ein. Wie bewerten Sie Beurteilungshilfen ganz allgemein?

LP3 Ja schon (.) also ich meine, (..) ich war schon froh, das zu sehen, ja. Weil wenn ich das jetzt alles anders machen soll und auf diese Kriterien Punkte vergeben muss, (.) ja dann, möchte ich schon auch eine Anleitung habe.

I Gut, also eine Stütze zum Anhalten empfinden Sie als positiv.

LP3 Ja genau aber schade sind halt diese ‚Wischiwaschi-Sätze‘.

I Ok, darauf werden wir auch noch eingehen. Aber prinzipiell zu den vier Kategorien des Rasters. Wie erscheint Ihnen diese Aufteilung?

LP3 Wenn ich jetzt einen Schreibauftrag habe, wie es halt so ist, ja dann muss ich schauen, ob er alle Punkte abgearbeitet hat, dann ist es schon mal positiv und dann habe ich den Spielraum mehr

Punkte zu geben. Das finde ich schon gut. Aufbau und Layout (.), ok, wenn es Formate gibt, dann müssen die natürlich auch umgesetzt werden und ich muss halt wissen, wie es aufgebaut sein muss. Also ich kann ja schon wirklich viele Punkte da sammeln, schon mit dem Aufbau, denn wenn ich mir das einlerne, dann kann ich ja doch schon etliches kriegen. Bei den sprachlichen Kategorien wird es dann natürlich schwieriger, wie gut kann ich schreiben, wie kann ich meine sprachlichen Möglichkeiten einsetzen, nicht?

I Gut und scheinen Ihnen die vier Kategorien als gleichwertig oder empfinden Sie Gewisses als weniger wichtig?

LP3 Ja, also (.) Aufbau und Layout ist, finde ich, überbewertet. Die Aufgabenstellung ist natürlich schon wichtig, wie auch die Sprache natürlich.

I Ja, gut. Und wir sind vorhin ja schon auf die sprachlichen Formulierungen im Raster eingegangen und wenn wir uns nun gewisse Deskriptoren genauer anschauen (..) ahm, vielleicht bleiben wir auch gleich bei der ersten Kategorie ‚Erfüllung der Aufgabenstellung‘, der zweite Deskriptor lautet in Band zehn ja ‚Formuliert Titel/Betreff/Abschnittsüberschriften/Anrede/Grüßzeile treffend‘ und dann in Band acht ist es statt ‚treffend‘

als „gut“ formuliert und bei Niveaustufe sechs als „angemessen“. Wenn wir uns jetzt diese Formulierungen ansehen, inwiefern sind diese voneinander abgrenzbar? Ist hierbei für Sie ein großer Unterschied auszumachen?

LP3 Also angemessen heißt für mich, er hat es halbwegs hingekriegt und somit ist es positiv, ja. Bei gut, für mich (.) fehlt da halt noch ein bisschen was und bei treffend ist es dann ganz richtig (...) muss ja so sein! Ja.

I Beziehen Sie das eher auf die sprachlichen Formulierungen oder auf die Punkteanzahl, die diesem Niveau zuzuordnen ist.

LP3 Hm (...) Naja, das passt schon so (.) also zum Zuordnen in die Niveaustufe. Also ich orientiere mich schon oft an den Punkten, weil (..) ich muss ehrlich sagen, mich macht das ja auch verrückt diese Umschreibungen, also (.) ich merke mir das ja auch nicht alles und so viel. Ich muss jedes Mal durchlesen (.) und das ist mir schon ein wenig lästig. Ich hätte da vielleicht lieber nur ganz kurze und präzise Anweisungen, weil das wiederholt sich dann ja auch immer wieder. Ich muss ehrlich sagen, ich schaue immer kurz hin und mache das aber eigentlich nach Gefühl. Mir ist das zu blöd.

I Also orientieren Sie sich dann eher an den Punkten als den Formulierungen an sich.

LP3 Ja, schon eher an den Punkten und schaue dann (.), wenn ich im Zweifel bin, ja, was steht denn da und dort und dann entscheide ich. Im Zweifel.

I Mhm, ok. Interessant, ja. Also sind Sie nun der Ansicht, dass der Raster subjektiven Interpretationsraum offen lässt.

LP3 Ja. Eindeutig. Natürlich. Ja.

I Aha, ok. Gut. Und wir haben vorhin schon die individuelle und soziale Bezugsnorm angesprochen und verstehe ich Sie richtig, wenn ich davon ausgehe, dass Sie meinen, dass der Raster sachlich unterstützt aber die anderen Bezugsnormen ganz klar mit einfließen.

LP3 Ja, ganz genau. Der Raster ist eine Argumentationshilfe, das kann manchmal sehr gut sein. Eine transparente Anleitung, auf die kann man sich auch berufen und dafür bin ich auch dankbar.

I Vielleicht gehen wir noch darauf ein, wie Sie den Raster anwenden. Beim Korrigieren, verwenden Sie ihn da immer bei

- jedem Text oder nur bei Schularbeiten und der standardisierten Reife- und Diplomprüfung?
- LP3 Ja (.) schauen Sie. (lacht)
- Nein, also nicht immer. Um Gottes Willen, also da würde ich verrückt werden, wenn ich ihn bei allen Hausübungen verwenden würde. (lacht) Also da verlasse ich mich schon auf mein Bauchgefühl oder meinen Kopf beim Verbessern und auf meine Erfahrung. Bei der Schularbeit und nur bei der Schularbeit (..) außer ich weiß, jemand will in Französisch maturieren, dann schon, denn da muss das trainiert werden, weil die Matura wird natürlich auch so beurteilt. Das finde ich so auch vernünftig. Warum sollte ich die anderen auch damit quälen? Und das ist meine Zeit, die da viel intensiver beansprucht wird. Ich meine, die jungen Leute, die arbeiten so aber ich nicht.
- I Wenn Sie aber Hausübungen bewerten, ziehen Sie dann die vier Kategorien des Rasters heran oder lassen Sie ihn gänzlich weg?
- LP3 Oja, ich habe es natürlich im Kopf aber den Raster direkt verwende ich nicht. So halbwegs also, als Leitfaden könnte man sagen, verwende ich ihn.
- I Mhm, ja, sehr interessant. Und ahm, sie haben schon Seminare und Fortbildungen angesprochen, haben Sie schon welche zum Thema Beurteilungsraster besucht?
- LP3 Wir hatten – zum Glück – eine Schulung hier bei uns im Haus. Ich hatte teilweise schon sehr ungute Erfahrungen mit Seminaren aber ja das eine hatten wir und da wurde einem schon eingetrichtert wie objektiv das nicht ist aber ja, das ist natürlich nicht ganz so, wie behauptet wird. Aber eine Fortbildung hilft sicher. Wichtig ist auch, dass wir eine gute Französisch-Fachgruppe sind, die sich sehr unterstützt, und das ist besonders wichtig, finde ich.
- I Ok gut. Und wenn wir nun Bezug auf die Schülerperformanz nehmen, der hier bei der Forschung herangezogen wurde, dann haben Sie die Erfüllung der Aufgabenstellung gut bewertet mit acht Punkten und bei den anderen ein Mittelmaß von 5 Punkten. Wie haben Sie den Schülertext allgemein wahrgenommen und empfunden?
- LP3 Naja, ich glaube, das Vokabellernen gefällt ihm nicht so sehr, denn da fehlt es eben und auch so grundlegende Grammatik wie auch teilweise sprachliche Feinheiten aber ja (.)

- I Als Note haben Sie ein Genügend gegeben. Wenn Sie jetzt davon ausgehen, dass mehrere Lehrpersonen diesen einen Text beurteilen, meinen Sie, dass sehr ähnliche Noten dabei herauskommen oder weniger?
- LP3 Ich könnte mir vorstellen, dass manche vielleicht sogar Fünf sagen und das geben, ja.
- I Also die Forschung zeigt ja auf, dass natürlich hier mehrere unterschiedliche Noten vergeben werden und nicht immer gleiche Ergebnisse zustande kommen.
- LP3 Ja, das glaube ich gerne. Aber ich habe mir gedacht: „Ja, ok, du hast dich bemüht aber es fehlt doch noch einiges.“ Also ist das für mich ein Vierer gewesen. So hätte ich es gesehen.
- I Ja. Wenn wir diesbezüglich jetzt den Ansatz der Handlungsorientierung im Fremdsprachenunterricht heranziehen. Sind Sie der Meinung, dass der Schüler kommunizieren kann?
- LP3 JA! Natürlich, das kann er.
- I Wie schätzen Sie den Stellenwert der Handlungsorientierung im Fremdsprachenunterricht ein?
- LP3 Ja eigentlich, ist das eh wichtiger. Wenn der mal ins Ausland fährt, muss er was sagen können und die können sich das schon zusammenreimen und ob der da *des autres* oder *d'autres* sagt, ist auch egal. Also kommunizieren kann er.
- I Ist dann die Handlungsorientierung wichtiger als die Sprachrichtigkeit?
- LP3 Schon. Für den Anfang schon. Also für die Schule, ja. Weil man möchte, dass wenn er wohin fährt, dass er sagt, das hat er doch gelernt und kann er doch. Will er die Sprache studieren oder braucht sie im Beruf, dann muss er es eh verbessern, damit es ausreichend gut und richtig ist. Aber so sage ich, gerade Französisch, wird ja mittlerweile auch bald schon eine Exotensprache vergleichsweise zum Englischen oder den anderen und da ist das dann weniger wichtig.
- I Ja sehr interessant. Dürfte ich vielleicht noch nachfragen, ob Sie davon ausgehen, dass Sie für sich selbst eine intrapersonelle Bezugsnorm zu den Punkten, also eine Erwartungshaltung zugelegt haben, die immer dieselbe ist, woran sie SchülerInnentexte messen und in Bezug setzen?

LP3 Wahrscheinlich. Wahrscheinlich passiert das, ohne dass man sich dessen bewusst ist. Ja also eine gewisse Richtlinie schon.

I Gehen Sie auch davon aus, dass diese persönliche Richtlinie davon abhängig ist, in welchem Rahmen man agiert. Also man spricht vom sogenannten Referenzrahmeneffekt in der Forschung, wobei gemeint ist, dass sich die Erwartungen einer Lehrperson seinem Umfeld anpassen, beispielsweise steigen die Erwartungen dann in sehr leistungsstarken Klassen oder umgekehrt.

LP3 Ja natürlich ist das so. Das ist ja auch die Illusion, die alle haben, dass es immer gleich ist aber natürlich ist das abhängig und nicht immer vollkommen objektiv. Was soll man machen? Man muss das Beste herausholen und sich dabei natürlich seinem Umfeld anpassen.

I Gut, prinzipiell hätte ich keine spezifischen Fragen mehr und würde Ihnen Raum lassen, falls Sie noch einen Gedanken oder sonstiges mitteilen möchten.

LP3 Ja, (..) also prinzipiell, wenn man den Raster überarbeiten könnte und noch präziser machen würde, das wäre schon schön, denke

ich jetzt (..) für mich. Eine Anleitung prinzipiell finde ich schon sehr gut und wünschenswert aber in dieser Form nicht wirklich.

I Ok, vielen Dank für die interessanten Einblicke und ich freue mich, dass ich das Gespräch für meine Forschungszwecke heranziehen darf.

