

# Dicht oder lang

## - Überzeugungskraft freier Arbeiten DaF II

Rudolf Reinelt

### 4. 4. Ansätze zur Beurteilung

#### 4. 4. 1. Vorbemerkungen

Am Ende des einen Studienjahres können die Studenten nicht alles schreiben, was sie schon einmal gelernt haben: Das ist einfach zu viel. Schon im SS schreiben einige bis über 300 Wörter oder Phrasen, so dass das zum WS-Abschluss Geschriebene auf jeden Fall eine Auswahl ist. Allerdings schreiben einige Studenten mehr, einige aber extrem wenig. Unterschiede ergeben sich auch in den Jahren: Schon im SS17 haben viele Studenten in „Mein Deutsch“ über 150 Punkte gesammelt, während dies im SS2016 nur sehr wenige erreichten.

Das heißt aber, dass die Lerner Beispiele aus dem Unterricht bzw. damit zusammenhängenden Materialien bringen oder aber selbst erweitertes Sprachmaterial aufführen. Methodisch handelt es sich dabei nicht um offene Antworten zu Umfragen (Ainley 2014, Braun & Clarke 2006, Sara 2017), obwohl für die Analyse einige gleiche Inhaltspunkte gelten.

Dabei können einige Studenten sehr kurz sehr gute Beispiele dafür bringen, dass sie im Kurs etwas dem Kursziel des Zielsprachenlernens Entsprechendes gelernt haben. Dann wirkt ihre Arbeit dicht und überzeugend. Andererseits können andere Lerner in ihren Arbeiten selbst bei manchmal sogar viel größerer Länge und entsprechend mehr Wörtern und Phrasen einfach nicht überzeugend wirken. Da diese Überzeugungskraft aber Auswirkungen hat auf die Benotung, sollen im folgenden Beispiele für Beurteilungen gezeigt und dabei berücksichtigt werden, wie anfällig diese sind für Unterschiede in der Überzeugungskraft der schriftlichen Arbeiten.

Tests wie die hier Vorgestellten ermöglichen nicht nur den Lehrern, sondern auch (im Sinne eines *formativen Assessment* (Wicking 2016, 2017) den Studenten sich selbst zu zeigen, wieviel sie schon können. Andererseits sind solche Tests sehr selten, weil sie extrem schwer zu beurteilen sind, und auch nach längerer Suche wurde keine einschlägige Literatur gefunden. Andererseits benötigt der Kursleiter ein Beurteilungsverfahren, das sehr schnell angewendet

werden kann, auch wenn dabei kleine Unstimmigkeiten vorkommen, was bei der Beurteilung von schriftlichen Texten von vornherein nicht auszuschließen ist.

Vor diesem Hintergrund stellen wir vier Beurteilungsverfahren vor.

Es sei noch angemerkt, dass vor einer Beurteilung eigentlich zu checken ist, wieviel Inhalt aus dem Unterricht ist, und wieviel der Lerner selbst erstellt hat. Andererseits ist eine hohe Repeitivität bei lexikalischen Items aufgrund der beschränkten Mittel einer Sprache von vornherein vorgesehen, so z.B. bei der Perfektbildung.

Wir gehen hier also davon aus, dass für die Prüfung erst einmal alle benutzten sprachlichen Elemente (Wörter, Phrasen, Grammatik) legal, wenn auch nicht immer richtig sind.

Ein weiteres Problem ergibt sich aus den Kriterien bei den offenen Arbeiten. Jedes Beurteilungssystem kann diese selbst wählen, die *bottom-line* ist nur, dass ein daraus gebildetes System nicht systemimmanent negativ für die Lerner sein sollte.

#### 4. 4. 2. Punkte geben

Unmittelbar nach Abgabe der Papiere müssen diese beurteilt werden. Das heißt aber, dass ein schnelles und trotzdem möglichst faires Verfahren verwendet werden muss, um die Leistungen zu beurteilen. Lange Zeit hat der Verfasser eine erweiterte Version des folgenden Verfahrens benutzt (Details s. Reinelt 2018):

- Jedes Wort (s. aber o. die Ein-Buchstaben-Fehler-Regel) ist ein Punkt.
- Jede funktional richtig eingesetzte Phrase ist ebenfalls ein Punkt.

Das Beispiel in Tafel 5 gibt die Punktezahl pro Zeile an Wörtern und an Funktionen jeweils in Klammern am Ende der Zeile.

Tafel 7: Neujahrsdialog (1)

Übrigens, (9+1)	was	haben Sie	Noeiyer	denn so	alles gemacht
	An	Noeiyer,	habe ich	Hagoita	gespeilt.
Wo	(5+1) denn? (1+1)	Okayama			
Was	((2-1=) 1+1) denn? ((4-2=) 2+1)	Japanisch	Tennis		(2+1)
Mit	wem? (2+1)	Mit	mine	Swestar.	

Ist	(1+1) das	Spaß?		(3+1)
T=	(37)	Ja,	sehr.	(2+1)

*Diskussion (Z=Zeile):*

Alle Äußerungen können als funktionsgerecht in einem so geführten Gespräch angesehen werden. Deshalb kann man im Beispiel in Tafel 7 jeweils pro Satz +1 geben.

Z1 ein Punkt auch für *Noeiyer*, obwohl Neujahr bekannt sein sollte.

Z2 *Noeiyer* bringt in der Wiederholung keine Punkte.

Z3 *Denn* kommt häufig in dieser abschwächenden Funktion vor, deshalb, außer beim ersten Mal, kein Punkt.

Z4 *Eigentlich IN*, dieser Punkt wird nicht gegeben.

Z5 Gemeint ist wohl: *Was ist das denn?* = 4 Punkte.

Z6 Ohne Flexion, trotzdem ein Punkt, auch wegen der guten Idee.

Z8 *Mine* gibt es als deutsches Wort, *Swestar* hat zu viele Fehler.

Z9 *Ist* fälschlich für *macht*, beeinträchtigt aber das Verständnis nicht, deshab der Punkt.

Insgesamt erhält der Lerner hier 37 von 41 Punkten. Um eine Gesamtnote zu erstellen kann man alle Punkte einfach zusammenzählen, und wenn man noch strenger sein möchte, die ausgelassenen Wörter, Phrasen, usw. abziehen, was aber normalerweise keinen großen Unterschied ausmacht.

Nach dieser Zählweise haben die Arbeiten der Studenten 2, 6, und 3 in Tafel 4 bis 6 die folgenden Punkte (Tafel 8). Leider können die im Kommentar angegebenen Auffälligkeiten durch diese Punktegebung nicht gebührend berücksichtigt werden. Selbst eine doppelte Punktegebung für Teile aus dem WS würde nur bedingt helfen, weil im WS natürlich auch Sprachmaterial aus dem SS benutzt wird und klare Trennungen deshalb nicht so einfach möglich und vielleicht auch gar nicht sinnvoll sind.

Tafel 8: Charakteristika der ausgewählten Texte

T=Textnummer, P=erreichte Punkte, A= Auffälligkeit, K=Kommentar

T	P	A	K
2	199	nur wenige Teile aus dem WS	
6	158	nur etwa drei Viertel so lang wie 2	etwa die Hälfte der Sätze hat Inhalte aus dem WS.
3	89	noch kürzer als 6	fast nur Textteile aus dem WS oder einer Kombination aus Sommer- und WS

Der Vorteil dieser Beurteilung liegt darin, dass allen studentischen Deutschbemühungen entsprechende Punkte gegeben werden können. Damit kann man die Gesamtbemühungen der Klasse ebenfalls gut feststellen. Schwierigkeiten ergeben sich aus Wiederholungen und wenn dieselben Wörter und Chunks in verschiedenen Positionen vorkommen, so dass doppelte Punktzuerkennungen erfolgen können, die die Objektivität und die Reliabilität reduzieren. Außerdem sagt die Wortzählung oft nichts über die Qualität der Einträge aus, auch wenn z.B. bei Unverständlichkeit Funktionspunkte nicht zuerkannt werden. Schließlich ist die Zählung selbst kraft- und zeit-raubend. Nach dieser Beurteilungsmethode sind Lerner mit langen Beiträgen sicher im Vorteil, eine „inhaltliche Dichte“ kann nicht angemessen berücksichtigt werden. So könnte ein Lerner im Extrem einfach die Zahlen von null bis tausend aufschreiben und würde dafür eine ebenso große Punktezahl bekommen. Dieser Ansatz kann gerade nicht zwischen inhaltlich guten und weniger guten Arbeiten unterscheiden.

#### 4. 4. 3. Pro Itemblock ein Punkt

Die im Unterricht behandelten Inhalte bzw. Inhaltsblöcke kann man jeweils als einen Item betrachten. Dann kann man nachsehen, wieviele und wie gut ein Student diese Blöcke in seinem Testpapier behandelt hat. Für jeden Item-Block kann man dann bis zu einem Punkt geben, und entsprechend weniger, wenn es Probleme gibt. Ich danke M. Albers für einige Diskussionen in dieser Hinsicht.

Wir nehmen dasselbe Beispiel wie in Tafel 7:

Tafel 9: Neujahrsdialog als ein Item

Übrigens,	was	haben	Sie	Noeiyer	denn	so	alles	gemacht
Wo	denn?	An	Noeiyer,	habe ich	Hagoita	gespeilt.	(1)	
Was	denn?	Okayama						
Mit	wem?	Japanisch	Tennis				(1)	
	(0,7)	Mit	mine	Swestar.				
Ist	das	Spas?	Ja,	sehr.			(1)	T1=0,8;
T2=3,7/4=0,9								

#### **Diskussion**

Zwei Zählweisen bieten sich an:

T1 als Ganzes gesehen, mit ein paar Abzügen für die nicht ganz perfekten Zeilen fünf und acht, ergibt 0,8 Punkte für das Beispiel.

T2 ergibt unter Berücksichtigung des etwas fehlerhaften Blockstarts und der vier Weiterungen (Frage-Antwortsequenzen), die dieser Lerner hinzugefügt hat, obwohl sie nicht beim ersten Lernen dabei waren, eine Gesamtzahl von etwa 3,7 von 4 Punkten, also einen Durchschnitt von etwa 0,9. T1 und T2 sind sehr gute Ergebnisse. Schließlich kann man wiederum den Durchschnitt der Durchschnitte aller von dem Studenten im Test behandelten Themen ermitteln und damit zu einer Endnote kommen.

Anwendung auf die Texte 2, 6 und 3

Die Anwendung auf die Beispiele 2, 3 und 6 ist allerdings viel schwieriger. So ergibt in Beispieltext 2 die Reihung *Guten Morgen, Guten Tag, Guten Abend* einen Punkt, und in Zusammenhang mit einem Gespräch als dessen Anfang mit dem Gespräch zusammen wiederum einen Punkt. Im Beispiel sind bei einigen Sätzen die Satzanfangsteile weggelassen, so dass der Teil von *Guten Tag* bis *ich studiere* 0,8 Punkte ergibt. Solche Überlagerungen kann man nicht ausschließen und sind Teil unseres Sprachverhaltens, nur eben schwer zu bewerten. Sprachliche Elemente, die nicht syntagmatisch (wie oben z.B. als Reihen) oder paradigmatisch (z.B. als im Dialog ersetzbare Äußerungen) verbunden (und damit bedeutungsmäßig eingeordnet werden können, erscheinen dann als „Inseln“ und erhalten keine Bewertung (obwohl das Aufschreiben dieser Elemente (Wörter, Phrasen) ja auch schon eine minimale Leistung ist im Vergleich zu Studenten, die dies nicht gemacht haben).

Zählt man die Punkte aller Blocks zusammen und teilt diese durch die Anzahl der Blöcke, erhält man eine Zahl zwischen null und eins, die als „Durchschnitt“ für diesen Text angesehen werden kann. Die Beispieltexte erhalten die folgenden Punkte (Tafel 10):

Tafel 10: Punkte aus Blöcken

T=Text, P von B Punkte aus Blöcken, D = Durchschnitt

Text	P von B	D
2 lang:	7, 3 von 12	0,6
6	9,1 von 11	etwa 0,8
3	9,9 von 11	etwa 0,9

Das Problem bei dieser Beurteilungsmethode ist, wie bei 4.4.2 auch, dass es schwierig ist, die Gesamt(höchst)zahl an möglichen Äußerungen aus den im Unterricht behandelten sprachlichen

Teilen und deren Kombinationen, d.h. eine Gesamtzahl der möglichen Punkte, festzulegen, relativ zu der die erreichte Punktzahl zu berechnen ist. Dies mag im SS mit den noch begrenzten sprachlichen Möglichkeiten noch möglich sein, aber im WS kann schon die – ja im Sprach-

unterricht gewünschte – Kombinationsfähigkeit der Lerner zu -zig möglichen - und dann wie einberechenbaren? - Varianten führen. Noch selten wird es sogar vorkommen, dass ein ganzer Block mehrfach angerechnet werden muss (*wo wohnen Sie* als Teil des Anfangsgesprächs und in der Wegbeschreibung).

Ein weiteres Problem ergibt sich, wenn Studenten Listen auführen, z.B. die im WS behandelten Kleidungsstücke *die Hose, das Hemd, die Schuhe*, usw. Das Wissen und deren Darbietung, z.B. mit richtigem Geschlecht, ist sicher eine Leistung, aber im Rahmen dieser Beurteilung schwer zu erfassen. Mit dieser Beurteilungsmethode sind Lerner, die ihren Text „dichter“, d.h. mit mehr Inhalten in weniger Worten, verfassen, sicher im Vorteil, weil gleichbedeutende Elemente nicht zusätzlich zählen.

Schließlich bleibt noch ein Problem bei der Gewichtung. Um die Wichtigkeit des WSs zu betonen, könnte man alle Teile aus dem WS z.B. anderthalbfach punkten. Da im WS aber nicht nur WS-spezifische Inhalte gelernt werden, ist es schwer diese Zuteilung durchzuhalten, wenn solche Elemente z.B. in Sätzen aus dem SS verwendet werden.

Immerhin könnte man eine bestimmte Anzahl von thematischen Blöcken festlegen, die in der Arbeit vorkommen sollten. Dies macht die Beurteilung ähnlich wie in Tafel 3.

In den bisherigen Beurteilungsverfahren wurden jeweils Elemente aus den von den Lernern produzierten Texten in eine bestimmte Relation zu einer Zuerkennung von Punkten gesetzt.

Die beiden bisherigen Verfahren ließen sich leicht automatisieren, wenn die Arbeiten der Studenten maschinell eingelesen werden könnten. Dann wären diese mit dem als behandelt geltenden Unterrichtsfile vergleichbar: Wörter könnten direkt gesucht und Phrasen als *regular expressions* auf ihre Verwendung gecheckt werden. Dies wird mit Verfahren des Textmining, d.h. den Text als „Visual text“ zu analysieren (Text mining n.d.) ermöglicht. Ein besonders für Texte ausgelegtes Tool ist AntConc (AntCONC 2016). Leider sind die Textumwandler wie OCR noch nicht so gut, dass dieses Verfahren anwendbar wäre.

So sind die bisher vorgestellten Verfahren umständlich, und ob sie genauer sind als solche, in denen der Eindruck der Beurteiler herangezogen wird, ist alles andere als klar. Im folgenden betrachten wir ein Beispiel für den letzteren, in westlichen Ländern und bei vielen Einschätzungsverfahren üblichen Typ.

#### **4. 4. 4. Ganzheitlich nach Kriterien**

Einige Probleme der Beurteilungsverfahren in 4.4.3. und 4.4.4. können wir umgehen, wenn wir nicht die einzelnen Items im Detail ansehen, sondern die Gesamtleistung, diese aber nach einem Kriterienraster beurteilen, so dass wenigstens ein bisschen von der Feinheit erhalten bleibt.

Um zu zeigen, dass diese Methode zu ähnlichen Ergebnissen wie die beiden vorherigen führt, haben wir die Mitarbeit zweier Muttersprachler gewonnen, denen hier auch dafür ausdrücklich gedankt sei.

Schon als Schüler haben wir gelernt, dass etwas als Leistung angesehen und benotet wird. Mit der Zeit haben wir auch ein Einfühlungsvermögen dafür entwickelt, ob etwas gut oder schlecht ist, und viele Menschen können noch einige Grade dazwischen unterscheiden. Viele Lehrer (und z.B. beruflich auch Profiler) können zusätzlich noch nach einigen weiteren Kriterien getrennt unterscheiden, und inwieweit diese erfüllt sind. Das nächste Beurteilungsverfahren basiert auf diesen Annahmen und ist wegen seiner „Normalität“ weit verbreitet. Es ist eine Mischung aus ganzheitlicher und kriterienorientierter Beurteilung. Beim einen ist normalerweise der ganze Text Gegenstand der Beurteilung, und nicht nur einzelne Blöcke, wie in 4.4.3. Andererseits können/müssen die einzelnen Kriterien vorgegeben werden, nach denen der Text beurteilt werden soll. Diese können ganz verschiedener Art sein, aber viele Sprachtests bieten Listen an und der Verfasser dieser Arbeit hat eine eigene sinnvolle und handhabbare Liste erstellt. Schließlich muss die Abstufung der Beurteilung festgelegt werden. In diesem letzten Punkt habe wir alle den Vorteil, dass wir uns auf unser Vorwissen verlassen, d.h. im einfachsten Fall die uns vertrauten Schulnoten verwenden können (in Deutschland 1 bis 5, 6 wird normalerweise nicht gegeben).

Da bei nicht-professionellen Beurteilern die Unvoreingenommenheit wichtiger ist, als z.B. das Einhalten eines Noten“bandes“, wurde kein spezielles Training durchgeführt, sondern an das eigene Vorwissen und Vorverständnis der Beurteiler appelliert. Jede noch so gute Anleitung ist immer auch ein Unsicherheitsfaktor, weil abgesehen von professionellen Beurteilern niemals sichergestellt werden kann, dass verschiedene Beurteiler – außer mit beträchtlichem Aufwand - das gleiche unter einer Note (nstufe) verstehen.

Konkret heißt das, es wurden den Beurteilern keine genauen Ziele vorgegeben und auch keine Beschreibungen für die Notenstufen (vgl. aber dazu unten den Rubric in Teil 4.4.6).

Da - um einen Gesamteindruck zu ermöglichen - den Muttersprachlern nicht im voraus angegeben werden konnte, was inhaltlich behandelt worden war und was von den Lernern erwartet wurde, wurden einige Erläuterungen vorangeschickt. Dabei wurden, um Einflüsse aus der graphischen Gestaltung oder anderen Besonderheiten auszuschließen, einige Vereinheitlichungen vorgenommen. Technisches und Themenstellung und Erläuterungen im Kopfeintrag finden sich in Tafel 11 und die Kriterien in Tafel 12:

Tafel 11: Kopfeinträge des Fragebogens: Technisches und Themenstellung und Erläuterungen

Technisches	Themenstellung und Erläuterungen
<p>x-für leere Anfänge                      B=Bild oder Graphik                      NICHT ZU                      Berücksichtigen da                      aufgrund technischer                      Probleme ersetzt oder                      weggelassen:                      ohne ä,ö, ü und ss oder.                      oder, am Satzende.                      Du und Sie auch                      gemischt.                      Um graphische                      Vorbelastung des                      Eindrucks zu vermeiden:                      - pro Wort ein Feld,                      - pro Satz/Einheit eine                      Zeile                      - Dialoge sind aufgelöst!                      Listen als Listen                      geschrieben, nicht als Folgen                      Fehler so getippt, wie sie                      gemacht sind!</p>	<p>Thema: Ich                      (aber eigentlich kann man natürlich alles schreiben: Es gibt                      keine Themenbindung!)                      Die lockere Anweisung: Sich selbst mit den erhaltenen deut-                      schen Mitteln, auch dialogisch, so darzustellen, dass ein                      Firmenchef bei einer Bewerbung den Verfasser sofort nimmt                      d.h. einstellt.                      In „Ich“ werden Inhalte aus dem WS erwartet, solche aus                      dem SS sind aber auch o.k.                      Sprachliche Elemente aus dem WS sollten vorkommen, um                      zu zeigen, dass man daran teilgenommen hat, bzw. diese                      beherrscht.                      Sprachliche Elemente aus dem SS (Guten Tag, Wie gehts,                      Freizeit, pers. Daten, Essen, Trinken, Familie, heute abend                      (Gegenwart/Futur), Waren (preis) sind auch o.k. aber                      nicht wichtig.                      Es geht eher um die Elemente/Inhalte aus dem Wintersemes-                      ter, z.B. Wegbeschreibung, Wohnung (wie + Adjektiv,)                       haben/es gibt, Kleidung, Modalverben, Farben, und alle                      Perfektformen (gestern, letztes Jahr, usw.) (Dativ und                      Präpositionen nur am Rande erwähnt), Verabredung)                      Noch nicht: Nebensätze jeder Art, Dativ-und-Akk-verbren</p>

Tafel 12: Kriterien (jeweils wie in Deutschland (1 bis 6) zu beurteilen):

Art der Beurteilung	Kriterium
Einzel	Wortschatz
einzel	Rechtschreibung
Einzel	Grammatik
einzel	Inhalt
normorientiert	Ziel erreicht?
holistisch	Interessant?
holistisch	Mein Gesamteindruck
	Kommentare

Um festzustellen, ob die Werte nicht reiner Zufall sind, muss man bei mehreren Beurteilern immer den Übereinstimmungsgrad ermitteln (*ICC Inter/ intrarater correlation coefficient*). Dabei sind alle Werte der Beurteiler für alle Studenten zu berücksichtigen. Die Werte wurden aus Gründen der Vereinheitlichung und auch um sie für die hiesige Ehime Universität relevant zu machen, automatisch auf deren 100 Punktesystem umgerechnet.

Für die 11 Texte gaben die beiden Beurteiler (und RR zur Kontrolle) folgende Punktzahlen bzw. Prozentpunkte (Tafel 13), die dann in den ICC eingehen.

Tafel 13: Punkte von OIA, LW (jeweils nach Einzelkriterien und holistisch) und zur Kontrolle RR

	LiWi	LiWi	LiWi	OIA	OIA	OIA2	RR
	Einz.	hol2	hol %	Einz.	Hol2	Hol%	
1	76	3	33.3	88	1.5	66.7	71
2	79	2.5	40	77	3	33.3	66
3	89	1.5	66.7	96	1	100	61
4	71	3	33.3	73	3	33.3	75
5	68	3.5	28.6	52	4.5	22.2	93
6	81	2.5	40	96	1	100	98
7	84	2	50	83	2.5	40	62
8	76	3	33.3	69	4	25	71
9	84	2.5	40	82	2.5	40	98
10	73	3	33.3	75	3.75	26.7	50
11	89	2	50	82	2	50	92
	ICC		ICC	ICC		ICC	ICC

Zwischen OIA und LW war der Wert mit 0.79 außerordentlich hoch. Wurde die Benotung von RR einbezogen, ergab sich immer noch ein gut akzeptabler 0.68 (jeweils von 1).

Da also bei der Übereinstimmung keine Probleme entstehen, kann man auch die Einzelergebnisse betrachten. Aus der Gegenüberstellung der 5 Urteile nach Kriterien und 3 holistischen Urteilen von den beiden Muttersprachlern ergeben sich folgende Ergebnisse für die Studenten 2,6, und 3 (s. Tafel 14). Außerdem haben die beiden Beurteiler freundlicherweise Kommentare zu ihren Noten mit angefügt, die das Zustandekommen der Note erklären sollen, Beurteiler A für jedes Kriterium, Beurteiler B für jeden Text einen Kommentar.

Tafel 14: Noten und Kommentare von zwei Muttersprachlern zu 2,3, und 6

St	S=Rechtschreibung	G=Grammatik	L=Vokabular St=Student	D=dialogisch/als Reihe H2=insgesamt	V=Vielfalt H1=interessant?	H1	H2	K=Kommentar
2	3+, nichts Originelles, aber keine Wortschatzfehler	3, sehr viele Rechtschreibfehler, allerdings bei sehr vielen Sätzen	3, recht viele Grammatikfehler	2-3, sehr viele Sätze insgesamt, aber nur wenig Originelles	Eher 2 oder 3, sehr viele Sätze, die wohl aus dem Semester sind (gut 50 Sätze), viele Themen abgedeckt, teilweise aber nicht ganz vorhanden (Perfekt)	3	3	sehr viele Sätze vorhanden (50+), aber fast nichts Originelles, sondern Behaltenes aus den Stunden mit vielen Fehlern
2	2	4	3	2	2	2	2.5	guter Wortschatz, aber viele Rechtschreibfehler, besonders bei den Farben. Aufgrund der vielen Infos aber interessant zu lesen.
3	1-2, viele Eigenständiges	1, kaum Fehler hier	1-2, kaum grammatische Fehler	1-, einige schöne Eigenständige Sätze	Eher 1-2, Grammatiksaß bei den Perfektformen mal ja, mal nicht,	-1	1	im Gegensatz zu anderen sehr viel Eigenständiges, viel Gelerntes und wenig Fehler. Tendenziell auch ein paar sehr einfache Sätze. Dafür aber ein zusammenhängender Text
3	2	2	1	2	2	1	1.5	fast fehlerfreie Grammatik und Rechtschreibung, die paar Perfektformen sind auch gut.
6	1-2, viele eigenständige Vokabeln	1-2, kaum rechtschreibfehler	1-2, mehr Grammatikfehler, allerdings durch viele eigenständige und lange Sätze	1, überwiegend eigene brauchbare Sätze zu den selbstgewählten Thema, sehr viel Eigenständiges	Eher 1	1	1	im Gegensatz zu anderen sehr viel Eigenständiges, offensichtlich gelernte und selbst erdachte Sätze statt nur unvollständig Wiederholtes
6	3	2	3	2	2	2	2.5	das mit dem Perfekt klappt gut.

#### Anmerkungen zu Tafel 14

Es fällt auf, dass unbestimmte Mengewörter (*viel, mehr, etwas*) häufig benutzt werden. Originalität scheint auch wichtig zu sein, wie das Nicht-Abschreiben aus dem Lehrmaterial. Insgesamt erhalten die Studenten also folgende Punkte von den zwei deutschen Muttersprachlern:

Text 2 / 57 Punkte

Text 6 / 79 Punkte

Text 3 / 87 Punkte

Das hier vorgestellte Verfahren verbindet Noten für einzelne Kriterien mit einem Gesamteindruck. Die Ergebnisse sind nachvollziehbar und auch mit denen in 4.4.2 und 4.4.3 vergleichbar. Allerdings wird dabei nur *über* bestimmte Aspekte der Texte etwas gesagt, aber nicht *die einzelnen Textteile selbst* beurteilt.

Was nicht an der Oberfläche erscheint, und erst in den Kommentaren klar wird, ist, wie die Noten zustande kommen, d.h. ob der Lerner dichter oder länger oder sonstwie auffällig gut oder schlecht geschrieben hat. Dies ist eine Andeutung, wie die Studenten die Beurteiler zu überzeugen

gen versuchen. Allerdings werden einzelne Besonderheiten nicht ausführlich berücksichtigt. Das folgende Verfahren versucht diesen Nachteil zu erfassen.

#### **4. 4. 5. Strategisch Schreiben: Auffälligkeiten**

Eine Möglichkeit, um mehr als nur die offensichtlich erfassbaren Kriterien einzubeziehen, ist, jede einzelne Leistung als Auffälligkeit anzusehen, z.B. im Gegensatz zu Fällen, in denen sie nicht da ist. Diese Leistungen können dann im Sinne der Ziele der Lerner als mehr oder weniger gelungene Strategien aufgefasst werden. Bei der Bewertung bieten sich die folgenden Ansatzpunkte an:

- In den Texten sind Zeilenwechsel ein Kennzeichen einer von den Lernern vorgenommenen Einteilung für eine Leistung, die nicht nur dazu benutzt wird, in einem Dialog einen neuen Sprecher usw. anzuzeigen, sondern natürlich auch anzuzeigen, wenn etwas als Anfang eines neuen (Teil)items angesehen werden soll (s. 4.4.3).
- Innerhalb eines als inhaltlich zusammenhängend anzusehenden Abschnittes oder Blocks oder einer syntagmatischen oder einer paradigmatischen Reihung können die sprachlichen Teile nach ihrer Richtigkeit, usw. bewertet, werden sowie auch danach, wie weit sie sinnvolle sprachliche Handlungen darstellen. Somit kann auch die Qualität ansatzweise berücksichtigt werden.
- Schliesslich können Besonderheiten angemerkt werden und dafür Belohnungen (oder Abzüge) gegeben werden.

Die Berechnung erfolgt dann wie folgt:

Prinzipiell wird pro Zeile (oder entsprechender Einheit) ein Punkt vergeben, dieser aber bei besonderer Leistung z.B. auf 1.5 aufgestockt, oder bei besonders problematischen Inhalten entsprechend reduziert. Am Ende wird der Durchschnitt an Zeilenpunkten ausgerechnet.

Da das Notenband der Ehime Universität nur 100 Punkte vorsieht, viele Studenten aber besser schreiben, als nur das, was sie im Unterricht gelernt haben, muss man einen Bereich jenseits dieser Grenze vorsehen. Dies ist schon eine immanente Eigenschaft aller Sprachen: Mit begrenzten Mitteln (theoretisch) Unbegrenztes ausdrücken.

Durch die Möglichkeit, mehr als 100 Punkte zu erreichen, entsteht die Notwendigkeit, einen bestimmten Bereich für solche Studenten offenzuhalten. Dabei kann man z.B. den Bereich ab 90 Punkte, d.h. die letzten 10 Punkte mit einem Logarithmus verteilen lassen. Dabei ergibt sich die Notwendigkeit, eine Obergrenze festzulegen. Dies wird im nächsten Teil weiter ausgeführt, wir können aber hier einmal 0.4 Punkte als Grenze zum Übergang zur nächsten Stufe ansetzen. Die Begründung ist, dass Studenten, die besonders gut schreiben, dies normalerweise

auch an verschiedenen Stellen tun, so dass man eine größere Stufe annehmen kann. Diese Stufe ist natürlich variabel, muss aber aus Fairnessgründen zumindest in einer Beurteilungseinheit (Klasse, usw.) gleich gehalten werden.

Bei dieser Zählweise ergeben sich die folgenden Punktwerte für die oberen Bereiche: 0-89= die erreichte Prozentzahl, d.h. 0.89, dann z.B. 90 Punkte ab 0.94; 91 P. ab 0.98; 92 P. ab 1.2; 93 P. ab 1.6, usw. Ab 4 Punkte (99) ist keine Unterscheidung mehr möglich, so dass es danach 100 Punkte gibt.

Als Beispiel sieht Text 3 mit den Anmerkungen des Verfassers als Hinweis und bewertet dann so aus wie in Tafel 15.

Tafel 15: Text 3 nach Zeilenpunkten und Auffälligkeiten bewertet

(3)	W	Ich	kan	Klavier	spielen.				
	W	Am	letzte	Wochenende	habe	ich	Klavier	gespielt.	
R		Zwei gute Beispiele, überzeugen von zwei Konstruktionen							
		Das	ist	sehr	sehr	gut.			
-----3	+3	(können, Perfekt, Bewertung) sinnvolle Reihe							
		Im	Februar	fahre	ich	nach	Kyoto	mit	mienem
		Freundin							
		Ich	gehe	Kiyomisdera	und	Kinkakuji	und		
		Arashiyama							
		In	Kyoto	esse	ich	Yatsuhashi.			
		Yatsuhashi	ist	lecker.					
R		Hervorragende Reihung!							
-----4		(Reise, Ziele,Spezialität, Geschmack, aber alle nicht besonders schwer, deshalb normale Punktzahl) -0.2 (leichte Fehler)							
	W	Ich	habe	einen	Tisch	und	ein	Bett	und eine
		Uhr.							
	W	Der	Tisch	ist	weiss.				
	W	Ich	kaufe	den	bei	Nitori.			
		Der	ist	gut.					
R		Hervorragende Reihung!							
-----4+4		(2 syntagmatisch und 2 paradigmatisch, Geschlecht beibehalten (schwer), Akkusativ, Det.-verwendung sehr gut).							
		Erreichte Punktzahl: 17,8							
		Anzahl der Zeilen: 11							
		Endprozentzahl: 1,3							
		Punkte im 100 Punktesystem (für diesen Teil): 92							

*Anmerkungen*

- Durch die Schwierigkeit der Einteilungen, da Zeilen nicht immer genau zuteilbar sind, etwas schwer und umständlich.
- Durch die Zeilenzählung verschwindet der Mengeneffekt teilweise. Ebenso kann die Qualität

sowohl der einzelnen Zeilen wie auch der Äußerungen – trotz der Möglichkeit von Zusatzpunkten - nur bedingt berücksichtigt werden. Ein Vorteil ist die einfache Berechnung, weil nur der Durchschnitt bemüht wird. Man könnte auch noch eine Mindestanzahl an Zeilen/ Items/ Blöcken festlegen, was das System aber verkomplizieren würde.

Die drei Texte 2,3, und 6 erreichten jeweils die in Tafel 16 dargestellten Punkte. Zugleich werden Strategien der Studenten sichtbar, die in den Anmerkungen vermerkt sind.

Tafel 16: Textvergleich

S=Studenten, D=Punktedurchschnitt, P=Punkte EhU, A=Anmerkungen

S	D	P	A
2	0.8	80	Durch die Länge werden Schwachstellen ausgeglichen
6	0.9	90	Einige Zeilen/Abschnitte über 1.0, deshalb trotz mittlerer Länge hohe Punkte
3	1.3	92	Viele Punkte bei hoher Qualität und kleiner Quantität.

Mit der Idee, einen Noten/Punktebereich für einen Bereich an Ergebnissen freizuhalten, in dem die Abstände nicht gleich wie in dem anderen Bereich sind, haben wir den wichtigsten Schritt zur Berücksichtigung höherer Werte gemacht. Natürlich ergibt sich dadurch eine Ungerechtigkeit gegenüber den Punktabständen im „Normalbereich“, weil vergleichsweise größere Anstrengung nötig sind, um die nächsthöhere Stufe zu erreichen. Andererseits wird dadurch auch in höheren Punktbereichen eine Unterscheidung möglich. Im Endeffekt ist es eine gesellschaftliche Entscheidung, ob man gleiche Abstände haben möchte und dann im obersten Bereich alle zusammengefasst werden, oder ob man einen Bereich mit ungleich größeren Abständen und dann eben auch in höheren Punktbereichen unterscheidbaren Noten ermöglicht. Welches System vorgezogen wird, ist sicher von Gesellschaft zu Gesellschaft verschieden, und vielleicht noch von vielen anderen Faktoren abhängig, im gegenwärtigen Japan wird wohl das letztere vorgezogen.

#### 4. 4. 6. Ein gemischtes Beurteilungssystem: Punkte und Rubric

Um einerseits den Leistungen der Studenten, z.B. um ihnen im Sinne des *formative assessment* zu sagen, wie sie weiter lernen sollten (Wicking 2016), besser gerecht zu werden, andererseits aber auch um Unterschiede auch bei höheren Punktzahlen hinreichend berücksichtigen zu können, schlagen wir in diesem Abschnitt ein gemischtes Beurteilungssystem vor.

##### 4. 4. 6. 1. Das System

Dieses System besteht aus zwei Beurteilungen, die aber im Anschluss aneinander erfolgen, so

dass kein zweites Durchsehen nötig ist:

- einem Punktesystem, ganz ähnlich wie in 4.4.2., und
- einer ganzheitlichen Beurteilung nach bestimmten Kriterien (wie in 4.4.4.) mit einem neu entwickelten und hier zum ersten Mal vorgestellten Rubric.

Die meisten Inhalte dieser Beurteilungsmethode wurden in Reinelt (2017) vorgestellt, wo sich auch weitere Ausführungen zu den Hintergründen und zur Entwicklung dieses System finden.

Bei den Beurteilungen wird wie folgt vorgegangen:

1) Zuerst wird wie in Teil 4.4.2. die Punktezahl anhand der Zählung von *Wörtern*, *Phrasen*, und *Chunks* usw. ermittelt. Dabei wird das Geschriebene nach Möglichkeit auch als Teil eines Blocks usw. berücksichtigt, z.B. in einer Begrüßung. Schießlich handelt es sich einerseits um *free writing* (Tanner 2016), andererseits geht es auch darum, das Gelernte in sinnvollen Zusammenhängen darzubieten, und auch die Lernzeit, das WS, genügend zu repräsentieren. Dadurch können die Studenten produktive Lernprozesse darbieten, wenn z.B. ihre Texte über das Normale aus dem Unterricht oder im Lehrbuch hinausgehen (und sie dafür dann Sonderpunkte bekommen).

-- Exkurs --

Bei den Beurteilungen gehen wir nach den Annahmen der Grounded Theory vor (Grounded theory (Harper 2015)), indem wir Annahmen über verschiedene Stufen machen, nur dass diese nicht explizit getrennt vorgenommen, sondern mit dem Aufkommen von Items erweitert werden. Für die Beurteilung haben aber dann die folgenden vier Kriterien nach Übertragung auf das Fremdsprachenlernen in ihrer Ausprägung in dem produzierten Text große Bedeutung:

- *fit* (d.h. passen die Lexeme im Zusammenhang),
- *relevance* (sind die Teile situationsrelevant),
- *workability* (sind sie funktional einsetzbar), und
- *modifiability* (verschiedenfach anwendbar).

Jeder Textteil sollte sie erfüllen. Wie dies überprüft wird, zeigen die nachfolgenden Teile.

-- Ende Exkurs --

Hier sind einige Ausführungen zur Einzelpunktbewertung angebracht:

- Pro Wort ein Punkt beim ersten Auftreten, danach nicht mehr, wohl aber bei der Benutzung richtiger Flexionsformen (*kommen vs komme*) davon.
- Es gilt die Ein-Fehler-Regel: Bei nur einem Rechtschreibfehler gibt es noch Punkte.
- Bei Redewendungen pro Wort ein Punkt, da es auf der Anfangsstufe Redewendungen gibt, die sich nur in einem Punkt unterscheiden ((das tut mir leid (*sympathy*) vs. *es tut mir leid* (*weak apology*)). Bei der veränderten Wiederholung von *chunks* bekommt nur der veränderte Teil

einen Punkt (*Mitte August* vs. *Mitte September*).

- Kurze Phrasen erhalten 2 Punkte, egal ob Varianten davon drei oder ein Wort haben (*es geht* vs. *nicht so gut* vs. *schlecht*).
- Abzüge bei fehlenden Wörtern (*ich fahre Hiroshima*).
- Wenn schon alle Woerter anderweitig bekannt sind, gibt es einen Punkt für die Struktur oder nur für den Satz.

Dieser Teil trägt 50% zur Gesamtnote bei (s.u).

2) Unmittelbar danach geht es mit einem Rubric weiter, der dann die Bewertung abschließt.

Die Benutzung eines Rubric kanalisiert in gewisser Weise die Einschätzungen, die beim Verfahren in 4.4.4. gemacht wurden, ohne sie allzu sehr zu beschränken. Die Entwicklung von Rubrics für offene Prüfungen steht noch ganz am Anfang, und einen ersten Versuch der Entwicklung hat Rakedzon (2017) vorgelegt. Wir benutzen im folgenden eine Anwendung des von und für mündliche Prüfungen entwickelten *Score sheets* für Prüfer und Beurteiler auf geschriebene Texte (Reinelt 2017).

Tafel 17: Die Übertragung: Dieselben Kriterien benutzen

Kriterium links aus der mündlichen Prüfung entspricht rechts beim Schreiben:	
Aussprache >	Rechtschreibung
Grammatik >	Grammatik
Wortschatz >	Wortschatz
Dialogizität >	Stoffe dialogisch aufgeschrieben oder als solche verständlich
Flüssigkeit >	viele verschiedene Themen

Der Rubric sieht in der Anwendung wie in Tafel 18 aus. Dabei haben wir nur die Extreme bezeichnet, die Deskriptoren für die Zwischenstufen sind noch in Erarbeitung.

Tafel 18: Benotungsrichtungen und Kriterien

Benotungsrichtung: Positiv	>	Benotungsrichtung: Negativ
R Rechtschreibung	>	Rechtschreibung
Fehlerfrei		Viele Fehler, Englisch, Japanisch
G Grammatik: sehr viele Fehler	>	Grammatik: wenig oder keine Fehler
W Wortschatz viel, reichhaltig	>	Wortschatz: gering, Internationalismen
D Dialogizität		
In Dialogen, Stoffe im Zusammenhang	>	zusammenhanglose Teile, „Inseln“
V Flüssigkeit:		
viele Themen/Variationen	>	wenig Themen/Variationen

Für jedes Kriterium gibt es 5 bis 10 Punkte. Punkte: schlecht bis fantastisch, wie in der Schule in Deutschland. Das System kann man auch splitten: Eine Note für die textinterne Anwendung der Kriterien, eine andere Note für die Leistung relativ zu z.B. dem Unterrichtsinhalt.

Insgesamt werden bis zu 50 Punkte gegeben, d.h. ebenfalls 50% der Note.

Aus beiden, der Punktzahl in 1 und der in 2 setzt sich dann die Endnote zusammen. Der folgende Teil zeigt eine Beispielrechnung

#### 4. 4. 6. 2. Ein Beispiel

Eine Anwendung dieses Beurteilungsverfahrens auf Teil 6 demonstriert einige Vor- und Nachteile. Zuerst wird die Note für die Wortpunkte ermittelt. Die Punktezuteilung für Wörter/Phrasen findet sich in Tafel 19.

Tafel 19: Wort-und Phrasenpunkte für Text 6

2	I+C4:M44c	heiße	S	Y																
3	Ich	komme	aus	Matsuyama																
2	Ich	wohne	in	Anjojimachi																
5	Das	ist	Norden	von	Matsuyama															
4	und	noldwestl	von	Ehime	Uni															
3	in	meiner	Freizeit																	
2	ich	spiele	Futsal.																	
3	Futsal	ist	kleiner	Fussball																
6	Ich	gehe	mit	meinem	Freund	zu	spielcenter.													
2	Ich	spiele	Computerspiel.																	
3	Ich	fahre	mein	Auto																
9	Übrigend,	in	letztem	Oktober	habe	ich	den	Führersch	gemacht.											
3	dann,	ich	koche	Torte																
4	Und,	Ich	bin	jedzt	Student															
3	dann,	ich	studiere	Jura.																
5	Apropos,	meiner	Geburts	ist	13.10.	(dreizehnt	zehnte)													
5	Ich	habe	1997	(neunzehn	13.10.	(dreizehnt	geboren.													
7	Was	haben	Sie	in	letztem	Jahr	gemacht?													
5	Im	Januar	habe	ich	Centralprü	gemacht.														
3	Im	April	habe	ich	in	Ehime	Uni	gekommen.												
7	Im	August	habe	ich	vielen	Bücher	gelesen.													
1	Im	Oktober	habe	ich	den	Führersch	gemacht.													
4	Im	Novembe	habe	ich	am	Anfang	mein	Auto	gefahren.											
6	Im	Weihnach	habe	ich	meinem	Bruder	ein	Geschenk	gehenken											
4	Was	essen	Sie	gern?																
5	Ich	esse	am	liebsten	die	Wurst.														
4	Übrigend.	Ich	esse	lieber	Fisch	so	Fleisch													
1	Ich	wohne	in	Anjoji																
4	Wie	komme	ich	dahin?																
6	Von	hier	gehen	wir	nach	Westen.														
4	Da	ist	ein	Fuji.																
6	Und	dann,	von	da	gehen	wir	1km	nach	Holden.											
4	Da	links	ist	es.																
4	Jetzt	habe	ich	einen	Pullover	an.														
5	Der	ist	dunkel	blau.																
5	Wo	haben	Sie	den	gekauft?															
2	In	Shimamura.																		
1	Wann..?																			
2	An	diesem	Mai.																	
1	Warum?																			
3	Der	steht	mir																	

Der Student bekommt für diesen Text 158 Punkte. Da dies mehr als die von der Ehime Universität vorgesehenen 100 Punkte sind, verwenden wir hier das in 4.4.4. eingeführte Verfahren. Dabei erhalten alle Studenten, die 0 bis 90 Punkte haben, diese als Note für 1: So entsprechen z.B. 80 Punkten aus dem Text 80 Punkten, und da die Gesamtpunktzahl aus zwei Teilen besteht, dann davon die Hälfte: 40 Punkte für Teil 1.

Im vorliegenden Fall hat der Student aber weit mehr Punkte. Um aber auch jenseits von 90 Punkten eine Abstufung zu bekommen, benutzen wir den Logarithmus. Dieser liefert zwar kaum genaue Werte, ist aber für die Abstufung allemal ausreichend.

Dazu müssen wir eine Obergrenze festlegen, um festzustellen, welchen Logarithmus wir nehmen können.

Harper (2016) stellte fest, dass ihre Studenten an der Universität in freien Übungen zum Englischen (nach 6 Jahren!) 300 bis 400 Wörter schrieben. Zwei Studenten haben ebenso viele Wörter in ihrem Deutsch nach nur einem Jahr in ihrem Text „Ich“ geschrieben, so dass wir dies hier als Höchstgrenze annehmen können. Die 410 Punkte von 90 bis 500 projizieren wir dann mit einem Logarithmus auf den 10-Punkte-Bereich 91 bis 100. Das sieht folgendermaßen aus: Der  $\log_{1.82}$  kommt bei 410 (fast) genau auf 10. Wir können nun diese Abstände für die Berechnung der Punkte benutzen, und zwar mit den folgenden Stufen:

- 1) Der Student hat mindestens 90 Textpunkte = 90 Punkte
- 2) plus 68 über 90. Das sind bei  $\log_{1.82} = 7$  Punkte (Die Formel lautet:  $=\text{LOG} (68;1.82)$  wobei 68 die Zelle ist, in der die Wortpunktzahl steht.
- 3) Daraus ergeben sich für den Studenten für den Bereich 1 97 Punkte und damit für Teil 1 die Hälfte davon, d.h. 49, weil die Punkte etwas darüber liegen.

Im Teil 2 geht es darum, dass ein Urteil über die Gesamtarbeit relativ zu der Erfüllung bestimmter Kriterien abgegeben wird. Hierbei kann die Einschätzung nach Beurteiler weit schwanken, andererseits können auch qualitativ oder quantitative Probleme ausgeglichen werden. Die angenommenen Kriterien sind auch nicht ganz gleich in ihrer Reichweite: Rechtschreibung und Grammatik sind aus pädagogischen Gründen fast immer nur relativ zu dem zu beurteilenden Text zu beurteilen, während das Kriterium Wortschatz aufgrund der Möglichkeit des Selbstsuchens, das auch teilweise als Hausaufgabe gegeben ist, relativ zum Gelernten oder zu freien Annahmen der Beurteiler zu beurteilen ist. Die Dialogizität ist wiederum ein textinternes Phänomen, und dabei kann es auch Reihungen geben, die ebenso nur textintern zu beurteilen sind. Die Variabilität ist eigentlich relativ zur Menge der im Unterricht behandelten Themen zu beurteilen, wobei allerdings wiederum die Unendlichkeitsannahme gilt, dass man schon mit wenigen grammatischen und lexikalischen Möglichkeiten unendlich viel sprachlich durchfüh-

ren kann.

Schließlich kann natürlich das Gewicht verschoben werden, wenn man das Gewicht der Beurteilereinschätzung reduzieren möchte.

Tafel 20 enthält Vorschläge und Kommentare des Verfassers für Text 6.

Tafel 20: Kommentierter Rubric für Text 6

Rechtschreibung: 8 Ein paar Fehler stören den Lesefluss nur unwesentlich;  
Grammatik: 9 Im großen und ganzen o.k, teilweise auch gewagt;  
Wortschatz: 8 Viele Wörter, auch selbst gesuchte;  
Dialogizität: 9 Keine Inseln und es könnte so als Gespräch ablaufen;  
Vielfalt: 9 Fast alle im WS behandelten Themen sind berücksichtigt.  
Der Student erhält also für Teil 2 von Text 6 43 Punkte.

Zusammen mit den 48 Punkten aus Teil 1 ergeben die 91 Punkte ein sattes *sehr gut*.

Bei den beiden Texten 2 und 3 ergeben sich einige Unterschiede: Text 3 bekommt bei Kriterium V wegen der Kürze nur 6 Punkte. Text 2 kommt, trotz der Länge, bei Kriterium D wegen einiger Inseln nur auf 6 Punkte.

#### **4. 4. 6. 3 Nachbetrachtungen**

Zum Abschluss dieses Teil seien einige Vor- und Nachteile erwähnt.

Der größte Vor- und zugleich Nachteil dieser Beurteilungsmethode ist, dass zwei Bewertungen kurz nacheinander erfolgen (außer wenn man wirklich viel Zeit hat). Einerseits ist dies zeitsparend, andererseits bleibt das Problem der gegenseitigen Beeinflussung der Faktoren.

Ein weiterer Vorteil dieses Vorgehens ist, dass es Inseln berücksichtigen kann. Das heißt, wenn Wörter nur zusammenhanglos hingeschrieben werden, weil sie gerade eingefallen sind, führt dies zu einem Punkteabzug. Bei den bisherigen Verfahren stören Inseln die Beurteilung, weil der Beurteiler nicht weiß, wie er sie einordnen soll.

### **5. Zusammenfassung und Forschungsausblick**

Im Fremdsprachenunterricht geben im Sinne des *formative assessment* offene Tests (gegenüber normalerweise geschlossenen Tests) mit mehreren Themen den Studenten bessere Chancen, zugleich die eigene Leistung zu zeigen, wie auch für sich selbst zu sehen, wie weit man – ganz im Sinne der Optimierung (Reinelt 2017b) - schon kommen kann. In diesem Beitrag wurden mehrere Beurteilungssysteme vorgestellt, die vielleicht helfen können, Beurteilungen auch bei

solchen Arbeiten noch schneller und einfacher und dabei gerechter zu machen und zugleich einen Ausgleich zwischen DICHT und LANG zu schaffen.

Dabei sind Bewertungen bei gemischten Systemen, und wie man damit umgehen kann, noch alles anders als klar. Zumindest aber muss man bisherige Beurteilungen hinterfragen.

### Literatur

- Ainley, E. , Jenny King, Susanne Käsbauer, Robyn Cooper (2014) A framework analysis of free-text data from the neonatal survey 2014. <https://doi.org/10.1016/j.jnn.2017.09.003>.
- AntCONC 2016 (<http://www.laurenceanthony.net/software/antconc/>).
- Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. *Qual. Res. Psychol.*, 3(2), pp. 77–101.
- Harper, J. (2015) Movie Viewing Academic Writing 2015 TLT 39,4 3-7.
- Reinelt, R. (2017) The second time around: Is ""Mein Deutsch II"" necessary? The 12<sup>th</sup> Matsuyama Mini-conference on Foreign Language Learning and Teaching, Matsuyama Sept. 23.
- Reinelt, R. (2017b) Lehrerbetreuung für optimierten Unterricht I, Ehime Universität Ronsou 19 (im Erscheinen).
- Reinelt, R. (2018) Is Mein Deutsch II necessary. Matsuyama, The12thmatsu17 Compendium (im Erschienen).
- Rubric (n.d.) Rubric Development | University of West Florida <http://uwf.edu/offices/cutla/supporting-pages/rubric-development/>.
- Sara Spinellia, Caterina Dinnellaa, Camilla Masia, Gian Paolo Zobilib, John Prescottacc, Erminio Monteleonea (2017) Investigating preferred coffee consumption contexts using open-ended questions. *Food Quality and Preference* 61,63-73, <https://doi.org/10.1016/j.foodqual.2017.05.003>.
- Tanner, P. (2016) Freewriting: Don't Think Twice, It's All Write *The Language Teacher - Issue 40.3; May 2016*, 9-16, <http://jalt-publications.org/node/2/articles/5155-freewriting-don%E2%80%99t-think-twice-it%E2%80%99s-all-write>.
- Text mining (n.d.) <https://www.predictiveanalyticstoday.com/top-free-software-for-text-analysis-text-mining-text-analytics/>.
- Tzipora Rakedzon, Ayelet Baram-Tsabari (2017) To make a long story short: A rubric for assessing graduate students' academic and popular science writing skills, In: *Assessing Writing*, 32, 28-42, ISSN 1075-2935, <https://doi.org/10.1016/j.asw.2016.12.004>. (<http://www.sciencedirect.com/science/article/pii/S1075293516300873>).
- Wicking, P. (2016) The role of Formative Assessment in Global Human Resource Development. *Jalt Journal* 38, 1, 27-43.
- Wicking, P. (2017) Learning Oriented Assessment in a Testing Oriented Culture. *The Language Teacher* 41, 4, 9-13.