

Leistungsbeurteilungskompetenz von Russischlehrkräften in Deutschland, Österreich, der Schweiz und Südtirol (DACHS): Zwischen Status Quo und aktuellen Bedürfnissen

Anastasia Drackert¹ und Wolfgang Stadler²

The focus on competence-orientation in the Russian language classroom still confronts teachers with new challenges not only when teaching, but also in their evaluation of students' performances. We report on the results of a needs analysis, the aim of which was to survey the assessment practices of in-service teachers of Russian as a foreign language in Austria, Germany, Switzerland and South Tyrol as well as to discover their needs in terms of assessing student performances in Russian. The analyses of 198 questionnaires show that the needs of Russian teachers differ from those uncovered in previous research in other languages (cf. Vogt 2010, 2011) and are found in other areas of classroom assessment practice. We conclude by presenting a list of recommendations for teacher training of pre-service and in-service teachers of Russian.

1. Einleitung

Untersuchungen zu *Language Assessment Literacy (LAL)*, d.h. zu Sprachbeurteilungskompetenz, sind – v.a. im deutschsprachigen Raum – nach wie vor selten, obgleich dieser Begriff bereits vor mehr als zehn Jahren Eingang in die Testforschung fand (vgl. Jeong 2013: 346). Eine der wenigen Untersuchungen zur Leistungsbeurteilungskompetenz von Fremdsprachenlehrkräften u.a. aus Deutschland stammt von Vogt (2010), die den Begriff *LAL* als Konstrukt definiert, welches mehrere Komponenten umfasst: "solides Wissen über Mechanismen und Prozesse der Beurteilung"; "Anspruch an eigene Professionalität zur Durchführung qualitativ hochwertiger Leistungsmessung und -beurteilung" und die "Fähigkeit zur kritischen Evaluation von Tests und zur Auswahl und Erstellung angemessener Formen der Leistungsmessung und -beurteilung" (Vogt 2010: 268). Diese Definition deckt sich mit Begriffsbestimmungen z.B. von Davies (2008) und Taylor (2009), die ebenfalls Wissen (*knowledge*), Fähigkeiten und Fertigkeiten (*skills*) sowie Grundsätze (*principles*) in den Mittelpunkt ihrer Definitionen rücken (vgl. z.B. Koh 2011: 256f.; Fulcher 2012: 115; Jeong 2013: 355;

1 Korrespondenzadresse: Prof. Dr. Anastasia Drackert, Ruhr-Universität Bochum, Seminar für Slavistik / Lotman-Institut, D-44780 Bochum, E-Mail: anastasia.drackert@rub.de

2 Korrespondenzadresse: Prof. Dr. Wolfgang Stadler, Leopold-Franzens-Universität Innsbruck, Institut für Fachdidaktik / Institut für Slavistik, A-6020 Innsbruck, E-Mail: wolfgang.stadler@uibk.ac.at

Malone 2013: 331). Als erste haben Stiggins (1999) und Brindley (2001) den Begriff *assessment literacy* in der Literatur verwendet: Ersterer verstand unter *LAL* die Grundsätze einer fundierten Beurteilung (*principles of sound assessment*); letzterer sah Beurteilungskompetenz v.a. in Verbindung mit der Kenntnis des Curriculums (*curriculum-related assessment*) (vgl. Fulcher 2012: 116).

Ohne im Rahmen dieses Beitrags im Detail auf weitere Begriffsbestimmungen von *LAL* eingehen zu können, sei hier lediglich angeführt, dass *LAL* als Terminus eingesetzt wird, um zusammenzufassen, was Fremdsprachenlehrkräfte über Leistungsbeurteilung (*assessment*) wissen sollten (vgl. Boyles 2005; Stiggins 1997; Stoyhoff & Chappelle 2005; zit. nach Malone 2008: 225). Popham (2009: 4) sieht Leistungsbeurteilungskompetenz zwar als *conditio sine qua non* für die Arbeit einer Fremdsprachenlehrkraft, dennoch konstatieren Leighton et al. (2010: 8f.), die sich auf Studien in den USA, Kanada und Großbritannien berufen, dass den meisten Lehrpersonen fundierte Grundkenntnisse im Bereich Beurteilungsprinzipien und -praktiken fehlen.

Eine der wenigen Analysen (vgl. Hasselgreen, Carlsen & Helness 2004; N=914) zur Beurteilungskompetenz unterschiedlicher *Stakeholder*-Gruppen, von denen in der Untersuchung Lehrkräfte als repräsentative Gruppe allerdings nicht adäquat (N=364) vertreten waren – so Vogt (2011: 206) –, widmet sich dem *LAL*-Ausbildungsstand der Befragten, unterrichtsbezogenen und konzeptuellen Fragen der Leistungsmessung und -beurteilung sowie dem Bedarf nach Fortbildung in diesem Bereich. Vogt (2010) und Vogt & Tsagari (2014) waren unter den ersten, die *LAL* bei europäischen Fremdsprachenlehrkräften untersuchten. Sie adaptierten zu diesem Zweck den Fragebogen nach Hasselgreen et al. (2004) und versandten ihn an Lehrkräfte (N=781) in sieben europäischen Ländern. Gefragt wurde nach der Ausbildung und Lehrtätigkeit, dem Stand der Ausbildung im Bereich Leistungsbeurteilung und dem Bedarf nach Fortbildungsmaßnahmen. Für die Situation in Deutschland wurden zudem Interviews mit Fremdsprachenlehrkräften geführt (N=9, Vogt 2010). Im Folgenden sind die Ergebnisse dieser Befragung allein für Deutschland (N=113) in aller Kürze zusammengefasst:

- Ein klares *LAL*-Konzept für Deutschland fehlt (ebenso im europäischen Kontext).
- Eine Ausbildung und Vorbereitung im Bereich der Leistungsmessung und -beurteilung findet kaum statt.
- Wenn Testen behandelt wird, werden Besonderheiten der Leistungsbeurteilung im linguistischen Bereich (Morphologie, Lexik, Syntax, Orthografie) abgedeckt.
- Fortbildungsbedarf wird in neuen Formaten der Leistungsbeurteilung (z.B. Portfolio) gesehen.

- Es besteht kein Interesse an Fortbildungen im Bereich Klassenarbeiten und Notengebung.
- Fortbildungsangebote sollen auf die Bedürfnisse einzelner Lehrerinnen und Lehrer (im Folgenden LuL) zugeschnitten sein.

Die Notwendigkeit, sich mit der Beurteilung von Schülerleistungen in der Aus- und Weiterbildung zu befassen, ist auf Grund dieser Ergebnisse gegeben. So ist die dringende Auseinandersetzung mit *LAL* in verschiedenen *Stakeholder*-Gruppen nicht nur eine Konsequenz des PISA-Schocks (Harsch 2015), sondern auch auf die Erkenntnis zurückzuführen, dass das Testen und die Leistungsbeurteilung fremdsprachlicher Kompetenzen ein zentraler Aspekt nicht nur der Lehrkompetenz, sondern auch der täglichen Unterrichtspraxis ist. Eine Lehrkraft wendet einen wesentlichen Teil ihrer Arbeitszeit für die Beurteilung von Leistungen auf (Stiggins & Conklin 1992). Dabei geht es nicht nur um das Bewerten von Inhalten und Kompetenzen, sondern auch um die diagnostische Feststellung von Stärken und Schwächen einzelner Schülerinnen und Schüler (SuS) und die darauf aufbauende individuelle Förderung. Da es in der Lehrerbildung nur vereinzelt Veranstaltungen zu *LAL* gibt (Hasselgreen et al. 2004; Popham 2011; Vogt 2010), müssen fast alle LuL Schülerleistungen bewerten, ohne jemals gelernt zu haben, was gutes *assessment* ist, und nicht selten basieren ihre Bewertungspraktiken auf subjektiven Erfahrungen mit Tests aus einer Zeit, als sie selbst noch Lernende oder Studierende waren (Gunn & Gilmore 2014).

Besonders herausfordernd ist die Situation in 'kleinen' Sprachen wie Russisch. Russisch wird in den alten Bundesländern Deutschlands und in Österreich in der Regel als dritte oder vierte lebende Fremdsprache – viel seltener als zweite – gelehrt, sehr oft als Wahlfach. In den neuen Bundesländern hingegen, die etwa 70% der Russisch lernenden SuS in Deutschland auf sich vereinen, wird Russisch wesentlich öfter als zweite Fremdsprache angeboten. Ausnahmen bilden Waldorfschulen, wo Russisch auch als erste Fremdsprache gelernt wird, sowie einzelne Grundschulen. Je nach Region kann es vorkommen, dass nur eine Lehrkraft an der Schule Russisch unterrichtet. Trotz seines Rufes als 'kleines Fach' wird Russisch allein in Deutschland gegenwärtig von 111.185 (Daten des Statistischen Bundesamts für das Schuljahr 2015/2016) und in Österreich von ca. 7.600 Schülerinnen und Schülern an unterschiedlichen Schultypen (Stand 2014/15, Quelle: Tiroler Tageszeitung vom 27.9.2017) gelernt und nimmt daher als Schulfremdsprache den fünften Platz ein (vgl. Mehlhorn 2016: 536). Diese Faktoren mögen die Leistungsbeurteilungs- und Bewertungskompetenz sowie die Korrekturpraktiken von Russischlehrkräften insofern beeinflussen, als nur wenig kollegialer Austausch möglich ist.

Außerdem stehen für Russisch als Fremdsprache keine Vergleichsarbeiten zur Verfügung bzw. werden keine Lernstandserhebungen wie zum Beispiel VERA³ durchgeführt. Die in Russland entwickelten Prüfungen für Russisch als Fremdsprache (*Testirovanie russkogo jazyka kak inostrannogo; TRKI*) weichen in mancher Hinsicht von den Vorstellungen ab, wie Tests am GeR zu kalibrieren sind – ein aufwändiges Procedere, das im Manual (*Council of Europe 2009*) ausführlich beschrieben ist. Darüber hinaus gibt es für Russisch als Fremdsprache im Vergleich zu Englisch als Fremdsprache kaum Fachliteratur aus der Testforschung und nur sehr wenige *Best-practice*-Beispiele. Die fehlende Forschungsliteratur mag für Russischlehrende ein vernachlässigbares Manko sein, doch die mangelnden standardisierten Vergleichsarbeiten und Testsammlungen fördern das 'Schmoren im eigenen Saft' und die individuelle Produktion von Testaufgaben für schülergerechte Klassenarbeiten.

2. Die Studie

Die folgende Studie widmet sich der *LAL* von Russischlehrkräften, einer Gruppe, für die es nach unserem Wissen bisher keine Untersuchungen gibt. Ziel der Untersuchung ist es, Bewertungspraktiken von Russischlehrenden zu erheben und Einzelheiten über ihre speziellen Bedürfnisse hinsichtlich des Testens und Bewertens zu erfahren.

2.1 Probanden

An der Umfrage haben Russischlehrkräfte im schulischen Kontext aus Deutschland (D), Österreich (A), der Schweiz (CH) und aus Südtirol (S) (=DACHS) teilgenommen. Insgesamt haben 198 Lehrkräfte den Fragebogen ausgefüllt und retourniert,⁴ davon 109 aus Deutschland, 65 aus Österreich, 17 aus der Schweiz und 7 Lehrpersonen aus Südtirol. Der folgende Beitrag widmet sich in einzelnen Punkten vorrangig dem Vergleich der Ergebnisse zwischen den zwei größten Gruppen, den deutschen und österreichischen Russischlehrenden. Zur Orientierung werden aber auch die Gesamtergebnisse (DACHS) angegeben.

An der Studie nahmen Lehrende mit unterschiedlicher Berufserfahrung teil (vgl. Tab. 1). Die meisten Befragten ($N = 141$) unterrichten an Gymnasien.

3 Vgl. <https://www.iqb.hu-berlin.de/vera>

4 Leider wurden nicht alle 198 Fragebögen vollständig ausgefüllt. Einige Teilnehmer/innen haben die Beantwortung des Fragebogens an unterschiedlichen Stellen abgebrochen. Aus diesem Grund wird bei der Auswertung jeder Frage die exakte Zahl der Antwortenden (N) angegeben.

Tabelle 1: Unterrichtserfahrung der Befragten

	DACHS (N=172)	D (N=94)	A (N=62)
1-5 Jahre	39	18	16
6-10 Jahre	28	14	10
> 10 Jahre	38	18	17
> 20 Jahre	28	19	7
> 30 Jahre	39	25	12

Ca. 70% der befragten Lehrkräfte sind an ihrer Schule an der Erstellung, Abnahme und/oder der Korrektur der Reife- (SRDP)⁵ bzw. Abiturprüfung im Fach Russisch beteiligt, ca. 30% sind es nicht.

2.2 Datenerhebung und -auswertung

Der Fragebogen⁶ wurde nach dem Studium von *LAL*-Fragebögen in der Forschungsliteratur zu *LAL* (u.a. Plake & Impara 1993; Hasselgreen et al. 2004; Fulcher 2012; Vogt & Tsagari 2014) von den Autoren für die eigenen Zwecke adaptiert und entwickelt. Anschließend wurden die Entwürfe mit *Master-of-Education*-Studierenden an der Ruhr-Universität Bochum und mit Russisch- und anderen Fremdsprachenlehrenden in Deutschland und Österreich pilotiert. Bei einzelnen Fragen geschah dies mit dem Ansatz des 'Lauten Denkens'.

Der Fragebogen besteht aus drei Teilen. In Teil I (Fragen F1-F12) werden persönliche und berufliche Daten erhoben, u.a. die L1 der Lehrkräfte, Unterrichtsort, Schultyp, Schulstufe, in der Russisch unterrichtet wird. Außerdem wird nach der Erfahrung bezüglich Testerstellung, Beteiligung an der SRDP- bzw. Abiturprüfung sowie der Behandlung des Themas 'Testen, Prüfen und Evaluieren' in der Ausbildung gefragt. In Teil II (Fragen F13-F25), der den Kern des Fragebogens bildet, wird in offenen und geschlossenen Fragen nach Aspekten der Leistungsbeurteilungskompetenz/der Bewertungspraktiken mit einem Fokus auf *classroom assessment* gefragt. Teil III (Fragen F26-F28) erhebt die Fortbildungswünsche der LuL. Von 29 Fragen wurden zehn Fragen im offenen und 19 im geschlossenen Format gestellt. Die Kombination von offenen und geschlossenen Fragen erachten wir im Sinne einer methodeninternen Triangulation als sinnvoll. Eine Begrenzung auf ausschließlich geschlossene Fragen hätte nur strukturierte Antwortoptionen zugelassen. Durch die offenen Fragen konnten die Lehrpersonen neue Inhalte einbringen und ihre persönliche Meinung äußern.

5 Standardisierte kompetenzorientierte Reife- und Diplomprüfung

6 Der komplette Fragebogen ist auf der Webseite www.uibk.ac.at/slawistik/pdf/pdf2017/leistungsbeurteilungskompetenz_russischlehrer.pdf abrufbar.

Der Fragebogen wurde in eSurv.org digitalisiert und an Russischlehrkräfte über mehrere regionale Verbände (u.a. Verband Russisch und Mehrsprachigkeit, Russischlehrerverband NRW, Hamburger Russischlehrerverband, Verband der Russischlehrer Österreichs) sowie an dem Autor und der Autorin bekannte Russischlehrende an einzelnen Schulen verschickt. Die Datenerhebung fand zwischen November 2016 und Februar 2017 statt.

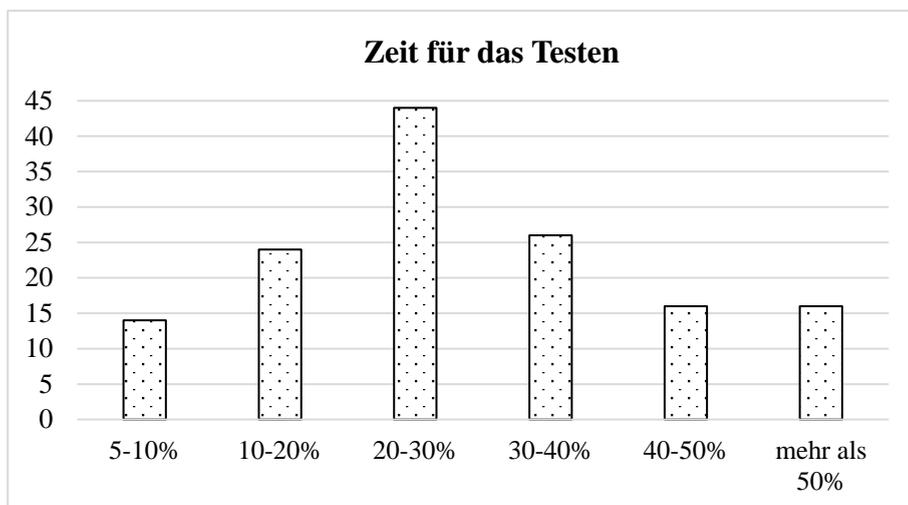
Die Angaben zu den geschlossenen Fragen wurden direkt in SPSS eingegeben und ausgewertet. Die Antworten auf die offenen Fragen gruppieren beide Autoren zunächst unabhängig voneinander nach inhaltlichen Schwerpunkten, was einer weiteren Form der Triangulation, der Forschertriangulation (vgl. Knorr & Schramm 2016: 92), gleichkam. Dadurch konnte die Reliabilität der Kategorisierung verbessert werden. Erst in der gemeinsamen Diskussion über die individuelle Gruppierung entstand ein Zusammenführen und in Folge eine Klassifikation der Antworten. Es wurde versucht, nach dem Prinzip 'so viel wie nötig, so wenig wie möglich' vorzugehen. So wurde bei der Frage nach den Zielen des Russischunterrichts entschieden, lediglich mit zwei Kategorien ('Kommunikative Kompetenz' und 'Handlungskompetenz') zu operieren. Obwohl Handlungskompetenz nicht ohne Kommunikation ausgeübt werden kann, sind nicht alle Kommunikationsaufgaben handlungsorientiert. Da dies zum Teil einen Einfluss auf die Bewertungspraktiken haben kann, wurden alle Nennungen, die die Wichtigkeit einer situationsgemessenen Sprachanwendung betonen, der Kategorie '(Vermittlung von) Handlungskompetenz' zugeordnet.

Im Anhang wird als Beispiel eine Klassifizierung der Antworten auf Frage F17 (Ein 'guter' Test: Was bedeutet das für Sie?) präsentiert, die die tatsächlichen Antworten der Russischlehrenden wiedergibt.

2.3 Ergebnisse

2.3.1 Zeitaufwand für das Testen

Um zu erfahren, wie viel Zeit Lehrkräfte der Evaluation von Schülerleistungen generell widmen, wurde gefragt, wie viel Zeit sie insgesamt – sowohl in der Schule als auch zu Hause – mit der Planung von Beurteilungsmaßnahmen für den Unterricht, der Entwicklung von Testaufgaben, dem Korrigieren von Testaufgaben und dem Besprechen von Testergebnissen verbringen. Ein Drittel der befragten Lehrkräfte (44 von $N=140$) gibt an, etwa 20-30% der Zeit dem Testen zu widmen. Abb. 1 zeigt, dass es – auf- und absteigend – mehr oder weniger gleich große Gruppen von Russischlehrenden gibt, die einerseits sehr wenig Zeit (5-10%) und andererseits sehr viel Zeit (mehr als 50%) für das Testen aufwenden.

Abbildung 1: Zeitaufwand für testbezogene Tätigkeiten $N=140$

2.3.2 Testen in der Ausbildung

Um das Hintergrundwissen von Russischlehrer/innen im Bereich des Testens besser einschätzen zu können, wurde erhoben, ob diese in ihrer universitären Ausbildung oder während des Referendariats in Deutschland bzw. des Unterrichtspraktikums in Österreich Kurse belegen konnten, in denen sie etwas über Leistungsbeurteilung gelernt haben (F11).

Insgesamt 62% der Probanden bejahen diese Frage (vgl. Tab. 2). In Deutschland sind es 63%, in Österreich 58% der Befragten. 38% aller Befragten absolvierten allerdings keine(n) Kurs(e) im Bereich 'Testen und Bewerten'. Berücksichtigt man auch die Zeit der Ausbildung, so wird zudem ersichtlich (Abb. 2), dass in den letzten fünf Jahren das Thema 'Testen' deutlich öfter als früher einen Platz in der Ausbildung einnimmt.

Tabelle 2: Testen in der Ausbildung in %

	DACHS ($N=171$)	D ($N=92$)	A ($N=62$)
Ja	62	63	58
Nein	38	37	42

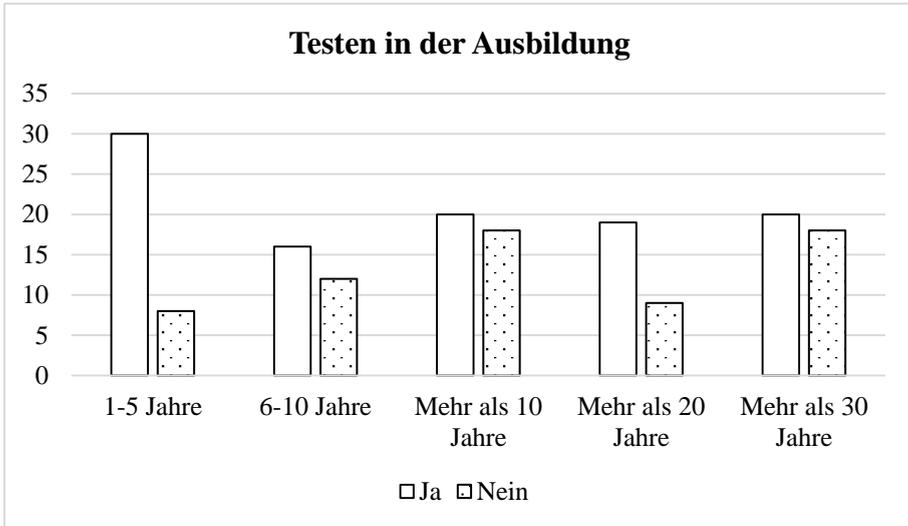


Abbildung 2: Testen in der Ausbildung nach Unterrichtserfahrung $N=171$

Was die Inhalte dieser Lehrveranstaltungen bzw. Kurse betrifft (F12) (vgl. Tab. 3), so nannten die Russischlehrer/innen in DACHS vorwiegend das 'Korrigieren von Fehlern' (30,3%) und das 'Erstellen von Aufgaben' (28%). In Österreich scheint Textkorrektur zudem ein wichtigeres Thema gewesen zu sein als Aufgabenerstellung. Leistungsbeurteilung in der universitären Lehrerbildung wurde und wird retrospektiv immer noch vorrangig von Fehler- und Textkorrektur dominiert (51,8%) und weniger mit Aufgabenerstellung und dem Verstehen von Lernprozessen assoziiert.

Tabelle 3: Kursinhalte in %

	DACHS ($N=100$)	D ($N=55$)	A ($N=34$)
Fehlerkorrektur	30,3	30,7	29,9
Aufgabenerstellung	28,0	29,3	23,9
Textkorrektur	21,6	18,6	31,3
Verstehen von Lernprozessen	20,3	21,4	14,9

2.3.3 Vorgaben für die Leistungsbeurteilung

Im schulischen Kontext gibt es Vorgaben, die die Leistungsbeurteilung der LuL regeln. Ob die Lehrkräfte diese Dokumente kennen bzw. wahrnehmen, wurde in der offenen Frage F20 gefragt. Die Antworten sind in Tab. 4 zusammengefasst.

Tabelle 4: Vorgaben für die Leistungsbeurteilung in %

Vorgaben	DACHS (N = 122)	D (N = 69)	A (N = 41)
Gesetze oder interne Vereinbarungen	41,6	42,5	48,9
Nein	22,6	20,0	17,8
Ja (ohne Angabe)	9,5	11,3	4,4
Übernahme aus anderen Fächern	8,8	11,3	4,4
Sonstiges	8,8	10,0	6,7
SRDP/Abitur	8,0	5,0	15,6
Lehrbuch	0,7	0	2,2

Weniger als die Hälfte der Russischlehrkräfte sowohl in DACHS gesamt (41,6%) als auch in Deutschland (42,5%) und Österreich (48,9%) betrachtet Gesetze und interne Vereinbarungen als eine gewisse 'Bindung'. 22,6% der Befragten geben an, keine Vorgaben als Rahmenbedingungen zu haben. In Deutschland negieren 20% der Lehrkräfte bestehende Vorgaben, 9,5% bestätigen diese als gegeben. 17,8% der österreichischen Russischlehrenden führen an, es gebe in ihrem Lehr-/Lernkontext keine Vorgaben, die ihre Beurteilungspraktiken beeinflussten; 4,4% sagen, dass es diese sehr wohl gibt. Die SRDP oder das Abitur als 'Vorgabe' für Beurteilungspraktiken nennen österreichische Russischlehrkräfte dreimal so oft wie ihre deutschen Kolleginnen und Kollegen. So gut wie keine Rolle spielt in diesem Zusammenhang das Lehrbuch.

2.3.4 Ziele des Russischunterrichts

Beurteilungspraktiken im Unterricht sollten von den angestrebten Lehr-/Lernzielen abhängen. Deshalb wurden die Russischlehrkräfte nach den Hauptzielen ihres Unterrichts gefragt (F13). Insgesamt machten 156 Befragte 414 Angaben zu dieser offenen Frage. Die Angaben wurden nach dem in Abschnitt 2.2 erläuterten Verfahren in 12 Kategorien eingeteilt und dementsprechend kodiert (vgl. Tab. 5).

Wie man Tabelle 5 entnehmen kann, gehören 'Kommunikative Kompetenz' (17,6%) sowie 'Sprachwissensvermittlung und Grundlagen' (15,9%) zu den am häufigsten genannten Zielen des Russischunterrichts, betrachtet man die Ergebnisse der Befragten gesamt. An dritter und vierter Stelle liegen in ähnlicher Häufigkeit 'Mehrsprachigkeit' (12,3%) und 'Fertigkeiten nach GeR' (11,6%). Es folgen 'Landes- und Kulturkunde' (9,7%), 'Spaß am Russischlernen' (8,5%) sowie die 'Vermittlung positiver Einstellungen und das Interesse an der russischen Sprache und Kultur' (7,2%).

Interessanterweise unterscheidet sich die Priorisierung der Ziele vom Gesamtbild, wenn man die Ergebnisse nach Ländern getrennt betrachtet. Während sowohl in Deutschland als auch in Österreich 'Kommunikative Kompetenz' als

oberstes Lernziel am häufigsten genannt wird, unterscheidet sich der zweite Platz: In Deutschland liegt 'Mehrsprachigkeit' an 2. Stelle (15,9%), während diese in Österreich nur Platz 7 (8,1%) belegt. 'Spaß am Russischlernen' wiederum liegt in Deutschland an 7. Stelle (6,9%); in Österreich ist es das viertwichtigste Ziel (11,1%).

Tabelle 5: Ziele des Russischunterrichts in %

Vorgaben	DACHS (N=156)	DE (N=84)	A (N=55)
Kommunikative Kompetenz	17,6	18,0	16,3
Sprachwissensvermittlung und Grundlagen	15,9	15,5	14,8
Mehrsprachigkeit	12,3	15,9	8,1
Fertigkeiten nach GeR	11,6	12,5	12,6
Landes- und Kulturkunde	9,7	8,2	8,9
Spaß am Russischlernen	8,5	6,9	11,1
Positive Einstellungen zur und Interesse an der russischen Sprache und Kultur	7,2	6,4	9,6
Methodenkompetenzen	5,3	6,9	2,2
Handlungskompetenz	5,1	5,2	4,4
Gate-keeping-Funktion	4,3	3,0	7,4
Schüleraustausch	1,4	1,3	2,2
Sonstiges	1,0	0,5	2,2

2.3.5 Ein 'guter' Test

In F17 wurden die Russischlehrenden nach ihren subjektiven Sichtweisen hinsichtlich eines 'guten' Tests gefragt. Diese offene Frage konnten sie in Stichworten beantworten. Den meisten Befragten in DACHS (19,3% aller Antworten) zufolge ist eine Leistungsüberprüfung 'gut', wenn der Test 'dem Unterricht entspricht', mit anderen Worten, wenn er inhaltsvalide ist und das Kompetenzniveau der Lernergruppe abbildet (vgl. Tab. 6). Inhaltsvalidität bzw. Angemessenheit der Niveaustufe der Lernergruppe ist somit die meistgenannte Antwort in DACHS, doch auch andere Kriterien sind – den befragten Russischlehrkräften zufolge – wichtig für einen 'guten' Test. So ist ein Test 'gut', wenn er z.B. 'unterschiedliche Kompetenzbereiche' berücksichtigt (11,6%) oder 'leistungsdifferenziert' (11,2%) ist. Letzteres bedeutet, dass der Test zwar generell "gut ausfallen soll", aber auch "mehrere Aufgaben steigender Schwierigkeit aufweisen soll, damit leistungsschwächere SuS aufhören können, falls ihnen die Aufgaben zu schwer werden" (vgl. auch die Zitate im Anhang). Mit anderen Worten, den befragten Russischlehrenden zufolge sollen sowohl leistungsschwache als auch leistungsstarke Lernende ihre Möglichkeiten ausschöpfen können, nur dann kann ein Test als 'gut' definiert werden.

An vierter Stelle der genannten subjektiven Vorstellungen von 'guten' Tests steht Transparenz (10%). Diese wird je nach Land von mehr (10,7% in Deutschland) oder von weniger Befragten (7,1% in Österreich) genannt. Unterschiedlich bewertet wird auch die Konstruktvalidität eines Tests. Der Begriff 'Konstruktvalidität' wurde nicht von den Lehrkräften erwähnt, sondern er wurde von den Autoren benutzt, um folgende Antworten zusammenzufassen: "Ein 'guter' Test zeigt effektive/wirkliche Kompetenzen der SuS"; "der Inhalt sollte nicht vorhersehbar sein (z.B. vorbereitete Texte für eine Schreibaufgabe auswendig lernen)"; "er testet das, was er testen soll (z.B. nur Hörverständnis)". In Österreich ist Konstruktvalidität das zweithäufigste Merkmal eines guten Tests, in Deutschland hingegen ist sie kaum relevant: Nur 5% nennen diese Kategorie.

Eine wichtige Erkenntnis aus der Umfrage unter den DACHS-Russischlehrpersonen ist, dass nur zwei von 134 Lehrkräften in vier Ländern das Kriterium 'Objektivität' für einen 'guten' Test anführen. Erstaunlicherweise entfallen auch nur 2,8% der Antworten auf das Kriterium 'Nützliches Feedback an Schülerinnen und Schüler'.

Tabelle 6: Kriterien eines 'guten' Tests aus der Sicht der Befragten

Gütekriterien	DACHS (N=134)	DE (N=73)	A (N=47)
entspricht dem Unterricht und der Niveaustufe der Lerngruppe	19,3	20,7	22,6
berücksichtigt unterschiedliche Kompetenzbereiche	11,6	12,1	9,5
ist leistungsdifferenziert	11,2	12,2	10,8
ist transparent (Kriterien, Formate, Punkteverteilung)	10,0	10,7	7,1
wendet unterschiedliche Aufgabenformate an	9,6	10,7	7,1
ist konstruktvalid	9,2	5,0	11,9
ist praktikabel	9,2	10,0	9,5
gibt Rückmeldung für den Unterricht	7,6	6,4	9,5
ist spannend und interessant für die SuS	7,6	9,3	6,0
ermöglicht nützliches Feedback an die SuS	2,8	2,1	2,4
ist objektiv	0,8	0	2,4
ist angemessen gestaltet	0,8	0,7	1,2

2.3.6 Zweck von Evaluierung im Russischunterricht

Die Frage, zu welchem Zweck Russischlehrkräfte im Unterricht ihre SuS evaluieren, wurde zweimal gestellt: in einer offenen (F15) und einer geschlossenen Frage (F21).

Der am häufigsten genannte Grund für die Evaluierung in F15 (vgl. Tab. 7) ist die erforderliche 'Leistungs- und Lernzielkontrolle' mit 19% der Antworten, gefolgt von 'Rückmeldung zum eigenen Unterricht' (16,7%) und 'Wissensüberprüfung' (15,8%). Sowohl bei 'Leistungs- und Lernzielkontrolle' als auch bei 'Wissensüberprüfung' handelt es sich um eine punktuelle Überprüfung von Kompetenzen und von grammatikalischem bzw. lexikalischem Wissen – allerdings ohne explizite Nennung, zu welchem Zweck die Evaluierung durchgeführt wird.

Die Priorisierung der Ziele von Leistungsbeurteilungsmaßnahmen in Deutschland und Österreich unterscheidet sich vom Gesamtbild DACHS: In Deutschland wird 'Wissensüberprüfung' häufiger als Zweck der Evaluierung genannt bzw. als 'Rückmeldung zum eigenen Unterricht' (20,6% vs. 15,3%). Das wichtigste Ziel in Österreich dagegen ist die 'Rückmeldung zum eigenen Unterricht' (17,2%), gefolgt von 'Wissensüberprüfung' (ca. 11,8%). 'Leistungs- und Lernzielkontrolle' liegt in Österreich mit 10% der Nennungen knapp unter 'Notenvergabe' und 'Disziplinierung und Aktivierung der Lerner' (11%).

Tabelle 7: Ziele des Testens im Russischunterricht in %

Zweck von Evaluierung	DACHS (N = 138)	D (N = 74)	A (N = 49)
Leistungs- und Lernzielkontrolle	19,0	22,9	10,0
Rückmeldung zum eigenen Unterricht	16,7	15,3	17,3
Wissensüberprüfung	15,8	20,6	11,8
Notenvergabe	8,7	5,3	10,9
Überprüfung der Lernfortschritte	7,7	7,6	9,1
Rückmeldung zu individueller Entwicklung	6,8	7,1	7,3
Wiederholung und Festigung des Gelernten	5,8	2,9	8,2
Disziplinierung und Aktivierung der Lerner	5,5	2,9	10,9
Selbstevaluation durch die Schüler/innen	5,1	7,1	3,6
Vorgesehen im Lehrplan	3,5	1,8	5,5
Vorbereitung auf Abitur oder SRDP	2,6	2,9	2,7
Erhöhung der Motivation	2,3	2,4	2,7
Schulinterne Evaluation	0,6	1,2	0

Interessanterweise scheinen Zwecke der Evaluierung, die zu einem Mehrwert eines Beurteilungsverfahrens führen, wie 'Überprüfung der Lernfortschritte (7,7%)', 'Rückmeldung zu individueller Entwicklung (6,8%)' und 'Selbstevaluation durch die Schüler/innen (5,1%)' eine nachrangige Rolle zu spielen. Werden allerdings die Nennungen zu formativen Aspekten der Evaluation addiert, ergibt sich insgesamt ein Wert von 19,6%. Dieser Wert käme in der Priorisierung der Ziele fast gleichauf zu liegen mit 'Leistungs- und Lernzielkontrolle (19,0%)'.

Neben der offenen Frage konnten die Befragten auch im geschlossenen Format auf einer Skala von 1 ('überhaupt nicht wichtig') bis 6 ('sehr wichtig') vorgegebene

Gründe für das Ziel einer schulischen Leistungsüberprüfung einstufen (vgl. Tab. 8). Die Tabelle zeigt die Ergebnisse in DACHS insgesamt, nach Deutschland und Österreich getrennt und führt jeweils auch den Mittelwert (M) und die Standardabweichung (SD) an.

Tabelle 8: Zielsetzungen für schriftliche / mündliche (Mitarbeits-)Überprüfungen

Ich überprüfe schulische Leistungen im Russischunterricht, ...	DACHS (N=128) M (SD)	D (N=71) M (SD)	A (N=43) M (SD)
damit Lernende sehen, was sie schon alles können und wo sie noch Probleme haben	5,3 (1,3)	5,2 (1,3)	5,3 (1,4)
um Lernenden eine individuelle Rückmeldung zu ihren Stärken und Schwächen geben zu können	5,3 (1,4)	5,2 (1,4)	5,3 (1,5)
um mir ein Bild machen zu können, ob meine Lernenden das Ziel erreicht haben	5,1 (1,4)	5,0 (1,4)	5,1 (1,6)
damit Lernende über ihren eigenen Lernprozess und ihre Fortschritte reflektieren	5,1 (1,3)	5,0 (1,4)	5,1 (1,4)
um gezielt auf Fehler eingehen zu können	4,7 (1,4)	4,6 (1,4)	4,6 (1,5)
weil die SuS dann besser und mehr lernen	3,8 (1,4)	3,8 (1,3)	3,7 (1,5)
weil Mitarbeitsüberprüfungen kontinuierlich vorzunehmen sind	3,5 (1,6)	3,5 (1,6)	3,4 (1,6)
um Eltern spezifische Hinweise zu den Leistungen ihrer Kinder geben zu können	3,4 (1,6)	3,4 (1,7)	3,6 (1,4)
weil die Lernenden dies von mir erwarten	2,9 (1,5)	3,2 (1,5)	2,5 (1,5)
um Disziplin in die Lerngruppe zu bringen	2,6 (1,6)	2,7 (1,6)	2,6 (1,6)

Die Zielsetzung 'damit Lernende sehen, was sie schon alles können und wo sie noch Probleme haben', hat insgesamt den höchsten Mittelwert (5,3) und wird zwischen einem 'wichtigen' und 'sehr wichtigen' Ziel einer Leistungsüberprüfung eingestuft. Generell zeigen die Ergebnisse, dass jene Kriterien das Ranking anführen, die eng mit einer diagnostischen und formativen Beurteilung, also mit *assessment for learning* einhergehen (z.B. 'um Lernenden eine individuelle Rückmeldung zu ihren Stärken und Schwächen geben zu können' ($M=5,3$) oder 'damit Lernende ihren eigenen Lernprozess und ihre Fortschritte reflektieren' ($M=5,1$). Ähnlich wie in der offenen Frage erscheint Leistungskontrolle ('um mir ein Bild machen zu können, ob meine Lernenden das Ziel erreicht haben') als wichtiges Ziel der Evaluation im Russischunterricht ($M=5,1$). Leistungsüberprüfung zur Steuerung der Disziplin im Klassenzimmer hat für die Befragten in Deutschland und Österreich gleichermaßen wenig Bedeutung. Die Mittelwerte liegen bei 2,7 in Deutschland bzw. bei 2,6 in Österreich.

2.3.7 Einschätzung von Komponenten der Beurteilungskompetenz

Von zwölf Komponenten der Leistungsbeurteilungskompetenz (F23), die die Lehrkräfte auf einer Skala von 1 ('überhaupt nicht wichtig') bis 6 ('sehr wichtig') bewerten sollten, erachten die Befragten insgesamt fünf Komponenten als besonders wichtig für ihre Tätigkeit (vgl. Tab. 9). Der höchste Wert entfällt auf die Teilkompetenz 'Beurteilung transparent und nachvollziehbar machen' ($M=5,6$), gefolgt von 'Lernende fair beurteilen können' ($M=5,4$); 'Auf der Basis von Testergebnissen Feedback an die Lernenden geben können' ($M=5,3$); 'Auswirkungen des Unterrichts einschätzen können' ($M=5,2$); 'die Stärken und Schwächen der Lernenden kennen' ($M=5,2$). Auch im Ländervergleich haben diese fünf Teilkompetenzen in jedem Land einen Wert über 5. Im Vergleich Deutschland – Österreich zeigt sich, dass die österreichischen Russischlehrkräfte es für die Leistungsbeurteilungskompetenz als wichtiger erachten, den GeR und das Europäische Sprachenportfolio (ESP) einsetzen zu können ($M=4,1$) als ihre deutschen Kolleginnen und Kollegen ($M=3,5$). Den geringsten Mittelwert ($M=3,3$ in Deutschland; $M=3,4$ in Österreich) in der Liste der Leistungsbeurteilungskomponenten weist der Punkt 'keinen Fehler übersehen' auf.

Tabelle 9: Einschätzung von Komponenten der Beurteilungskompetenz

Komponenten der Leistungsbeurteilungskompetenz	DACHS ($N=126$) M (SD)	D ($N=71$) M (SD)	A ($N=42$) M (SD)
Beurteilung transparent und nachvollziehbar machen	5,6 (1,1)	5,5 (1,1)	5,5 (1,3)
Lernende fair beurteilen können	5,4 (1,3)	5,4 (1,3)	5,5 (1,4)
Auf der Basis von Testergebnissen Feedback an die Lernenden geben können	5,3 (1,1)	5,3 (1,1)	5,2 (1,2)
Auswirkungen des Unterrichts einschätzen können	5,2 (1,2)	5,2 (1,1)	5,0 (1,4)
Die Stärken und Schwächen der Lernenden kennen	5,2 (1,3)	5,1 (1,2)	5,2 (1,5)
Selber Testaufgaben erstellen können	4,9 (1,4)	4,8 (1,5)	4,9 (1,4)
Noten geben können	4,8 (1,4)	4,9 (1,5)	4,7 (1,4)
Die Fehler von Lernenden verbessern können	4,8 (1,3)	4,7 (1,3)	4,7 (1,5)
Auswirkungen der verwendeten Beurteilungsverfahren einschätzen können	4,5 (1,5)	4,5 (1,4)	4,4 (1,7)
Informelle Beurteilungsmethoden verwenden können	3,7 (1,5)	3,7 (1,5)	3,8 (1,6)
GeR und ESP einsetzen können	3,7 (1,4)	3,5 (1,3)	4,1 (1,3)
Keinen Fehler übersehen	3,3 (1,6)	3,3 (1,6)	3,4 (1,6)

2.3.8 Häufigkeit der Überprüfung von Fertigkeiten und Kompetenzen und Verwendung von Testformaten

Verschiedene Fertigkeiten und Kompetenzen nach GeR konnten auf einer 4-Punkte-Likert-Skala hinsichtlich der Häufigkeit der Überprüfung von 'nie' (=1) bis 'oft' (=4) eingestuft werden. Die produktiven Fertigkeiten Sprechen ($M=3,7$) und Schreiben ($M=3,6$) werden lt. Angaben am häufigsten überprüft, gefolgt von den linguistischen Kompetenzen Lexik ($M=3,6$) und Grammatik ($M=3,5$). Sprachmittlung, interkulturelle Kompetenz, Hör-/Sehverstehen sowie soziopragmatische Kompetenz haben die niedrigsten Mittelwerte ($M=2,6$ bzw. $M=2,3$), d.h. diese Kompetenzen werden nur 'selten' bis 'manchmal' überprüft.

Aus insgesamt 24 in F24 angegebenen Formaten verwenden 92,8% der Russischlehrkräfte das halboffene Format 'Fragen zum Inhalt eines Textes in Stichwörtern beantworten'; 7,2% verwenden es nicht. Spitzenreiter unter den geschlossenen Testformaten sind Mehrfachwahlantworten (*Multiple Choice*) (82,4%) und Zuordnungsaufgaben (80,8%). Rollenspiel und Vokabeltests liegen gleichauf in der Verwendung (79,2% der Russischlehrkräfte setzen diese Formate im Unterricht ein); nur die Hälfte der Befragten verwendet Diktate als Testformat, ein Drittel setzt Portfolios ein. Etwas mehr als ein Viertel verwendet *Editing*-Aufgaben des Typs 'Finde das überflüssige Wort'.

2.3.9 Erstellung von Klausuren

Tab. 10 zeigt, dass sowohl in Deutschland als auch in Österreich ca. ein Viertel der gegebenen Antworten auf die Option 'Ich erstelle die Aufgaben allein' entfallen. Nur in 11% der Fälle werden die Aufgaben in einem Team mit Kolleginnen und Kollegen erstellt, in Österreich werden 17% der Klausuren kollektiv ausgearbeitet. Ca. 11% der Antworten entfallen auf die Möglichkeit 'Ich kopiere Aufgaben aus Büchern bzw. Testsammlungen'.

Tabelle 10: Vorgehen bei der Erstellung von Klausuren in %

Erstellung von Klausuren	DACHS (N=125)	D (N=70)	A (N=42)
Ich erstelle die Aufgaben allein.	25,9	26,6	23,8
Ich übernehme/adaptiere Aufgaben aus anderen Lehrwerken oder dem Lehrerhandbuch.	22	22,8	19,7
Ich verwende Aufgaben aus meinen alten Tests.	16,4	15,6	16,3
Ich erstelle die Aufgaben mit Kolleg/inn/en.	12,4	11	17
Ich übernehme/adaptiere Aufgaben aus dem Internet.	11,9	12,7	12,2
Ich kopiere Aufgaben aus Büchern (z.B. aus Testsammlungen).	11,4	11,4	11

2.3.10 Fortbildungswünsche im Bereich des Testens und Bewertens

100 Befragte gaben 227 Antworten auf die offene F26, welche Themen sich die Russischlehrkräfte für eine Fort- bzw. Weiterbildung wünschen (vgl. Tab. 11). Das am meisten gewünschte Fortbildungsthema (25,6% aller Antworten) ist 'Erstellung und Bewertung von Aufgaben zu den vier Fertigkeiten'. Als zweithäufigstes Thema wünschen sie sich auf die Bedürfnisse ihres Unterrichts zugeschnittene Themen (17,6%). Ca. 10% aller Antworten betreffen eine Vorbereitung auf neue Abiturformate. Nur von sehr wenigen Lehrkräften werden Fortbildungswünsche zu den Themen *Online Testing*, Erstellen guter Klassen-/Schularbeiten, Diagnostisches Testen und Einstufungstests für Herkunftssprecher/innen geäußert. Der größte Unterschied zwischen den zwei Ländern Deutschland und Österreich besteht darin, dass in Deutschland 12% der Fortbildungswünsche das Thema Sprachmittlung betraf; in Österreich sind es hingegen 0%. Für die 'Beurteilung mündlicher Sprachproduktion' interessieren sich in Österreich doppelt so viele Lehrkräfte wie in Deutschland. Die Erstellung differenzierter Tests interessiert keine/n der Befragten in Österreich, aber immerhin 4,5% der deutschen Lehrkräfte.

Tabelle 11: Wünsche bezüglich Fortbildungen in %

Fortbildungswünsche	DACHS (N=100)	D (N=58)	A (N=32)
Erstellung und Beurteilung von Aufgaben zu den vier Fertigkeiten	25,6	27,6	24,4
Mit dem Unterricht verbundene Themen	17,6	15,7	21,8
Vorbereitung auf die neuen Abiturformate	9,7	8,2	11,6
Beurteilung mündlicher Produktion	9,3	6,1	12,8
Beurteilung schriftlicher Produktion	7,5	7,5	6,4
Beurteilung von Sprachmittlung	7,5	11,9	0
alternative Formen des Testens	5,2	6,0	5,1
kompetenzorientiertes Testen	4,0	3,7	3,8
Sonstiges	4,0	3,0	6,4
Erstellung differenzierter Tests	2,6	4,5	0
Beurteilung interkultureller und soziolinguistischer Kompetenzen	2,6	2,2	3,8
<i>Online Testing</i>	1,8	0,7	1,3
Erstellen von guten Klassenarbeiten	1,3	0,7	2,6
diagnostisches Testen	0,9	1,5	0
Einstufungstests für Herkunftssprecher	0,4	0,7	0

Unter den Fortbildungsthemen, die in der geschlossenen F27 angeboten wurden und die in allen Ländern als wichtig für die Tätigkeit einer Russischlehrkraft erachtet werden, finden die Themen 'Entwicklung und Beurteilung von Klassenarbeiten' ($M=4,9$); 'Entwicklung und Beurteilung von Abiturprüfungen' ($M=4,8$)

sowie 'Arbeit mit Beurteilungsrastern für Sprechen, Schreiben und Sprachmittlung' den größten Zuspruch. Weniger wichtig scheinen der Einsatz von *Self-Assessment* ($M=3,6$) und *Peer-Assessment* ($M=3,5$) im Russischunterricht zu sein.

Beinahe ein Viertel der Russischlehrenden könnte ihrer Einschätzung nach bei einer Fortbildungsveranstaltung selbst das Thema 'Erstellen von Abituraufgaben' kompetent anbieten (F28). 20% würden fachlich kompetent eine Veranstaltung zu 'Offenem Lernen' bzw. zu 'Arbeit mit Medien' anbieten können; nur 10% trauen sich eine Fortbildung zu konkreter Textarbeit zu. 30% der Lehrkräfte, die diese Frage beantwortet haben, gaben allerdings an, kein Thema nennen zu können.

3. Zusammenfassung und Interpretation ausgewählter Ergebnisse

Die Ergebnisse der Untersuchung zeigen deutlich, dass Testen und Bewerten ein wichtiger Bestandteil des Lehrerralltags ist, und sie belegen die Angaben von Stiggins & Conklin (1992) in der Fachliteratur. Obwohl die Russischlehrkräfte ihrer Selbsteinschätzung nach ca. 25% der Zeit für das Testen und Bewerten aufwenden, weisen beinahe 40% aller Befragten keine Ausbildung in diesem Bereich auf, auch wenn sich die Ausbildungssituation in den letzten fünf Jahren diesbezüglich verbessert hat. Dennoch unterstreichen die gewonnenen Ergebnisse die Feststellung in der Forschungsliteratur, dass *LAL* bisher eine zu geringe Rolle in der Ausbildung spielt (vgl. z.B. Popham 2009, Vogt 2010). Dieser Bereich ist dringend in die universitäre, curriculare Pflichtlehre sowie in Fort- und Weiterbildungsveranstaltungen von Russischlehrkräften aufzunehmen, zumal 70% der befragten deutschen und österreichischen Russischlehrenden in irgendeiner Form an der Erstellung, Durchführung und Auswertung von Abituraufgaben beteiligt sind. In Form von weiterführenden Interviews mit den Lehrkräften müsste in qualitativen Untersuchungen geklärt werden, ob bzw. wie sie auf diese kompetenzorientierten Aufgaben vorbereitet wurden bzw. werden.

Hatten die Russischlehrkräfte während ihres Studiums eine Ausbildung im Bereich Leistungsbeurteilung, so ist davon auszugehen, dass diese in erster Linie 'linguistische Aspekte' betraf und die Korrektur von Fehlern und Texten einschloss. Das mag ein Grund sein, warum die Gruppe jener Russischlehrpersonen, deren Ausbildung mehr als 20 Jahre zurückliegt, der Meinung ist, eine Ausbildung in Testen und Bewerten gehabt zu haben. Damals wurde – zumindest in Österreich – in der universitären Lehrerausbildung Leistungsbeurteilungskompetenz größtenteils mit Fehlerkorrektur (*error analysis*) gleichgesetzt. Jene Gruppe, die län-

ger als 30 Jahre unterrichtet, hatte u.U. nicht einmal diese Ausbildung. Individuelle exemplarische Befragungen könnten klären, ob das Fehlen einer solchen Ausbildung sprachspezifisch bedingt ist.

Etwas überraschend war die Erkenntnis, dass 25,4% der befragten Lehrkräfte meinen, keine Vorgaben zur Leistungsbeurteilung im Russischunterricht zu haben. Selbstverständlich gibt es Dokumente dieser Art, z.B. die Hinweise zur Korrektur von Klausuren im Land NRW⁷ oder die Leistungsbeurteilungsverordnung (LBVO)⁸ in Österreich. Es mag sein, dass die Lehrkräfte, deren SuS kein Abitur bzw. keine Reifeprüfung in Russisch ablegen, diese Rahmenbedingungen tatsächlich nicht kennen – diese Prozentzahl stimmt ungefähr mit der Prozentzahl jener Befragten, die keine Verantwortung für das Abitur in Russisch tragen, überein – oder sie ignorieren diese Vorgaben. Da diese Gesetze bzw. Handreichungen von verschiedenen Ministerien und Landes-/Stadtschulratsbehörden einerseits als Unterstützung für die LuL, andererseits aus Transparenzgründen entwickelt werden, sollten sie auf jeden Fall zur Kenntnis genommen bzw. bei Weiterbildungen besprochen werden.

Ausgehend von der bei vielen Fortbildungsveranstaltungen von Russischlehrkräften immer wieder geäußerten Meinung, dass schulischer Russischunterricht – schon allein auf Grund der morphologischen Besonderheiten der Sprache – 'besonders schwierig' und somit auch 'anders sei', 'sein müsse' als 'herkömmlicher' Fremdsprachenunterricht (in Englisch, Italienisch, Französisch oder Spanisch), ist ein Ergebnis doch begrüßenswert: Mehr als 15 Jahre nach dem Erscheinen der deutschen Ausgabe des GeR (Europarat 2001) wird auch von Russischlehrenden 'kommunikative Kompetenz' als oberstes Lehr-/Lernziel des Unterrichts genannt. Die früher nicht selten postulierte Forderung, Russischunterricht müsse grammatik- und lexikorientiert sein, sowie die damals häufig geäußerte Skepsis gegenüber der Kompetenz- und Handlungsorientierung, scheinen inzwischen der Akzeptanz von kompetenzorientiertem Unterricht gewichen zu sein; dennoch wird Sprachwissensvermittlung von den LuL nach wie vor als zweitwichtigstes Ziel im Russischunterricht genannt. Diese beiden Ziele, kommunikative Kompetenz und Wissen über Sprache, finden sich auch in den Antworten auf jene Frage, welche Fertigkeiten am häufigsten getestet werden: Auch hier liegen die kommunikativen Aktivitäten Sprechen und Schreiben knapp vor den linguistischen Kompetenzbereichen Lexik und Grammatik. Es lässt sich demnach eine gewisse Übereinstimmung zwischen den gesetzten Lernzielen und der Bewertung von Fertigkeiten und Kompetenzen im Russischunterricht

7 Vgl. <https://www.standardsicherung.schulministerium.nrw.de/cms/zentralabitur-gost/faecher/getfile.php?file=3777>

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

konstatieren. Den Daten nach dürften LuL kommunikative Kompetenz v.a. als aktive Beherrschung der Sprache in Wort und Schrift auslegen. Die Fertigkeiten Lesen und v.a. Hörverstehen werden im Russischunterricht weniger oft getestet, obwohl Hörverstehen im alltäglichen Leben als primäre rezeptive Fertigkeit die höchste Relevanz hat (vgl. Wagner & Werry 2015).

Sollen Kriterien eines guten Tests genannt werden, so wird – wie für *classroom assessment* nicht anders erwartet – auch von Russischlehrenden dem Kriterium der Inhaltsvalidität in den Antworten Genüge getan. Sie geben u.a. an, dass der Test nicht nur den Unterrichtsinhalten, sondern auch dem Niveau der konkreten Lerngruppe entsprechen soll. Das könnte bedeuten, dass trotz der Implementierung von kompetenzorientierten Lehrplänen, die die inhaltsorientierten Curricula ablösen, im Russischunterricht eher norm- als kriterienorientiert bewertet wird. Das gilt sowohl für die Zusammenstellung als auch die Korrektur der Tests. Eine zu starke Ausrichtung auf die Leistungsstärke oder -schwäche einer spezifischen Gruppe kann aber u.U. verhindern, dass das in den (Kern-)Lehrplänen angegebene Kompetenzniveau bzw. die geforderte Transparenz und Vergleichbarkeit von Kriterien erreicht werden.

Ein in der Testliteratur nicht erwähntes Phänomen bezüglich der beschriebenen Problematik, das aber sehr wohl in den Antworten der Lehrkräfte auf Frage F17 vorkommt, ist Leistungsdifferenzierung. Jedem zehnten Russischlehrenden zufolge sollen auch schwache Schüler/innen, selbst wenn sie das angestrebte Leistungsniveau nicht oder nur knapp erreichen, den Test bestehen (vgl. 2.3.5). Erst dann wäre der Test als 'guter' Test einzustufen. Eine 'flexible' Vorstellung von individuellen *cut scores* mag den Kriterien eines *dynamic assessment* entgegenkommen, wäre aber in der Welt der *high-stakes tests* (Dlaska & Krekeler 2009) undenkbar. Für die Schule ist – schon aus Gründen der Fairness – zu bedenken, ob die Note der Lernenden nicht doch kriterienorientiert und weniger normorientiert definiert sein soll. Die Note des einzelnen sollte nicht von der Gruppe abhängig sein. Gute Beispiele leistungsdifferenzierter Bewertung gibt es bereits in Form konkreter Vorschläge für die Praxis (vgl. Behr 2016; Wapenhans 2016). Die von Behr und Wapenhans vorgeschlagene Bepunktung von Leistungen orientiert sich nicht an einer sozialen Bezugsnorm, sondern an klar definierten Sachkriterien hinsichtlich einer freien bzw. unterstützten Aufgabenlösung. Durch die Inanspruchnahme von Hilfen hätten leistungsschwächere Schüler zumindest die Möglichkeit, bestimmte Aufgaben zu lösen, was aber durch einen Punkteabzug sanktioniert würde.

Laut der Befragung erstellen die Lehrkräfte die Klausuren in ca. einem Viertel der Fälle allein, nur in 12% der Fälle werden die Aufgaben gemeinsam mit Kolleginnen und Kollegen erstellt. Das kann einerseits daran liegen, dass an manchen Schulen nur eine Person Russisch unterrichtet, andererseits aber auch

daran, dass – wie oben ausgeführt – eine gute Klausur auf den eigenen Unterricht und die eigene Lerngruppe ausgerichtet wird. Eine gemeinsame Erstellung von Testspezifikationen und in Folge von Klausuren würde jedoch nicht nur eine potentielle Verbesserung von konkreten Testinstrumenten bedeuten, sondern auch zur Vergleichbarkeit des Gelernten über Klassen oder Lerngruppen hinweg beitragen (vgl. Stadler & Kremmel 2018: 53).

Ausgehend von den Antworten auf die offene Frage F15 werden Prüfungen, Tests und Evaluationen im Russischunterricht in den meisten Fällen zur Leistungs- bzw. Lernzielkontrolle eingesetzt. Die am zweithäufigsten genannte Antwort betrifft die Rückmeldung an die Lehrkraft zum eigenen Unterricht. Formative Aspekte der Evaluation spielen zudem eine wichtige Rolle, da die individuelle Rückmeldung an die SuS hinsichtlich Lernstärken und -schwächen als Zweck einer Überprüfung ganz weit oben steht.

Ein wesentlicher Unterschied zu der eingangs erwähnten Untersuchung von Vogt (2010) liegt darin, dass die von uns befragten Russischlehrkräfte in DACHS das Fortbildungsthema 'Entwicklung und Beurteilung von Klassen-/Schularbeiten' als sehr wichtig erachten. Vogts Untersuchung (2010: 276) hat jedoch gezeigt, dass Englischlehrkräfte in Deutschland kein Interesse an Themen aus dem Bereich Leistungsmessung und -beurteilung haben, die mit Klassenarbeiten und Notengebung zu tun haben. Das könnte dadurch erklärt werden, dass es im Englischen viele internationale Tests (*TOEFL*, *IELTS* usw.) gibt, in denen Beispiele guter Testaufgaben zu finden sind. Im Russischen ist dies nicht der Fall, was ein Grund dafür sein kann, dass auch das Thema 'Bewerten der vier Fertigkeiten' einen hohen Stellenwert unter den befragten Russischlehrkräften einnimmt. Als ebenso wichtig werden weitere 'klassische' Themen wie Vorbereitung auf neue Abiturformate, Bewertung mündlicher/schriftlicher Produktion und Sprachmittlung angesehen. Letzteres betrifft fast ausschließlich die Fortbildungswünsche der Lehrkräfte in Deutschland.

Aus den Ergebnissen der Befragung lassen sich folgende Empfehlungen für die Aus- und Weiterbildung von Russischlehrkräften formulieren:

1. obligatorische Verankerung des Testens in der Aus- und Weiterbildung
2. Behandlung des Themas 'Diagnostisches (formatives) Testen'
3. Sensibilisierung für objektives und faires Bewerten
4. Klärung der Wechselwirkung zwischen Inhalts- und Konstruktvalidität im Bereich des schulischen Testens
5. Kennenlernen der (gesetzlichen) Rahmenbedingungen
6. gemeinsames Erstellen von Klausuren (z.B. bei Fortbildungen)
7. gleiche Gewichtung der Kompetenzen im Unterricht und in Klausuren

4. Grenzen der Studie und Ausblick

200 Russischlehrende mittels eines Fragebogens zu befragen, ist ein schnelles und effizientes Instrument der Datenerhebung. Die Ergebnisse solch einer quantitativen Befragung, die immerhin zehn offene Fragen enthielt, müssen jedoch in Folge durch qualitative Methoden wie Lehrerinterviews, Unterrichtsbeobachtung oder Dokumentenanalyse genauer hinterfragt werden. Die offenen Fragen lassen zwar Raum für individuelle Meinungen der einzelnen Probandinnen und Probanden, doch können sie auch dazu führen, dass Antworten unvollständig bleiben, zu wenig ausführlich formuliert oder aus zeitlichen Gründen überhaupt nicht gegeben werden. Die methodeninterne Triangulation führte in manchen Fällen zu einer bewussten Verdoppelung von Fragen, wodurch sich die Länge des Fragebogens erhöhte, was manche Lehrpersonen zu einem vorzeitigen Abbruch bewegen haben mag. Zu bedenken ist auch die Möglichkeit von gewissen Antworttendenzen im Sinne der 'sozialen Erwünschtheit', die verzerrte oder bewusst positive Antworten auslöst, um gewissen Erwartungen oder Vorstellungen zu entsprechen.

Abschließend sei eine Liste von Ergebnissen angeführt, die noch einer näheren Analyse bedürfen. So nennen beinahe 30% der Russischlehrkräfte das 'Erstellen von Aufgaben' und 20% das 'Verstehen von Lernprozessen' als Inhalte ihrer universitären Ausbildung, mit der sie im Bereich des Testens und Bewertens konfrontiert waren (F12). Hier müsste konkret nachgefragt werden, um welche Art von Aufgaben es sich handelte bzw. müsste genauer untersucht werden, ob der Zeitpunkt der Ausbildung einen Einfluss auf die Auswahl der Aufgabenformate hatte. Die Nennung 'Verstehen von Lernprozessen' wird zwar nur von einem Fünftel der Lehrpersonen unter den universitären Inhalten angeführt. Es wäre aber ein anzustrebendes und lohnendes Ziel in der Aus- und Weiterbildung von Russischlehrkräften, das realisiert werden sollte. In der Forschungsliteratur äußern Leighton et al. (2010) mit Verweis auf Stiggins (1991) sowie Black und William (1998) allerdings große Zweifel daran, ob Sprachstandstests an der Schule mit dem Ziel erstellt werden können, etwas über die Lernprozesse bei SuS zu erfahren: "In fact, there is little evidence to suggest that existing classroom assessments are developed to inform students' learning processes" (ebd.: 17). Dennoch nennen die befragten Russischlehrkräfte die Reflexion der Lernprozesse und des damit verbundenen Fortschritts immerhin als viertwichtigstes Ziel (F21) der schulischen Leistungsüberprüfung. Ob es auf Seiten der Lehrkräfte für das Verstehen dieser Prozesse auch konkrete Belege gibt, müsste ebenfalls Teil weiterer Untersuchungen sein.

Zu klären wären auch gewisse bei der Auswertung aufgetretene Diskrepanzen. Bei der offenen Frage – Was bedeutet ein 'guter' Test für Sie? (F17) – haben nur

wenige LuL Objektivität als ein wichtiges Kriterium genannt. In der geschlossenen Frage lagen Transparenz und Fairness an erster Stelle. Heißt das, dass dieses Kriterium so selbstverständlich ist, dass es die LuL bei der offenen Frage überhaupt nicht erwähnen, oder ist es vielmehr so, dass im Vergleich mit den anderen genannten Kriterien Objektivität in der Praxis doch nicht so wichtig erscheint?

Des Weiteren konnten Fragen oder bestimmte Termini unterschiedlich aufgefasst worden sein. Selbst das Wort 'Test' in 'Ein 'guter' Test: Was bedeutet das für Sie?' wurde unterschiedlich ausgelegt: Bedeutet 'Test' Vokabeltest, Klausur oder jede Form einer schriftlichen oder mündlichen Leistungsüberprüfung? Der Fragebogen wurde zwar von zwei Forschern aus Deutschland und Österreich erstellt, dennoch unterscheiden sich gesetzliche Rahmenbedingungen und Terminologie in den beiden Ländern, sodass geklärt werden müsste, ob manche Ergebnisse auf Grund von Missverständnissen zustande gekommen sind.

Bei der Frage nach Gründen für die Evaluierung im Unterricht scheint formatives *assessment* unter Russischlehrkräften in Deutschland und Österreich einen hohen Stellenwert einzunehmen, wenn es generell um den Zweck von Evaluierungen (F15) oder um die Zielsetzung schriftlicher /mündlicher / (Mitarbeits-) Überprüfungen (F21) geht. 'Diagnostisches Testen' rangiert aber unter den Fortbildungswünschen der Russischlehrenden ganz weit unten. Bedeutet das, dass die Befragten sich zutrauen, Lernerleistungen diagnostisch zu bewerten, oder finden sie andere Themen viel wichtiger? Weiterer Nachfragebedarf zeigt sich auch hinsichtlich des Verhältnisses zwischen den Antworten auf die Frage nach den am häufigsten getesteten Kompetenzen (F22) und den in den Tests verwendeten Aufgabenformaten (F24). Laut den Antworten werden die Fertigkeiten Sprechen und Schreiben von den Befragten am häufigsten überprüft. Die verwendeten Testformate spiegeln dieses Ziel aber nicht wider: So ist unter den sechs Top-Aufgabenformaten, die 80-90% der Ja-Nennungen erhalten, nur ein einziges offenes Testformat, das Sprechen überprüfen könnte – nämlich das Rollenspiel. Lässt sich daraus schließen, dass Sprechen und Schreiben 'gelehrt' und im Unterricht bzw. für Hausaufgaben häufig eingesetzt, aber nicht 'getestet' werden?

Auch wenn aus der Untersuchung neue Fragen entstanden sind, können aus der vorliegenden Befragung – der ersten dieser Art mit Russischlehrkräften – wichtige Erkenntnisse gewonnen werden. So ist erfreulich, dass viele Russischlehrende einen kompetenz- und handlungsorientierten, kommunikativen Unterricht bevorzugen und den Fragen der Leistungsbeurteilung zunehmend abgeschlossen gegenüberstehen. Ob alle gewonnenen Erkenntnisse rein sprach- bzw. fachspezifisch sind, d.h. nur Russischlehrende oder auch LuL anderer Fremdsprachen betreffen, sollte in weiteren Untersuchungen geklärt werden.

Eingang des revidierten Manuskripts 19.12.2017

Literaturverzeichnis

- Behr, Ursula (2016): Eine Klassenarbeit mit Wahlmöglichkeiten. Differenzierte Aufgaben und Bewertung. *PRAXIS Fremdsprachenunterricht Russisch* 1/2016, 7-9.
- Black, Paul & William, Dylan (1998): Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice* 5(1), 7-74.
- Brindley, Geoff (2001): Language assessment and professional development. In: Elder, Catherine; Brown, Annie; Grove, Elisabeth; Hill, Kathryn; Iwashita, Noriko; Lumley, Tom; McNamara, Tim F. & O'Loughlin, Kieran J. (eds.): *Experimenting with uncertainty: Essays in honour of Alan Davies*. Cambridge: Cambridge University Press, 126-136.
- Dlaska, Andrea & Krekler, Christian (2009): *Sprachtests. Leistungsbeurteilungen im Fremdsprachenunterricht evaluieren und verbessern*. Baltmannsweiler: Schneider Hohengehren.
- Council of Europe (2009): *Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR). A Manual*. Strasbourg, Language Policy Division [Online: www.coe.int/lang 11.11.2017].
- Davies, Alan (2008): Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347.
- Fulcher, Glenn (2012): Assessment literacy for the language classroom. *Language Assessment Quarterly* 9, 113-132.
- Gunn, Alexandra C. & Gilmore, Alison (2014): Early childhood initial teacher education students' learning about assessment. *Assessment Matters* 7, 24-38.
- Harsch, Claudia (2015): *Assessment Literacy – Trend oder Notwendigkeit?* In: Böcker, Jessica & Stauch, Anette (Hrsg.): *Konzepte aus der Sprachlehrforschung – Impulse für die Praxis. Festschrift für Karin Kleppin*. Frankfurt a.M.: Peter Lang Edition, 489-509.
- Hasselgreen, Angela; Carlsen, Cecile & Helness, Hildegunn (2004): *European survey of language testing and assessment needs. Part one: General findings*. [Online: <http://www.ealta.eu.org/resources.htm> 19.08.2017].
- Jeong, Heejeong (2013): Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing* 30(3), 345-362.
- Koh, Kim Hong (2011): Improving teachers' assessment literacy through professional development. *Teaching Education* 22 (3), 255-276.
- Knorr, Petra & Schramm, Karen (2016): Triangulation. In: Caspari, Daniela; Klippel, Friederike; Legutke, Michael & Schramm, Karen (Hrsg.): *Forschungsmethoden in der Fremdsprachendidaktik. Ein Handbuch*. Tübingen: Narr Francke Attempto, 90-97.
- Leighton, Jacqueline P.; Gokiert, Rebecca J.; Cor, Ken M. & Heffernan, Colleen (2010): Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: implications for assessment literacy. *Assessment in Education: Principles, Policy & Practice* 17(1), 7-21.
- Malone, Margaret E. (2008): Training in language assessment. In: Shohamy, Elana & Hornberger, Nancy H. (Eds.): *Encyclopedia of Language and Education, Vol. 7: Language Testing and Assessment*. Springer Science + Media Business LLC, 225-239.
- Malone, Margaret E. (2013): The essentials of assessment literacy: contrasts between testers and users. *Language Testing* 30 (3), 329-344.
- Mehlhorn, Grit (2016): Russisch. In: Burwitz-Melzer, Eva; Mehlhorn, Grit; Riemer, Claudia; Bausch, Karl-Richard & Krumm, Hans-Jürgen (Hrsg.): *Handbuch Fremdsprachenunterricht*. 6., Völlig überarbeitete und erweiterte Auflage. Tübingen: A. Francke, 534-539.
- Plake, Barbara & Impara, James C. (1993): Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice* 12 (4), 10-12.

- Popham, James W. (2009): Assessment literacy for teachers: faddish or fundamental? *Theory into Practice* 48 (1), 4-11.
- Popham, James W. (2011): Assessment literacy overlooked: A teacher educator's confession. In: *The Teacher Educator* 46 (4), 265-273.
- Relating Language Examinations to the Common European Framework of Reference for Languages. A Manual* [Online: https://www.coe.int/t/dg4/linguistic/Source/ManualRevision-proofread-FINAL_en.pdf 19.10.2017].
- Stadler, Wolfgang & Kremmel, Benjamin (2018): Testprinzipien. In: Stadler, Wolfgang & Hinger, Barbara (Hrsg.): *Testen und Bewerten fremdsprachlicher Kompetenzen. Eine Einführung*. Tübingen: Narr Francke Attempto, 39-55.
- Stiggins, Richard (1991): Assessment literacy. *The Phi Delta Kappa* 72, 534-539.
- Stiggins, Richard & Conklin, Nancy (1992): *In teachers' hands: Investigating the practice of classroom assessment*. Albany, NY: SUNY Press.
- Taylor, Lynda (2009): Communicating the theory, practice and principles of language testing to test stakeholders: some reflections. *Language Testing* 30 (3), 403-412.
- Tiroler Tageszeitung (2017): Spanisch und Russisch sind im Trend – Englisch ist der Dauerbrenner. Nummer 267-ST, 27.9.2017, 20.
- Vogt, Karin (2010): Assessment Literacy bei Fremdsprachenlehrkräften – Bedarf und Bedürfnisse in Aus- und Weiterbildung. In: Altmayer, Claus; Mehlhorn, Grit; Neveling, Christiane, Schramm, Karen & Schlüter, Norbert (Hrsg.): *Grenzen überschreiten: sprachlich – fachlich – kulturell. Dokumentation zum 23. Kongress für Fremdsprachendidaktik der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF) Leipzig, 30. September – 3. Oktober 2009*. Baltmannsweiler: Schneider, 267-277.
- Vogt, Karin (2011): Leistungsmessung und -beurteilung – Kompetenzen europäischer Fremdsprachenlehrkräfte. *Deutsch als Fremdsprache* 48 (4), 205-211.
- Vogt, Karin & Tsagari, Dina (2014): Assessment Literacy of Foreign Language Teachers: Findings of a European Survey. *Language Assessment Quarterly* 11 (4), 374-402.
- Wagner, Erik & Werry, Hanno (2015): Kompetenzorientierter Fremdsprachenunterricht: Neue Formen der Leistungsmessung in Klassenarbeiten und ihre Wirkung auf den Unterricht. In: Böcker, Jessica & Stauch, Anette (Hrsg.): *Konzepte aus der Sprachlehrforschung – Impulse für die Praxis. Festschrift für Karin Kleppin*. Frankfurt a.M.: Peter Lang, 569-592.
- Wapenhans, Heike (2016): Rasskaži, što ... – einen Monolog im 2. Lernjahr bewerten. *PRAXIS Fremdsprachenunterricht Russisch* (1), 10-11.

Anhang: Kodierung der Antworten auf F17: Ein 'guter' Test: Was bedeutet das für Sie?

1) Dem Unterricht (Inhaltsvalidität) und dem Niveau entsprechend (Normorientierung)

Abprüfen von Inhalten, die in der vorhergehenden Unterrichtseinheit vermittelt worden sind; Spiegelt den Inhalt des Gelernten wider; faires Niveau; Ein guter Test sollte sich auf den Leistungsstand der Gruppe beziehen; Unterricht und Test müssen übereinstimmen (z.B. wenn ich kommunikativ unterrichte, muss auch das Prüfen danach ausgerichtet sein); umfasst das gelernte Stoffgebiet

2) Berücksichtigung unterschiedlicher Kompetenzbereiche

Alle Teilbereiche (Hören, Grammatik, Textproduktion und Lesen) sind immer vorhanden; fragt unterschiedliche Kompetenzen ab

3) Anwendung unterschiedlicher Aufgabenformate

SuS müssen unterschiedliche Aufgaben lösen, z.B. geschlossene Aufgaben (Einsetzübungen), halboffene Aufgabe (stark gelenkte Sprachproduktion) und offene Aufgaben (individuelle Sprachproduktion zu bestimmten Themen)

4) Transparenz (Kriterien, Formate, Punkteverteilung)

Transparente Beurteilungskriterien; hat ein für den Schüler vertrautes Format hinsichtlich der Aufgaben; hat klare Vorgaben und die SuS wissen, was sie erwartet (*Testspecs*); Punktvergabe offensichtlich; klare Punkteverteilung

5) Konstruktvalidität

Zeigt effektive/wirkliche Kompetenzen der SuS; Inhalt sollte nicht vorhersehbar sein (z.B. vorbereitete Texte auswendig lernen für die Schreibaufgabe); testet das, was er testen soll (z.B. nur Hörverständnis); Überprüfung eines bestimmten Schwerpunktes (nicht alle Kompetenzen gleichzeitig!)

6) Nützliches Feedback an die SuS

Er gibt den SuS eine Rückmeldung darüber, was sie gut können und was sie noch besser machen können; Schülern zu verdeutlichen, wo ihre Stärken und Schwächen im interkulturellen, fachlichen und sprachlichen Bereich sind

7) Leistungsdifferenzierend

sollte schwierigere und leichtere Items haben; hat steigende Komplexität, damit schwächere SuS früher aufhören können; fordernd und lösbar; Sowohl schwache als auch starke SuS können ihre Möglichkeiten ausschöpfen; wenige 5, einige 1, Großteil der SuS konnte den Test gut bewältigen, SuS arbeiten selbstständig und bestehen diesen; mehr als die Hälfte gute Noten (allerdings Utopie)

8) Objektivität

9) Praktikabilität

Zeitlich machbar; sollte zeitlich gut lösbar sein; zeitlich begrenzt, jedoch eher ein Zeitrahmen – genug Zeit auch für langsamere SuS; einfach zu korrigieren, schnell zu korrigieren, schnell erstell- und auswertbar

10) Rückmeldung für den Unterricht

zeigen, wo noch Übungsbedarf besteht; sollte ein Instrument zur Justierung des Unterrichts sein, zeigen, wo Schwächen der Lernenden sind, an denen im weiteren Unterrichtsgeschehen gearbeitet werden muss; gibt mir die Möglichkeit, den Wissensstand für die weitere Arbeit im Unterricht zu ermitteln

11) Spannend und interessant für SuS

Wenn alle SuS mit großem Interesse mitmachen; wenn ein Lächeln auf dem Gesicht des Schülers aufleuchtet und er Spaß am Test hat

12) Angemessene Gestaltung

übersichtlich, d.h. nicht zu viel / zu kleine Schrift auf einer Seite; 'schülerfreundlich' erstellt; ansprechend gestaltet (Layout, Grafiken/Bilder)