

Der RR-Sprechtest und Test-DaF I

Rudolf Reinelt

Überblick

Einleitung: Hintergrund, Ziel und Struktur der Arbeit

Vorwort.

1. Der RR-Sprechtest: Hintergrund, Entwicklung und jetziger Zustand

1. 1. Hintergrund

1. 2. Entwicklung

1. 3. Jetziger Zustand

2. TestAS und TestDaF

3. Vergleichspunkte

3. 1. Theoretische Kriterien

3. 1. 1. Drei wichtige Gütekriterien und die Testerstellung:

3. 1. 1. 1. Objektivität

3. 1. 1. 2. Reliabilität

3. 1. 1. 3. Validität

3. 1. 1. 4. Testerstellung: Qualitätssicherung und –standards

3. 1. 2. Weitere wichtige Kriterien

3. 2. Praxis: Durchführung

3. 3. Beurteilung

3. 4. Beispiel Sprechen

4. Ergebnisse und eventuelle Schwierigkeiten

5. Gefahren, Verbesserungen und weitere zukünftige Entwicklungen

Verwendete Abkürzungen TD= TestDaF, RS= RR-Sprechtest

Einleitung

In modernen Gesellschaften steht am Ende bestimmter Stufen des Fremdsprachenlernens (FL) oft ein Test. Für die Lerner in den Deutsch-Anfängerkursen im allgemeinbildenden Bereich der Ehime Universität in Matsuyama hat der Autor einen Test für die Fertigkeit Sprechen (hier RR-Sprechtest genannt und im weiteren mit RS abgekürzt) entwickelt. Die anderen Fertigkeiten werden separat geprüft.

RS wurde aus den behandelten Unterrichtsinhalten entwickelt, ist aber nicht einfach mit anderen in Japan durchgeführten Sprechtests (wenn diese Fertigkeit denn überhaupt getestet wurde) vergleichbar, weil die Lerner darin spontan sprechen und die Inhalte teilweise offen sind (s. Teil 2). Andererseits ist es für die Lerner wie auch für die Testentwicklung selbst sinnvoll zu vergleichen, wie ein Test gegenüber anderen Tests mit vergleichbaren Testgegenständen aufgebaut ist und welche Kriterien erfüllt werden oder nicht.

Am besten eignen sich für einen solchen Vergleich natürlich Tests, die schon *per definitionem* fast perfekt sein müssen: High-Stakes Tests, in denen es um die weitere Lebensgestaltung der Testteilnehmer geht, wie z.B. TestDaF, im folgenden TD abgekürzt, das die deutschsprachliche Eignung der Teilnehmer für einen Zugang zu deutschen Hochschulen abtestet. Dieser Test kann, mit seiner ganzen Entwicklung und der Sorgfalt der Durchführung, als einer der am weitesten entwickelten angesehen werden.

Das Ziel dieser Arbeit ist demnach, aus dem Vergleich mit TD für die Verbesserung von RS relevante Punkte zu gewinnen. Deshalb wird im folgenden so vorgegangen: Nach dieser Einleitung und einem Vorwort zur Gesamtsituation wird in Teil eins kurz die Entwicklung von RS, d.h. Hintergründe, Entstehung, und jetziger Zustand, vorgestellt. Teil 2 geht kurz auf wichtige Eigenschaften von TestDaF und Test As ein. Teil drei vergleicht dann RS mit Punkten, die von TD als wichtig angesehen und in dessen öffentlicher Präsentation erwähnt werden, und zwar in theoretischer (3.1) wie

auch praktischer Hinsicht (3.2). Teil vier diskutiert einige Ergebnisse und eventuelle Schwierigkeiten an und Teil 5 weist auf Gefahren, Verbesserungen und weitere zukünftige Entwicklungen hin.

In dieser Arbeit wird durchgehend eine vergleichende Perspektive eingenommen. Die Vergleiche können starke Ähnlichkeiten wie auch extreme Kontraste aufdecken, aber dies ist gerade der Sinn und das Ziel derartiger Gegenüberstellungen. Die Vergleichbarkeit selbst ergibt sich aus der Position der beiden Tests an einer wichtigen Übergangsstelle im Leben der Lerner, die allerdings konträre Merkmale aufweist: TD steht idealerweise am Anfang eines weiteren Studiums, bei dem Deutsch gebraucht wird, während RS am Ende des Deutschunterrichts eingesetzt wird, nach dem die Lerner wahrscheinlich nie wieder im Leben mit dieser Fremdsprache (FS) zu tun haben werden. In beiden Fällen zeigen die Lerner aber, was sie bis dahin erreicht haben.

Der Verfasser hat in Detailuntersuchungen viele Teile des RS unter verschiedenen theoretischen Voraussetzungen betrachtet, so dass hier auf eine Nennung aller Arbeiten verzichtet werden kann. Die neueste Zusammenstellung findet sich in Reinelt (2013a)

Vorwort

Sprachenlehren und –lernen und Testen hatten schon immer ein merkwürdiges Verhältnis: Zum einen war es in der alten Lateinschule ganz selbstverständlich, dass an bestimmten Stellen bzw. zu bestimmten Zeiten getestet wurde, und schon aus diesem Grunde dafür gelernt wurde (und nach dem letzten Kontakt auch noch schneller vergessen wurde). Personen, die von einem Privatlehrer eine Sprache unterrichtet bekamen, hatten durch diese Art der Sprachvermittlung laufend Zielsprachenkontakt, wenn nicht gerade der Grammatikübersetzungsansatz gewählt wurde. Ansonsten fand Fremdsprachenlernen im natürlichen Zielsprachenkontakt statt, und dabei waren Tests nicht nötig; Wenn jemand in der Fremdsprache erfolgreich war, war dies Beweis genug.

Mit der Zunahme der Verschulung hat sich dies geändert. Zum einen mussten oft Fremdsprachen gelernt werden, die für die Lerner keine irgendwie einsehbare Relevanz für ihr Leben hatten, so z.B. Latein in Deutschland und lange aber auch Deutsch in Japan. An den Universitäten in Japan konnte die hohe Belegung nur durch die Scheinverpflichtung und die sich daraus ergebende Notwendigkeit des Testens aufrechterhalten werden. Dementsprechend brach der zweite Fremdsprachenunterricht, insbesondere der Deutschunterricht, nach der Freigabe (schon seit 1996 nicht mehr Universitätsabschlussanforderung) in der Reform Anfangs dieses Jahrtausends weitgehend weg. Damit entfiel aber auch die Möglichkeit, im Rahmen des normalen Unterrichts ein anderes Land als den anglophonen Bereich vorzustellen.

Andererseits ist es unter den neuen Bedingungen erst recht sinnvoll wenn nicht gar eine Verpflichtung gegenüber denen, die freiwillig zum Unterricht kommen, diesen immer weiter, ja sogar so weit wie eben möglich, zu entwickeln. Eines der Ergebnisse ist der Unterricht beim Autor und der daraus entwickelte RS.

1 . Der RR-Sprechtest: Hintergrund, Entwicklung und jetziger Zustand

1 . 1 . Hintergrund

Am Anfang der Deutschkurse werden die Studenten nach ihren Wünschen befragt, und in den Antworten taucht immer wieder das Wort Konversation (*kaiwa*) auf, alleine oder in Verbindungen mit anderen Ausdrücken. Da es in Japan keine ministeriellen Vorgaben für einen zweiten Fremdsprachenunterricht im allgemeinbildenden Bereich an Universitäten in Japan gibt, hat der Verfasser seinen Kurs nach diesem Studentenwunsch ausgerichtet.

Dies war an sich keine große Umstellung, da schon seit längerer Zeit einige Lehrer an der Ehime Universität Sprechelemente in ihren Unterricht aufgenommen hatten und der Verfasser schon früher eine Sprechprüfung (Kursleiter und ein Student) durchgeführt hat.

Der Verfasser hat dabei in seinem dem Test vorausgehenden Unterricht Abstand von der Grammatik-Übersetzungsmethode genommen. Durch ein an der Einführung von *Chunks*, kurzen Gesprächsteilen, orientiertes Vorgehen mit (zeitlich) kurzen (etwa je 10-20 Minuten langen), abwechselnden, aber oft wiederholten, Inhaltsteilen konnten die Inhalte schließlich so erweitert werden, dass die Lerner in den ersten Semestern GER A1 oder sogar A2 (GER 2013)-Bereiche erreichen.

In einem solchen Fall ist es natürlich unerlässlich, die erreichten Fähigkeiten der Lerner zu überprüfen. Hier soll auch noch der motivierende Aspekt solcher Tests nicht unerwähnt bleiben. Nicht nur haben sich Studenten beschwert, wenn (als ein Versuch) keine Tests durchgeführt wurden. Es ist in einem Land wie Japan, in dem im Schulkontext Prüfungen mehrmals im Jahr etwas ganz Normales sind, so, dass der motivierende Aspekt vielleicht sogar die abschreckenden Aspekte von Prüfungen überragt.

1. 2. Entwicklung

Vor diesem Hintergrund hat der Verfasser an der Ehime Universität nach und nach einen Deutschkurs entwickelt, der alle vier Fertigkeiten berücksichtigt und an dessen Ende dann auch ein Sprechtest steht.

Schon von Anfang an in den ersten 80er Jahren wurde versucht im Unterricht mit den Kollegen im damals noch zwei Jahre dauernden Deutschunterricht alle vier Fertigkeiten zu berücksichtigen. Besonders Sprechen war damals - und ist es teilweise bis heute - ein Stiefkind. Andere Fertigkeiten lassen sich aber auslagern und den Studenten zum Selbstlernen überlassen.

Später standardisierte der Verfasser mit weit mehr als 2000 Studenten eine mündliche Prüfung von etwa zwei bis drei Minuten Länge mit ihm selbst und jeweils einem Studenten. Auch hier war das Thema spontan, d.h. nicht voraussehbar, aber natürlich auf das bis dahin Gelernte beschränkt.

Drei Faktoren führten zum nächsten Schritt:

- Das Ziel des FLs ist gerade mit einem anderen Benutzer der Zielsprache als dem Lehrer zu sprechen.
- Bei der Durchführung der Sprechprüfung mit über 2000 Studenten wurde der Verfasser müde, so dass Gleichheit nicht mehr garantiert war.
- Der Kontakt mit Austauschstudenten aus Deutschland (jeweils einer an der Ehime und an der Matsuyama Universität) verbesserte sich soweit, dass nach deren Mitarbeit gefragt werden konnte.

Nach ihrer Zusicherung der Mitarbeit entwickelte der Verfasser nach der Sichtung von international anerkannten Kriterienlisten ein *Scoreboard* mit den Kriterien Aussprache, Grammatik, Vokabular, Flüssigkeit und Dialogizität, zu beurteilen nach dem vertrauten deutschen System von 1 bis 6.

Zuerst wurden die Austauschstudenten als Beurteiler eingesetzt. 2007 wurde dann der erste Versuch mit einem Muttersprachler als Sprechpartner im Sprechtest am Ende des zweiten Semesters gestartet. Diese Konstellation wurde nach gleich beim ersten Mal erfolgreicher Durchführung bis jetzt durchgehalten. Dabei können sowohl Sprechpartner wie auch Beurteiler an der Ehime Universität in Matsuyama sein, wie auch über Skype irgendwo in der Welt. Alle Sprechpartner beurteilen alle Dyaden, also auch die, in denen sie nicht Gesprächspartner waren.

In der letzten Zeit wurden Versuche mit Studenten als Rater (Reinelt 2013b) und Sprechpartner (Reinelt 2013c) gestartet, die diese Prüfung abgelegt und weiter Deutsch beibehalten haben.

1. 3. Jetziger Zustand

Der Test wird am Ende des ersten Semesters (Juli) und am Ende des zweiten Semesters (Februar) nach 13 bzw 27 Kontaktstunden (=einmal (bzw. mit Partner zweimal) 90 Minuten pro Woche) durchgeführt, jeweils mit anderen Gesprächspartnertypen und leicht verschiedenen Zielen, und zwar in Abhängigkeit vom Fortschritt in der Fremdsprache. Genauer findet sich im Abschnitt 3. 2.

2. TestAS und TestDaF

In Deutschland ist, anders als z.B. in Japan und einigen anderen Ländern, der Abschluss der Hochschule mit besonderen Schwierigkeiten verbunden, während andere Länder das Hauptgewicht auf die Beschränkung des Zugangs legen. Deshalb möchte man in Deutschland sicherstellen, dass das Studium auch erfolgreich abgeschlossen wird. Damit man die Chancen absehen kann, verlangen viele Hochschulen in Deutschland TestAs, „diesen objektiven, treffsicheren, standardisierten Eignungstest zur Auswahl und Zulassung von ausländischen Studierenden“ (TestAS 2013: 2). Dieser „Eignungstest“ hat einen Kerntest, der die allgemeine Fähigkeit prüft, und Module, die die besonderen Fähigkeiten für einzelne Fächer überprüfen.

Allerdings ist nicht nur das fachliche Wissen wichtig, sondern es müssen auch die sprachlichen Voraussetzungen für die Aufnahme eines Studiums vorhanden sein. Der dafür entwickelte Test TestDaF (TD) wurde beim Lektorenfachseminar 2013 von einer offiziellen Vertreterin (Zimmermann 2013) vorgestellt, und die dabei erwähnten Punkte können als von den Erstellern als „erwähnenswert“, d.h. in besonderer Weise den Test kennzeichnend, angesehen werden. Einige dieser Punkte sollen als Start für den Vergleich im dritten Teil dieser Arbeit dienen.

Bei TD geht es nicht nur um die wissenschaftlichen Kontexte, sondern um alle Situationen, die an der Universität entstehen können und sprachlich zu lösen sind. So geht es sowohl um mündliche Beteiligung am Unterricht in den verschiedenen universitären Formen, wie z.B. Referat, Diskussionsbeitrag, Meinungsäußerung, Gespräch mit dem Professor usw., sondern auch um ganz alltägliche Beiträge wie Arbeit finden, mit Kommilitonen an einer Arbeit zusammenarbeiten, eine Meinung wiedergeben, usw. Wohl am umfangreichsten ist das Hörverstehen, schon in normalen Sprachlernprozessen sehr schwierig. Da Universitätsangehörige unbegrenzt fast alle Typen von Texten, vom Fachtext bis zum Urlaubsgespräch, verstehen müssen, sind Fähigkeiten gefordert, bei denen nicht selten Muttersprachler enorme Schwierigkeiten

haben, um es freundlich zu sagen. Etwas anders sind die Verhältnisse beim Lesen und Schreiben: Wer sich für eine deutsche Universität bewirbt, hat oft schon in seinem Heimatland eine überragende Bildungsvorgeschichte (bezogen auf das jeweilige Land) und da gehört nicht selten Lesen und Schreiben dazu, obwohl dies alles andere als selbstverständlich ist. In diesem Fall geht es eher um die Bewältigung von großen Materialmengen in besonders ökonomischer (und nicht zuletzt auch sportlicher) Weise. Das Fernziel beim TD ist fest auf die Zukunft (Universitätszulassung) ausgerichtet, der Test ist also eine Art *gate-keeper*. Bei dem RS ist es genau umgekehrt. Für die meisten Studenten ist er der letzte Kontakt mit Deutsch in ihrem ganzen Leben. Dann kann RS nur den Lernern zeigen, wieviel sie, nach oft fürchterlichen Erfahrungen mit sechs Jahren Englisch, in einer Fremdsprache in sehr kurzer Zeit erreichen können. Es kann dann nur darum gehen:

1. zu zeigen, was man sogar in einem Jahr erreichen kann, nämlich Gespräche mit Muttersprachlern, wenn man nur richtig anfängt;
2. die wichtigsten Sprachlernpunkte des Deutschen als Start für weiteres Fremdsprachenlernen vorzubereiten; und
3. zu lernen, wie man lernen kann, falls ein solches Fremdsprachenlernen später noch einmal nötig wird.

Während der RS nach nur einem Jahr nach 30/60 Unterrichtsstunden teilweise auf A2 des Referenzrahmens (GER 2013) erfolgt, erreichen die Studenten nach Einschätzung von Universitätslehrern nach 6-7 Jahren Englisch in etwa B1, gerade mal eine Stufe höher.

Während TD Sprachkenntnisse auf fortgeschrittenem Niveau abcheckt (C1), geht es beim RS gerade um etwas mehr als das nicht-minimal Niveau.

Bei TD sind die Ziele getrennt nach Fähigkeiten und Fertigkeiten. Auch beim RS werden die Fertigkeiten getrennt geprüft und in dieser Arbeit geht es nur um den Sprechtest.

Während der TD alle im Hochschulleben möglichen Sprachsituationen zu

berücksichtigen versuchen muss, geht es bei RS nur um Themen, die für eine erfolgreiche Bekanntschaftsaufnahme nötig bzw. hinreichend sind. Da außerdem keine Lebenspraxis voraussehbar ist, geht es eher um eine ideale Bereitschaft für eine, teilweise globale, Änderung bzw. Erweiterung des allgemeinen Wissens (Reinelt 2013d) bzw. des Kommunikationswissens (Reinelt 2013e).

Außerdem ist die Richtung der Benutzung des Wissens genau umgekehrt: Während TD alles Sprachliche für ein Studentenleben vorbereiten möchte und muss, muss RS auf das Studentenleben bis dahin aufbauen um überhaupt den Studenten verständlich zu sein, da Universitäten in Japan doch einige Unterschiede zu Deutschland aufweisen (Reinelt 2013d, e).

Als Ausgabe erfolgt bei TD ein Zeugnis mit Leistungen und differenzierter Auflistung während am Semesterende nur eine allgemeine Punktzahl angegeben werden kann, zu deren Berechnung das RS-Ergebnis nur einen kleinen Betrag leistet (zusammen mit den anderen Leistungen beim Kursleiter und denen beim Partnerlehrer). Außerdem wird zur gleichen Zeit bei RS ein schriftlicher Test geschrieben.

Die Länge des TD beträgt 30 Minuten vs. 3-5 Minuten bei RS.

Bei TD werden zur internen Qualitätssicherung empirisch Aufgaben erstellt und abgecheckt. Bei RS ist als Gegenstand nur der Bereich des Unterrichts bisher definiert, ansonsten gibt es um die Spontaneität zu erhalten keine Vorgaben.

Ein oberstes Ziel in Japan, die Gleichbehandlung, wird bei RS durch den gemeinsam gelernten Bereich abgesichert, wobei dann dem Lerner überlassen ist, was er in seine Dyade einbringt..

3. Vergleichspunkte

Einleitung

Beide Tests kann man sowohl hinsichtlich der theoretischen Vorgaben (3. 1.) wie auch der praktischen Durchführung (3. 2) vergleichen. Tests müssen, damit sie überhaupt

vergleichbar sind und damit sie Aussagekraft über einen kleinen Bereich hinaus haben, Gütekriterien erfüllen. Die drei wichtigsten sind Objektivität, Reliabilität und Validität und werden hier zuerst behandelt, ebenso wie Besonderheiten der Testerstellung (3. 1. 1. 1 bis 3. 1. 1. 4). Einige weitere wichtige Kriterien werden im Anschluss daran verglichen (3. 1. 2). Daran schließen Ausführungen zur praktischen Durchführung (3. 2.) an, sowie zur Beurteilung (3. 3.). Schließlich gehen wir auf ein Beispiel: Sprechen ein (3. 4.).

3. 1. Theoretische Kriterien

3. 1. 1. Drei wichtige Gütekriterien und die Testerstellung:

Während TestDaF die folgenden drei Kriterien streng ein (zu) halten (versuchen) muss, ist dies beim RS nicht so einfach.

3. 1. 1. 1. Objektivität

Nach diesem Kriterium sind die Testergebnisse immer gleich unabhängig von Zeit und Raum und von der Durchführung und der Auswertung. Eine Voraussetzung für das Erreichen dieser Ziele ist eine Standardisierung.

Standardisierung

Bei RS wurde die Standardisierung schon an verschiedenen Stellen innerhalb der Universität durchgeführt. Deshalb kann davon ausgegangen werden, dass wenigstens hinsichtlich des Ortes und der Zeit die nötige Objektivität gegeben ist. Dies wird natürlich bei Durchführungen an anderen Universitäten oder mit anderen Teilnehmerzusammensetzungen neu zu untersuchen und zu kalibrieren sein.

3. 1. 1. 2. Reliabilität

Dieses Kriterium „bezieht sich auf die Genauigkeit, mit der bestimmte Eigenschaften gemessen werden“ und zwar unabhängig von dem Zeitpunkt der Messung und der Testform.

Bei der RS sind beide Aspekte problematisch: Die Messung kann nur am Ende des zweiten Semesters in dieser Form stattfinden, und ein Erstgespräch mit einem

Muttersprachler kann *per definitionem* mit jedem Partner nur einmal geführt werden. Es wäre z.B. am Anfang des zweiten Studienjahres zu untersuchen, ob dasselbe Anfangsgespräch mit einem anderen Muttersprachler (noch) möglich ist. Andererseits ist die Behaltensrate nach den Frühlingsferien doch immerhin noch so hoch, dass schon nach etwa zwei Unterrichten die Inhalte wenigstens weitgehend wieder präsent sind. Bei Lernern, die RS nicht gemacht haben, ist dann gewöhnlich gar nichts mehr vorhanden außer ein paar Wörtern.

Quantifizierung

In der Praxis hängt die Reliabilität davon ab, ob die erbrachte Leistung zuverlässig quantifiziert wird.

Bei RS soll dies durch eine doppelte Beurteilung garantiert werden, einerseits durch eine holistische Bewertung, und zum anderen durch eine Beurteilung mit einer Scorecard. Diese wurde nach der Sichtung von zig Aufsätzen, Papieren und anderen Arbeiten vom Verfasser neu für die Verhältnisse an japanischen Universitäten im Bereich der Allgemeinbildung und im Bereich des Sprechens entwickelt. Da ferner bis zu einem gewissen Grade davon ausgegangen werden kann, dass alle deutschen Muttersprachler beim Einschätzen mit den Kriterien aus ihrem eigenen FL vertaут sind und auch mit der Skalierung umgehen können, sollte immer die gleiche Quantifizierung stattfinden.

Das Instrument wurde so entwickelt, dass je nach Wunsch auch verschiedene Gewichtungen vorgenommen werden können.

3. 1. 1. 3. Validität

Dieses Kriterium berücksichtigt, ob auch getestet wird, was der Test zu testen vorgibt, und zwar mit den folgenden Unteraspekten:

x-in Hinblick auf das Konstrukt (Konstruktvalidität)

Da es sich bei RS bei den Testnehmern am Ende des Wintersemesters bis jetzt durchgehend um Zweitsemester, also Studenten im ersten Studienjahr, handelt, die

eben nur das Deutsch gelernt haben, was im Unterricht gelehrt und geübt wurde, kann nichts anderes getestet werden, und der Test nur und genau dieses erfassen.

Andererseits kann es natürlich sein, dass der Test hinter dem zurückbleibt, was sein könnte. So kann ein Student einen deutschen Muttersprachler etwas fragen und nur in wenigen Fällen wird der Student eine ausführliche Antwort verstehen.

Andererseits sind kurze Antworten allein schon durch die Vorhersehbarkeit garantiert. Große Abweichungen sind bisher noch nicht vorgekommen, manchmal aber Wörter mit gleichen Bedeutungen zu übersetzen gewesen. Das gleiche betrifft die Themenauswahl. Da Anfangsgespräche, und um diese handelt es sich, normalerweise nur einen beschränkten Bereich von Themen ansprechen (können), testet der Test dann wirklich, was die Lerner gelernt haben. Obwohl es natürlich praktisch nicht ausschließbar ist, hat bisher keiner mit der Kritik der praktischen Vernunft oder ähnlichen abstrakten Themen angefangen.

Konstrukt fremdsprachliche Kompetenz: Dass es sich um einen Test der Kompetenz handelt kann man daran sehen, dass hier nur der mündlich produktive Teil angesprochen ist, andere Teile werden in einer gleichzeitigen schriftlichen Prüfung abgeprüft.

Hören ist in diesem Fall problematisch, aber das ist natürlich immer der Fall. Andererseits kann man Sprechen auch dirigieren, indem man den Partner etwas erzählen lässt (oder dahin bekommt). Aber als Sprechstrategie ist das natürlich legal, zumal im Test.

X- in Hinblick auf die Authentizität Aufgaben (Inhaltsvalidität):

Die Authentizität der Aufgaben ergibt sich bei RS daraus, dass alles *on-the-spot* da spontan produzierte Äußerungen sind, die allerdings mit festen Mustern gesagt werden können. Alle Dyaden bestehen nur aus Teilen mit im Unterricht behandelten Stoffen. Allerdings gibt es nicht die Anforderung, einen bestimmten Teil zu bringen. Insofern werden einige Teile als gleichberechtigt angesehen: *Über Besitztümer zu sprechen*

(haben, es gibt + Akk.) entspricht *über Kleidung zu sprechen* (anhaben/tragen + Akk.), usw.

X- in Hinsicht auf die Übertragbarkeit auf die Realsituation (Vorhersagevalidität)

Während TD die Vielzahl und die Vielfalt der authentischen Sprachhandlungen in der realen Welt berücksichtigen muss, ist die im RS gezeigte Leistung eben Performanz.

Diese Übertragbarkeit auf die Realität ist vielleicht der einzige Punkt, in dem der RS dem TestDaF uneinholbar überlegen ist. Alle Dyaden sind echt: Die Beteiligten kennen sich vorher nicht und werden durch dieses Gespräch nicht-mehr -ganz- unbekannte Teilnehmer (Augenscheinvalidität) und Testabnehmer. Die Studenten waren bisher unmittelbar nach der Dyade meist überrascht, dass sie selbst so lange in einer Fremdsprache sprechen können, insbesondere nach den oft enttäuschenden Erfahrungen mit dem Englisch in der Schule.

Bei Exit-befragungen der Testabnehmer/beurteiler wurde zwar eine gewisse Unzufriedenheit offenbar, insbesondere in Hinsicht auf die geringe Themenauswahl (wobei einige mögliche allerdings von den Studenten gar nicht benutzt wurden) und die Ermangelung des Beweises des Hörverstehens. Wenn wir aber nach einer Alternative und besonderen Wünschen (andere Themen, Formulierungen, usw.) fragten, wurden keine angegeben.

3. 1. 1. 4. Testerstellung: Qualitätssicherung und –standards

Qualitätsstandards bestimmen, was die Anbieter sicherstellen müssen, nämlich dass

- Zielsetzung und Konstrukt genau definiert sind, und
- Die Leistungen Rückschlüsse auf die Situation außerhalb der Testsituation zulassen.

Da TD feste Aufgaben verwendet, erfolgt die Testerstellung bei TD intern nach verbindlichen Vorgaben. Dabei erfolgt die Qualitätssicherung durch eine interne Expertengruppe.

Da bei RS die Inhalte sich erst im Gespräch ergeben, entfällt diese Problematik. Sie ist

aber durch die bis dahin behandelten Inhalte (und Strategien der Studenten zu deren Einhaltung „*noch einmal bitte*“) abgesichert.

TD wird teuer standardisiert und kann und muss dies auch werden als *high-stakes* (High-stakes testing 2013) test. Der RS muss einige Vorannahmen machen, die als erfüllt anzusehen sind, und erst dann kann eine Standardisierung ansetzen. So muss bei Sprechpartnern und Beurteilern die Tatsache, mit dem Notensystem 1 bis 5 sowie Sprechtests vertraut zu sein, vorausgesetzt werden. Ebenso muss angenommen werden, dass einige Inhalte gleich gut sind, so z.B. die ausführliche Behandlung von „gestern“ und die Behandlung der eigenen Gegenstände, beides Themen, die im WS behandelt und oft wiederholt wurden. Erst unter der Voraussetzung der Parallelität dieser Punkte kann die Standardisierung angegangen werden.

TD muss scharf aufpassen, dass die Kriterien strikt erfüllt werden. Dies ist nicht nur ein Teil des Geschäfts und des Vertrauens, das die Universitäten in TD setzen, sondern auch eine Frage der Fairness, die sogar eingeklagt werden kann. Im Falle des RS kann man nur hoffen, dass die Kriterien eingehalten werden. Deshalb müssen die Testleistungen auch mit einigen Abständen geprüft werden, ob sie noch die Stufen erreichen, also nach einem Jahr noch A2 sind.