

Praktikabilität und Relevanz

Zur Optimierung von universitären mündlichen DaF Prüfungen in Japan

Rudolf Reinelt

1. Einleitung

Eine (jede?) Theorie ist (spätestens?/schon?) veraltet, wenn Sie fertig (abgeschlossen?) ist. Dies gilt sicher für verschiedene Bereiche der wissenschaftlichen Forschung in unterschiedlicher Weise: Vielleicht weniger für abstrakte Theorienbereiche wie die reine Mathematik, um so mehr aber für in der menschlichen Praxis fundierte Bereiche wie das Fremdsprachenlernen.

Diese zeitliche Verankerung erfordert zugleich eine dauernde Optimierung zumindest der wichtigsten Bestandteile jeder Theorie. Als ein Beispiel dafür habe ich in Reinelt (2008d) das Ziel des Deutschunterrichts (stellvertretend auch für andere Fremdsprachen jenseits von Englisch und der Muttersprache) an japanischen Universitäten definiert als:

"Eine möglichst vielfältige Einführung in möglichst viele Fertigkeiten der Zielsprache und die Techniken zu deren Erwerb, sowie im Unterschied zu reinen Sprachkursen (..) eine minimale theoretische Einordnung der behandelten sprachlichen und damit zusammenhängenden kulturellen

Erscheinungen und aktuellen Entwicklungen" (2008d, p. 111).

Eine erst in der letzten Zeit mit dem raschen Fortschritt der technischen Medien möglich gewordene Entwicklung ist die einseitige Schwerpunktlegerung auf Sprechfähig- und fertigkeiten. Es sind nämlich NUR diese, die NUR im Frontalunterricht (und adäquaten Formen wie Fernunterricht mit simultaner audiovisueller Sprecherpräsenz von Lerner und Lehrer (also *Skype*, aber nicht *chat*!)) geübt werden können, die aber zugleich alle Schwierigkeiten des Fremdsprachenlernens in konzentrierter Form repräsentieren bzw. mitsichbringen, z.B. Schnelligkeit, Richtigkeit, Reichhaltigkeit, Flexibilität, Gegenseitigkeit usw.

Jede wissenschaftliche Theorie muss kontinuierlich mindestens in zwei Richtungen überprüft werden: Zum einen nach innen, z.B. auf Konsistenz und optimale Ausgestaltung wie, im Bereich des Fremdsprachenunterrichts, die Praktikabilität; und zum anderen nach außen auf ihren Bezug zu ihrem "Geltungsbereich" und in Abgrenzung zu anderen Bereichen und Theorien, d.h. ihre Relevanz.

Dementsprechend diskutiere ich in der vorliegenden Arbeit zuerst einen Versuch der inneren Optimierung durch Vereinfachung des Prüfinstruments, die mündliche Prüfung. Im Anschluss daran wird die Entwicklung der Relevanz des erwähnten Ansatzes in seiner zeitlichen Dimension überprüft. Hinweise auf weitere Forschungsaufgaben beschließen diese Arbeit.

2. Praktikabilität

2.1. Mündliche Prüfungen für Fremdsprachen in Japan entwickeln

Nach der einsichtsvollen, aber nicht immer eingehaltenen und einzuhaltenen These, dass die Erfolgsüberprüfung eines Kurses in der bis

dahin im Mittelpunkt stehenden Fähigkeit/Fertigkeit erfolge sollte, ist eine mündliche Prüfung ein notwendiger wie unvermeidlicher Abschluss eines Kurses, der das oben angegebene Ziel anstrebt.

Die Schwierigkeit in Japan besteht nun darin, dass die weltweit standardisierten mündlichen Prüfungen nicht einfach übernommen werden können (Jeffrey n.d.). Dies gilt schon deshalb, weil in keinem anderen Land das "echte", d.h. nicht-instrumentale, z.B. nicht zum Alltagsleben notwendige, Fremdsprachenlernen so weit im nicht-elitären Schul- und Alltagsleben integriert ist wie in Japan, was auch immer man über die Erfolge sagen mag. Dies erfordert die Erstellung neuartiger wissenschaftlich abgesicherter mündlicher Prüfungen, und selbst für den Bereich des Englischen sind diese erst in den Anfangsstufen (Jeffrey n.d. Smith & Nederend 1998). Deren Ziel ist dann, ein Gespräch zu beurteilen, in dem "ein williger Zielsprachenmuttersprachler bereit ist, mit dem Prüfling eine bestimmte Zeit mit Sprechen zu verbringen" (nach Jeffrey n.d.). Dieses Ziel ist aber besonders im Bereich weiterer Fremdsprachen mit seiner begrenzten Zahl von verfügbaren Muttersprachlern nur selten erfüllbar.

In Anlehnung an die obigen Arbeiten hat der Autor für den Bereich des Deutschen als zweite Fremdsprache in Japan eine Prüfung vorbereitet und über Jahre standardisiert, bei der zumindest auf der Lernerseite keine Beschränkungen vorliegen.

Beurteiler (im folgenden wie in der englischen Literatur: Rater) sehen ein nicht vorstrukturiertes Gespräch zweier zufällig zugeteilter Lerner in der Zielsprache von zwei bis drei Minuten Länge und beurteilen die jeweilige Leistung nachher auf einer Scorecard (s. Reinelt 2008c, p. 79) nach (reduziert von sieben) fünf Kriterien. Dabei unterscheiden sich international standardisierte wie selbst erstellte Prüfungen normalerweise nur hinsichtlich

der zu berücksichtigenden Kriterien, während die Notenbandbreite und -einteilungen auf die Rater abgestimmt sein sollten und über Berechnungen auf bestehende Systeme (z.B. das 60-100 Punktesystem der Ehime Universität (Reinelt 2008a) abgestimmt werden. bei mehreren Ratern wird die Korrelation anhand des Ulmer Rechners (Inter correlations n.d.) errechnet.

Tafel 1 Kriterien

(in Klammern die an der Ehime Universität vorgenommene Gewichtung)

Aussprache = pronunciation (15%)

Korrektheit = correctness, grammar (15%)

Wortschatz = (richness in) vocabulary (20%)

Flüssigkeit = fluency (35%)

Dialogizität = dialogicity (15%)

Eine Übersicht über die Entwicklung der Prüfungen gibt die folgende Tafel 2:

Tafel 2 Entwicklungsstufen

	Beteiligte: Prüfer	Beteiligte: Lerner	Durchführung	Beurteilungstyp	Vorteile	Nachteile	Anmerkungen
Abkürzungen: hol=holistisch, ko=kriterienbezogen MS=Muttersprachler RR=der Autor							
0 Ideal	1Muttersprachler	1Lerner	sprechend	hol und ko möglich	scht!	Bereitschaft der MS ist Voraussetzung	Unmöglich wegen fehlender Anzahl an Muttersprachlern
1 Reinelt 2000	RR	1Lerner	Lerner spricht mit RR	Mindestens 4 Wechsel, dann holistisch	mit sehr vielen Studenten möglich, sönlich, Prüfer kann helfen	unausgewogen, Beurteilung kann schwanken; Lehrer als Prüfer im gleichen Kurs sehr problematisch (Vorbelastung)	sehr stressig für den Prüfer, (Organisation, Prüfung, Beurteilung) Wartezeiten für Lerner
2 Reinelt 2008c	RR	2Lerner	2 Lerner sprechen miteinander etwa 2 min	Holistisch, ko u. U. möglich	schnell viele möglich, annähernd gleiche Bedingungen	Beurteilung kann noch sehr schwanken, Wartezeiten	Stressig für Prüfer (Organisation, Zeitnahme, Beurteilung, &ca.)
3 Reinelt 2008a	RR & 1 weiterer Rater	2Lerner	2 Lerner sprechen miteinander etwa 2 bis 3 min.	beide hol (RR) ko (weiterer Rater)	beide Beurteilungstypen (ideal nach Jeffrey n.d.)	Andere Beurteiler nötig Vorbereitungszeit (training) Finanzielle Entschädigung	Stress für Prüfer, nicht-professionelle Raterermüden schnell, Rater halten gleichen Beurteilungslevel?
4 Reinelt 2010	RR & 2 weitere Rater	2Lerner	2 Lerner sprechen miteinander etwa 2 min.	=oben	mehr Objektivität?	oben plus Correlation kann sinken	mehr Rater suchen
5 Reinelt 2009	RR & 1 weiterer Rater (verschiedene Typen)	2Lerner	2 Lerner sprechen miteinander etwa 2 min.	=oben	u.U. viele Rater verfügbar	Finanz. Aufwendungen	Alternativen vorbereiten bei Nichtkommen nötig; anderes

2.2. Reduzierung der Belastung

2.2.1. Vereinfachung der Prüfung selbst

Um die Belastung aller Beteiligten zu erleichtern, scheinen Alternativen in zwei Richtungen möglich, die zum Teil schon viel älter (und anderweitig standardisierter) sind als die in Tafel eins und zwei vorgestellte neu entwickelte Prüfung:

- die Prüfungsinhalte reduzieren und die Form ändern (s. Reinelt 2010)

Dazu gehören in Tafel 3 zum Beispiel die folgenden Prüfungs- bzw. Sozialformen und Inhalte. Spalte zwei gibt an, warum diese, trotz ihrer größeren Praktikabilität, gegenüber der vom Autor vorgeschlagenen mündlichen Prüfung eben doch *keine sinnvollen Alternativen* darstellen.

Tafel 3: Problematische Alternativen

Prüfungs- und/oder Sozialform, Inhalt	Problem
vorbereitete Dialoge	nicht spontan
(offene) Fragelisten abarbeiten	nicht sehr dialogisch
vorher festgelegte Partner	auswendig lernen
vorher bestimmter Inhalt	dann nicht dialogisch, sondern nur eine Sprech (Aussprache) prüfung
zu feste sprachliche und inhaltliche Vorgaben	im voraus planbar, zu "sicher"
zu wenig verschiedene Inhalte	auswendig lernen möglich; zu wenig Kreativität gefordert
zu langer Zeitrahmen, zu viel Zeit zur Verfügung	Sprechen und Sprecherwechsel in natürlicher Sprechgeschwindigkeit wichtig
Sprecherwechsel, Unterbrechungen nicht verlangt	Monologisch, fehlende Dialogizität
Benutzung von Materialien z.B. Lehrbuch	stehen beim normalen Sprechen auch nicht zur Verfügung
Bildbeschreibungen	monologisch; gut, aber keine Zeit, anderswo besser möglich

Sinnvoll (d.h. von Studenten eingefordert, aber nur schwer statistisch als hilfreich nachweisbar) sind andererseits im voraus Hinweise zur Durchführung der Prüfung und solche die sich auf die Gesprächsgestaltung

als Ganzes beziehen, z. B. die Prüfungszeit "als Begegnung zu gestalten", usw. Außerdem führten im Sommersemester 2009 Jahr Hinweise darauf, dafür zu sorgen, dass auch der mögliche Partner genug Deutsch kann, zu mehr Zusammenarbeit der Lerner untereinander vor der Prüfung als z.B. Ermahnungen, gegenseitig zu üben.

- eine weitere Vereinfachung kann durch die Reduzierung der Beurteilungskriterien versucht werden. Weil dies eine wesentliche praktische Verbesserung und eine wichtige theoretische Vereinfachung darstellt, wird ein solcher Versuch im folgenden Teil ausführlich vorgestellt.

2.2.2. Kriterienreduktion

Eine weitere Möglichkeit, die Praktikabilität zu verbessern, könnte darin bestehen, die Anzahl der Beurteilungskategorien zu reduzieren. Dies erscheint sinnvoll, weil Rater sicher drei Kriterien besser gleichzeitig beurteilen können als fünf. Dass dies allerdings auch nicht ganz unproblematisch, obwohl auf den ersten Blick sehr wünschenswert, ist, ist erst in letzter Zeit klar geworden.

Wie leicht vorstellbar, kann die Kriterienanzahl und -auswahl stark variieren, je nachdem welche Kriterien man als wichtig ansehen möchte. Die Anzahl liegt aber auch bei professionellen Tests normalerweise nicht über 7. Dies ist wahrscheinlich die höchste, gleichzeitig beurteilbare Anzahl, und wohl nur von beruflichen Ratern handhabbar. Nicht-berufliche Rater, wie es die meisten Lehrer sind, können bis zu fünf Kriterien gleichzeitig handhaben, und deshalb wurde diese Anzahl vom Autor für seine Arbeiten übernommen. Diese Anzahl erfüllt zugleich auch die Anforderung der Realität, d.h. es muss auch machbar sein ohne wesentliche, d.h. unter normalen Arbeitsbedingungen nur schwer leistbaren Zusatzanforderungen (Reinelt 2005).

In einem neuerlichen Versuch wurde die Anzahl der Beurteilungskriterien auf drei reduziert: Aussprache, Grammatik und Flüssigkeit.

Aufgrund einer weiteren Erleichterung auf der Inhaltsebene brauchten die Studenten nur vier bis fünf Frage- Antwortsequenzen aus vorher eingeübten Sequenzen produzieren, so daß die Fehlermöglichkeit ebenfalls sehr gering war.

Bei dieser Konstellation wurden einige Punkte offenbar, die ohne diesen Versuch gar nicht als Problem vorstellbar waren.

- Aussprache: Da alle Studenten zumindest eine der beiden Wochenstunden beim Muttersprachler Unterricht hatten und die Fremdsprache von Anfang an gelernt wurde, war die Aussprache durchgehend gut bis sehr gut, und Unterschiede nicht so gravierend, dass sie ein eigenes Kriterium erforderten. Dies war nicht von vornherein anzunehmen, erklärt sich aber aus der Erwerbssituation: Im Unterschied zu Englisch an japanischen Universitäten liegt bei Deutsch eine reine Anfangslernsituation vor. Deshalb entfallen normalerweise auch weitgehend die *false beginner* Probleme. Wenn dazu noch die Menge an Produktionen in der Prüfung nur sehr begrenzt ist und deren Inhalt weitgehend festliegt, so dass (es sich meist einfach um ein erinnertes Nachsprechen handelt, wo) wenig Unsicherheit entsteht, dürfte und könnte die Aussprache aufgrund der Güte und der zu geringen Unterschiede als zu beurteilendes Kriterium entfallen.

- Grammatik: In gleicher Weise waren nicht viele Grammatikfehler festzustellen, obwohl in diesem Bereich aufgrund der vielen Endungen im Deutschen schon eher Unterschiede auftauchten. Aufgrund der geringen Anzahl an verschiedenen Äußerungen waren aber auch bei diesem Kriterium die Möglichkeiten beschränkt, so dass Grammatik bei einer geringen Zahl an Äußerungen ein problematisches, wenn nicht gar einfach weglassbares, Kriterium bleibt.

- Die Flüssigkeit war, trotz sicher vorgenommener Übungen und Unterricht beim Muttersprachler noch ein Problem, und deshalb aus gutem Grund Gegenstand der Beurteilung.

Insgesamt ergaben sich folgende Minimalergebnisse(Tafel 4) :

Tafel 4: Standardabweichung und Varianz bei drei (3) und fünf (5) Kriterien

Stat/ Unt	Mo3(3)	Mo4(3)	Do5(3)	Fr3(3)	Di 3(5)	2 Rater 10.38 vs.11.38 (5)
Durchschnitt	0.73	0.74	1.72	1.36	2.15	2.18 vs. 2.28
Standardabweichung	0.73	0.74	0.52	0.76	0.49	0.66 vs. 0.78
Varianz.	0.56	0.55	0.78	0.57	0.24	0.44 vs. 0.6
Standardabweichung minus Varianz	0.17	0.19	-0.16	0.19	0.25	0.22 vs 0.18

Nach Haupt (2007) ist, stark vereinfacht gesagt, eine hohe Standardabweichung mit großer Trennschärfe verbunden. Diese wird in allen Drei-Kriterien-Klassen erfüllt. Andererseits führt hohe Varianz zu niedriger Trennschärfe. Besonders problematisch werden Fälle, bei denen der Unterschied unter . 20 Punkten (von der Standardabweichung aus gesehen) liegt, weil dann nicht mehr klar unterscheidbar ist, welche Notenstufe der Rater gemeint hat. Dies ist allerdings in allen obigen Drei-Kriterien-Beispielen ebenfalls der Fall. In den Durchführungen mit fünf Kriterien war dies nur in einigen wenigen Beispielen der Fall.

Tafel 5 Standardabweichung und Varianz bei einzelnen Kriterien

Mo4Fr3 (3)	Aussprache	Grammatik	Flüssigkeit
Average	1.39	1.39	1.56
Standardabweichung	0.61	0.7	0.92
Varianz	0.37	0.49	0.85

Mo3	Aussprache	Grammatik	Flüssigkeit
Average	1.26923077	1.153846	1.8461538
Standardabweichung	0.53349357	0.367946	0.9671528
Varianz	0.28461538	0.135385	0.9353846

Für die einzelnen Kriterien fiel wie in Tafel 5 das Ergebnis wieder sehr unterschiedlich aus. Während Aussprache und Grammatik hohe Trennschärfe aufweisen, ist die Varianz bei Flüssigkeit fast so hoch wie die Standardabweichung und damit fast keine Trennschärfe mehr vorhanden. Insgesamt ließ sich feststellen, dass aufgrund der Neu-Lernsituation einige Kriterien nicht sofort akut werden (Aussprache) oder aufgrund der geringen Menge an notwendigen Äußerungen und der Sicherheit wenig Grammatikfehler möglich waren. Die Flüssigkeit war aber, wie hohe Standardabweichung und Varianz zeigen, schon in allen Drei-Kriterien-Beispielen von Student zu Student sehr variabel und deshalb zu recht Beurteilungsgegenstand.

Im Vergleich dazu ist im Versuch mit fünf Kriterien die Standardabweichung zwar nicht so hoch, die Varianz aber vergleichsweise niedrig, d.h. die Kriterien sind gut voneinander abgegrenzt (Tafel 6):

Tafel 6 Standardabweichung und Varianz bei einzelnen der fünf Kriterien

Average	2.175	2.425	2.275	2.275	1.975
Standardabweichung	0.385	0.54948	0.5541	0.5057363	0.2762
Varianz	0.148	0.30192	0.3071	0.2557692	0.0763
Standardabweichung minus Varianz					
	0.24	0.24	0.25	0.25	0.2

An dieser Stelle liegen allerdings keine Alternativvorschläge für besser trennende Kriterien vor.

Während die beschränkte Anzahl an Äußerungen kein Problem für holistische Bewertungen darstellt, ist klargeworden, dass für Kriterienbasierte Bewertungen mehr Äußerungen und noch andere Kriterien notwendig sind für eine ausgewogene / gültige Bewertung, wenn man die Kriterienanzahl reduzieren möchte. Welche anderen Kriterien verwendbar sind, müssen allerdings erst weitere Versuche zeigen.

3. Relevanz

Die in der Einleitung vorgeschlagene Theorie des Fremdsprachenunterrichts muss außerdem nach außen geprüft werden, und zwar in ihrem Bezug zu ihrem angenommenen "Geltungsbereich", dem Universitätsunterricht und dessen Annahme darüber durch die Lerner.

Um bei dem vorgestellten Unterricht Zeit für Sprechen frei zu bekommen, wurden viele Teile ausgelagert (Reinelt 2008b). Es bleibt also zu fragen, wie die wichtigsten Elemente des *Outsourcing* (*Moodle*, *Active Mail*, Hausaufgaben) aufgenommen wurden und ob sich über die Jahre Veränderungen ergeben.

Zwei Besonderheiten seien kurz erwähnt:

- zum einen werden, aus Erfahrung, Neuerungen an der Universität von Studenten in Japan besonders dann nicht positiv beurteilt, wenn sie mit zusätzlicher Arbeit verbunden sind. Deshalb ist in den ersten Jahren von negativen Beurteilungen auszugehen.

- da der vorgestellte Unterricht keinen Parallelunterricht hat, und Lehrer nicht einfach darum gebeten werden können, ihre Studenten für Vergleichstests zur Verfügung zu stellen, gibt es einfach aus ethischen Gründen (bei voraussichtlich wenig positiven Resultaten) keine Kontrollgruppen, weder für die Unterrichtsversuche noch für die mündlichen Prüfungen, und auch nicht für das hier vorgestellte *Outsourcing*.

Eine Fünf- Punkt- Likert-Skala-ähnliche Umfrage jeweils in der letzten Unterrichtsstunde im Sommersemester 2008 und 2009 beleuchtet die Relevanz des Versuches. Bei der nachfolgenden Darstellung werden nur die Spitzen und die Tendenzrichtungen erwähnt, weil genauere Zahlenangaben einen klareren Eindruck vermitteln würden als der Fall ist.

Im folgenden präsentieren wir zuerst im Original den Fragebogen, der aber

nur den *Outsourcing*-Teil betrifft. Dies ist aber auch der Teil, der die wichtigste Voraussetzung für den erweiterten Sprechunterricht, und die an dessen Ende stehende mündliche Prüfung, darstellt. Der Fragebogen spricht die Annahme des *Outsourcing* innerhalb der Universität als Ganzes wie auch im Unterricht an.

Tafel 7 Fragebogen zum Outsourcing im gesprächsorientierten Deutsch für Anfänger Unterricht an der Universität Ehime 2008 und 2009 (次ページに)

Praktikabilität und Relevanz
Zur Optimierung von universitären mündlichen DaF Prüfungen in Japan

< Moodle ・ Activemill 2008 年度アンケート >
このアンケートは無記名です。電子教材の使用においては東大であればもっと性能のよいものがあるのですが、愛媛大学では選択が限られてくるため不便に感じることもあったと思います。ごきょうざまでした。そこで、今後この経験を生かすために、以下のアンケートにご協力ください。

【外国語以外の講義における電子教材の利用について】

1. CALL 教室で電子教材を利用したことはありますか。
よくある - 他 - 一部の講義でもまれにある - 大学において利用したことがない

2. CALL 以外の教室で電子教材を利用したことはありますか。
大学において利用したことがない - 他 - 一部の講義でもまれにある - よくある

3. 利用したことがある場合どこで使いますか。
① 大学内 (メディアセンター・図書室)
② 大学内 (クラブなどの部室・担当教員の研究室)
③ 大学外
④ 自宅以外のご自宅 (インターネットカフェなど)
⑤ 自宅
⑥ 携帯

4. 他の講義で Moodle あるいはそれ以外の (ような) ものを利用している講義はありますか。
その場合、その講義の担当教員はどの電子教材を使用していましたか。

【本講義における Moodle の利用について】

1. Moodle の授業ファイル (Unterricht) をどのくらいの頻度で見ましたか。
毎日 - 週 2 回 - 週 1 回 - 2 週間に 1 回 - それ以下

2. それ以外のファイル (宿題などの項目) についてはどのくらいの頻度で見ましたか。
それ以下 - 2 週間に 1 回 - 週 1 回 - 週 2 回 - 毎日

3. Moodle の利用 (使いやすさ) (現在は見やすくするために使用部分を制限してあるのですが) についてはどのくらいの理解ができましたか。
全く分からない - 初めは使いにくいが高くなるにつれて分かってきた - 初めは考えることがあったが慣れていくと普通に利用できた - 初めから使いにくいということはなかった - 簡単な方だ - 簡単すぎる - すぐに扱えた

4. Moodle の利用は勉強に役立ちましたか。
① 復習について
大変役に立った - 一すこしは役に立った - あまり役に立たなかった - 全く役に立たなかった
② 新しい内容の予習として
全く役に立たなかった - あまり役に立たなかった - 一すこしは役に立った - 大変役に立った
③ 課題について
大変役に立った - 一すこしは役に立った - あまり役に立たなかった - 全く役に立たなかった

5. Moodle に関連したことについて、担当教員や助手は手助けをしてくださいましたか。
手助けが多すぎる - よく助けてくれた - まあまあ助けてくれた - あまり助けてくれなかった - 全然たらない

6. 来年以降の言語 (外国語) 講義に Moodle の利用を勧めますか?
強く勧める - あった方がいい - 無い方がいい - 利用なし

7. あなたのメールアドレスを、宿題提出など、授業で使用してもよいと思いますか?
愛媛大学からもらったアドレスだけ OK - 自分の携帯なら OK - 自分のメールアドレス全て OK - 使用したくない

8. あなたから Moodle 利用に関してのヒントや希望などがあれば、自由に記入してください。

【宿題提出における Active Mail の利用について】

1. 宿題はしましたか?
100% - 80% - 60% - 40% - 20% - それ未満

2. 宿題の難易度はどうでしたか?
大変難しい - 難しい方 - 簡単な方 - 簡単すぎる

3. 宿題はドイツ語のよい練習になりましたか? 勉強に役立ちましたか?
とても良かった - 良かった - そうでもなかった - 良くなかった

4. 面白かったですか?
つまらなかった - あまり面白くなかった - 少しは面白かった - とても面白かった

5. 文化を理解するのに役立ちましたか?
とても役立った - まあまあ役立った - 少しは役立った - 全く役に立たなかった

6. 宿題の頻度はどうでしたか?
多すぎる - ちょっと多い - ちょうど良い - もう少しあってとも良い - もっとあってとも良い

7. この授業を今後の学生に推薦できると思いますか?
大いに推薦する - 推薦する - あまり推薦したくない - 全く推薦できない

【全体的な質問】

1. 授業に対する満足度は?
良い → 悪い
1 - 2 - 3 - 4 - 5

2. メディア使用について
良い → 悪い
1 - 2 - 3 - 4 - 5

【試験についての質問】

1. 口頭試験については今後の参考にするためにビデオ撮影を行っています。あなた自身の試験部分のビデオコピーを欲しいと思いますか?
2. もし、あなた自身が必要ではなくてもペアの学生が欲しいと思えば、相手がコピーを持つことを許可できますか?
この授業の全体 (詳細など) についてのご意見をお願いします

Einige auszugsweise Ergebnisse seien in Zusammenfassung vorgestellt:

- Eine langsame Zunahme des *Outsourcing* und der Computerbenutzung an der Universität im zweiten Jahr.
- Die Studenten benutzen das LernManagementSystem *Moodle* tendentiell etwas mehr, von einmal die Woche nach zweimal die Woche, aber der Höhepunkt ist noch nicht dort.
- Die Lerner sehen *Moodle* zu anderen Zwecken als dem Unterricht konstant zweimal die Woche an.
- Die Benutzung wurde unverändert meist im Nachhinein verstanden.
- Die Benutzung war von "etwas" ein bisschen in Richtung "sehr" hilfreich.
- Die Nützlichkeit für die Vorbereitung von neuem Unterricht veränderte sich von "nicht gerade" etwas in Richtung "hat ein bisschen geholfen".
- Für die Themenbehandlung und -aufgaben hat das *Moodle* sich leicht verbessert von "etwas geholfen" minimal in Richtung "sehr geholfen".
- Die Hilfe vom Lehrer wurde als etwas hilfreicher empfunden: von "nicht all zu viel" nach "es geht".
- Eine wichtige Veränderung ergab sich in der Empfehlung des *Moodle*-Einsatzes : Wo voriges Jahr noch "besser nicht" war, ist der Höhepunkt nach "besser vorhanden" gewandert.
- Die Hausaufgaben wurden zu 80% gemacht, wo jetzt auch die Spitze ist.
- "Haben die Hausaufgaben beim Lernen geholfen" hat sich von "nein" nach "gut geholfen" verbessert, wo jetzt auch die Spitze ist!
- Die Menge der Hausaufgaben wurde als "etwas zu viel" angesehen, geht aber leicht in Richtung nach "genau richtig".
- Eine Empfehlung dieses Unterrichts an nachfolgende Studenten wurde früher abgelehnt, geht jetzt aber in Richtung "schon empfehlen", ohne allerdings dort die Spitze zu haben.
- Die Zufriedenheit mit dem Unterricht ist auf einer Skala von 1 bis 5

ziemlich genau zwischen 3 und 4 mit leichter Tendenz nach oben. Insgesamt kann eine leichte Verbesserung der Einstellung zu dem Unterricht und dem Einsatz von *Outsourcing* festgestellt werden. Wenn man davon ausgeht, dass eine neue Routine normalerweise vier Jahre, d.h. einen Durchlauf von Studenten, zur Anerkennung braucht, kann man sagen, dass der Prozess wenigstens gut im Rennen liegt.

4. Weitere Forschungsaufgaben

Zusammenfassend kann man sagen, dass die vom Autor entwickelte Prüfung sowohl einer inneren wie auch einer äußeren Prüfung minimal standgehalten hat. Es bleibt aber immer noch die Möglichkeit der Optimierung der Prüfung.

Eine Reihe von weiteren Arbeiten ergeben sich aus den in diesem Beitrag vorgestellten Untersuchungen. Zum einen geht es darum, einheitlichere Daten zu bekommen, um bessere statistische Untersuchungen durchführen zu können und damit die Ergebnisse besser absichern zu können. Obwohl z.B. für die Überprüfung der inneren Gültigkeit wie in Teil eins eigentlich für eine minimale Standardisierung genug Probanden und Daten vorhanden sind, sind die Voraussetzungen der Lernrer, die Durchführungen und die Ergebnisse doch noch zu uneinheitlich.

Die Bewertung der äußeren Gültigkeit, wie sie in Teil zwei gezeigt wurde, müsste mit anderen äußeren Bewertungen verglichen werden, so z.B. der jeweiligen Semesterendumfrage der Universität. Erste Betrachtungen in den letzten Jahren zeigten teils gravierende Unterschiede in der Bewertung einzelner Punkte. Allerdings hat die Universität Ehime dieses System so umgestellt, dass nicht mehr automatisch alle Studenten gefragt werden, sondern das Erstellen einer Bewertung freigestellt ist und dementsprechend

nur noch von weit weniger als der Hälfte der Lerner erfolgt und damit fast keine Repräsentativität gegeben ist.

Technische Fortschritte können vielleicht direkt für die Durchführung der Prüfung selbst ausgenutzt werden, z. B. für Fernbeurteilung und -scoring, oder auch von Anfang an Gespräche mit Muttersprachlern ermöglichen. Welche theoretischen Folgerungen dies haben könnte, schon aufgrund des nicht-vorhandenen *false beginner* Nachteils, bleibt in der Zukunft zu erforschen.

Literatur

Haupt, C. (2007). *Vorlesung* Testtheorien. München 2007

Intraclass correlation (n.d.)

<http://sip.medizin.uni-ulm.de/informatik/projekte/Odds/icc.html>

Jeffrey, D. (o. J.): *The Challenges of Creating a Valid and Reliable Speaking Test as Part of a Communicative English Program*

<http://www.nuis.ac.jp/~hadley/publication/jeffrey/jeffrey-speakingtest.htm>

Reinelt, R. (2000). Mündliche Prüfungen im Unterricht DaF in Japan. *Der Deutschunterricht in Japan*. 118-123.

Reinelt, R.(2005). Einleitung. In: Balmus, P., Oebel, G., Reinelt, R.

Herausforderung und Chance - Krisenbewältigung im Fach Deutsch als Fremdsprache in Japan. Daf-Westjapan, Ryukyu Universität, Okinawa, Japan 12-14. Dezember 2003 München: Iudicium, 2005, p 13-22.

Reinelt, R.(2008a). Inter-rater reliability in native speaker German beginners course' oral examinations. In Kim Bradford-Watts (Ed.), JALT2007 Conference Proceedings. Tokyo: JALT p.1154-1166.

Reinelt, R. (2008b). 口頭技能強化に向けた未習外国語教育における部分的な Outsourcing の可能性 -Blackboard™ (BB) の試行から- 大学教育実践 ジャーナル 第6号. 35-45.

- Reinelt, R.(2008c) Muttersprachlerbeurteilung von Sprechprüfungen im Deutschunterricht. *ドイツ文学論集*41号 p.73-83.
- Reinelt, R. (2008d) Ex post facto Kurrikulum. *愛媛大学法文学部論集人文学科編*第25号111-124.
- Reinelt, R.(2009): Variously rating oral exams: Native speakers, non-native speakers and more. *Matsuyama: 4th FL Teaching and Research Mini-conference 2009 in Matsuyama* (Sept. 26. 2009), Program handbook p. 12.
- Reinelt, R.(2010, im Erscheinen) *Mündliche Prüfungen mit drei Ratern beurteilen?! II (3人監督による口頭試験評価II)* Ehime Universität Ronsou.
- Smith, A.F.V. & Nederend, W. (1998). Using Oral Interviews at a Junior College. *The Language Teacher*.
(<http://www.jalt-publications.org/tlt/files/98/apr/smith.html>, besucht Feb. 2008).

日本語要約

実用性と重要性

日本における外国語としてのドイツ語の口頭試験の改善のために

理論の改善には方法が二つある：内的な改善と外的な改善である。この論文では、筆者が受講生の希望により行っている会話を重視する外国語授業から、両方の試みの例をあげる。第一部の前書きに続いて、第二部では筆者によるこの授業の定義、及びその口頭試験による評価の実用性を紹介する。

会話中心授業の場合、期末試験に口頭部分を含めなければならない。負担軽減のため試験の平易化（予備準備、暗記、資料の持ち込みなど）が提案されることがあるが（2.2.1）、会話の自発性を失う恐れがあるため、ここでは適切ではない。同様に、試験の評価基準を減らす案も、基準と基準の間のはっきりし

た区別が失われてしまうため、有意義ではない (2.2.2.)。

第三部では、外的な改善として、口頭試験外の部分で、特に口頭練習のために時間をとる授業外学習を扱う。Moodleなどによる授業外学習は受講生にとってメリットがあったか、アンケートでたずねたところ、あらゆる面（利用頻度など）で去年より若干良くなっている傾向が見られた。第四部では、改善のために必要なこれからの課題について述べる。

Zusammenfassung

Praktikabilität und Relevanz

- Zur Optimierung von universitären mündlichen DaF Prüfungen in Japan -
Jede Theorie sollte in zwei Richtungen optimiert werden, nach innen bzw. nach außen. In dieser Arbeit zeige ich Versuche in beide Richtungen aus meinem nach dem Wunsch der Studenten die Sprechfertigkeit betonenden Unterricht.

Nach der Einleitung im ersten Teil stelle ich im zweiten Teil eine Definition (des Ziels) des obengenannten Unterrichts sowie die Praktikabilität der Überprüfung anhand einer von mir entwickelten mündlichen Prüfung vor. Zu deren Erleichterung werden oft Vereinfachungen wie "Vorbereiten" oder "Auswendiglernen" vorgeschlagen (2.2.1.). Diese eignen sich aber nicht, da dabei die gerade für Sprechen charakteristische Spontaneität abhanden kommt. Desgleichen ist eine Reduzierung der Kriterien in der Prüfung nicht sinnvoll, weil dann die Trennschärfe zwischen den verbleibenden Kriterien verloren geht (2.2.2).

Eine äußere Optimierung erfolgt im Teil drei durch Fragebogenbefragung der Studenten, ob sie verschiedene Aspekte der Auslagerung (des *Outsourcing* durch *Moodle* usw.) von anderen Sprechzeit freimachenden Unterrichtsteilen als vorteilhaft ansehen. Gegenüber dem Vorjahr wurde dabei eine leichte Verbesserung festgestellt. Teil vier erwähnt weitere

Aufgaben für die Optimierung.

ⁱ Da sich Fähigkeiten und Fertigkeiten beim Sprachlernen gegenseitig verstärken, ist es unter bestimmten Umständen sogar sinnvoll nachzusehen, wieweit eine andere Fähigkeit mitfortgeschritten ist; man denke nur an den Zusammenhang von Hören und Sprechen.

ⁱⁱ Der Autor dankt JS für die Zurverfügungstellung der wichtigsten Daten und die Diskussion der Hintergründe.