

Zweitveröffentlichung



Reitbrecht, Sandra

Bewertungen von Sprechflüssigkeit : Überblick über eine Forschungslinie mit einem Diskussionsbeitrag für die Prüfungspraxis

Datum der Zweitveröffentlichung: 23.07.2026

Akzeptiertes Manuskript (Postprint), Zeitschriftenartikel

Persistenter Identifikator: urn:nbn:de:bvb:473-irb-116300x

Erstveröffentlichung

Reitbrecht, Sandra (2017): Bewertungen von Sprechflüssigkeit : Überblick über eine Forschungslinie mit einem Diskussionsbeitrag für die Prüfungspraxis, in: ÖDaF-Mitteilungen, Göttingen: V&R, Jg. 33, Nr. 2, S. 96–107, doi: 10.14220/odaf.2017.33.2.96.

Rechtehinweis

Dieses Werk ist durch das Urheberrecht und/oder die Angabe einer Lizenz geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die für Sie geltende Gesetzgebung zum Urheberrecht und/oder durch die Lizenz erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

Für dieses Dokument gilt das deutsche Urheberrecht.

Bewertungen von Sprechflüssigkeit – Überblick über eine Forschungslinie mit einem Diskussionsbeitrag für die Prüfungspraxis

Sandra Reitbrecht

Der vorliegende Beitrag widmet sich der Bewertung mündlicher Leistungen in der Fremd-/Zweitsprache, konkret der Bewertung von Flüssigkeit. Gezeigt werden im Folgenden ein Überblick über die Forschungslinie zu Flüssigkeitsbewertungen und Ergebnisse einer eigenen diesbezüglichen Untersuchung, wobei vor allem Belege dafür herausgearbeitet werden, dass Flüssigkeitsbewertungen interindividuell verschiedene Begründungen zugrunde liegen und dass diese frei verbalisierten Begründungen über das perzipierte Sprechmaterial hinaus Schlüsse auf die Sprecher*innen, ihre kognitiven Aktivitäten sowie ihre emotionale Befindlichkeit oder ihre Sprecher*innenpersönlichkeit beinhalten. Diskutiert werden diese Belege in einem abschließenden Fazit in Bezug auf das konkrete Handlungsfeld des standardisierten Testens und Prüfens. Leitend ist dabei die in einem konstruktivistischen Verständnis von Wahrnehmung begründete Grundannahme, dass Objektivität im Sinne kontinuierlicher Qualitätssicherung ein immer anzustrebendes, aber zugleich nie erreichbares Ziel im Kontext standardisierter kommunikativ ausgerichteter Sprachprüfungen ist.

1. Methodische Zugänge innerhalb der Forschungslinie zu Flüssigkeitsbewertungen

In der Fremdsprachenforschung erfährt (Sprech-)Flüssigkeit als Forschungsgegenstand zu unterschiedlichen Fragestellungen Aufmerksamkeit. Einerseits wird sie mit Komplexität und Korrektheit unter der Abkürzung CAF (engl. complexity – accuracy – fluency) als Merkmal kompetenter mündlicher Sprachverwendung in der Fremd-/Zweitsprache verhandelt und empirisch erforscht (siehe dazu z. B. die Spezialausgabe der Zeitschrift *Applied Linguistics* 30/4, 2009). Ebenso liegen Studien vor, die Zusammenhänge zwischen Flüssigkeit und bestimmten Lernniveaus oder sprachlernbiographischen Variablen untersuchen oder die Wirksamkeit bestimmter Lernsituationen (z. B. Auslandsaufenthalte oder Kursformate) oder Übungsformate im Sinne einer Steigerung des Sprechflüssigkeitsgrades bei Lernenden überprüfen (z. B. Riazantseva 2001; Segalowitz/Freed 2004; de Jong/Perfetti 2011).

Andererseits ist in den letzten Jahren aber auch ein verstärktes Forschungsinteresse in Bezug auf Fragen der Bewertung von Sprechflüssigkeit zu verzeichnen. Studien dieser Forschungslinie sind zumeist so konzipiert, dass L1-Hörer*innen einer Sprache in einem ersten Schritt die Sprechflüssigkeitsgrade von L2-Sprecher*innen anhand elizitierter Sprechbeiträge beurteilen. In einem zweiten Schritt werden diese Bewertungsergebnisse in Bezug zu den Sprechdaten, konkret zu daraus ermittelten Flüssigkeitsparametern wie der Sprech-/Artikulationsrate, der mittleren Pausendauer oder dem durchschnittlichen Vorkommen bestimmter Häsitationsphänomene (Pausen eingeschlossen), gesetzt. Hervorzuheben ist bezüglich dieser Forschungslinie die Tatsache, dass trotz einer verhältnismäßig kurzen Dauer intensiverer empirischer Auseinandersetzung eine forschungsmethodische und forschungsparadigmatische Entwicklung erkennbar ist, die in Abbildung 1 schematisch skizziert ist und im Folgenden anhand konkreter Beispiele näher erläutert wird.

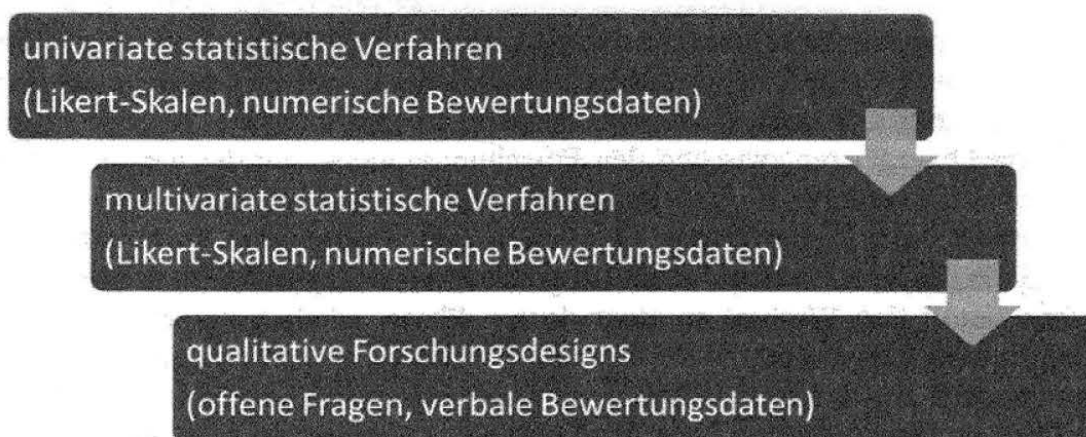


Abb. 1: Entwicklung in der Forschung zu Flüssigkeitsbewertungen

In einer ersten Phase entstanden vorrangig Studien mit quantitativem Design, die mittels Likert-Skalen numerische Bewertungsdaten erhoben und mittels univariater statistischer Verfahren (Korrelationstests) den Zusammenhang zwischen einzelnen der oben genannten Flüssigkeitsmerkmale in den Sprechbeiträgen und der Bewertung von Flüssigkeit ermittelten (Kormos/Dénes 2004; Rossiter 2009; Derwing/Munro/Thomson/Rossiter 2009; Osborne 2011). In einer später einsetzenden zweiten Phase folgten weitere im quantitativen Forschungsparadigma verortete Untersuchungen, die ihre Aussagen auf die Ergebnisse multivariater statistischer Verfahren stützen (Derwing/Rossiter/Munro/Thomson 2004; Bosker/Pinget/Quené/Sanders/de Jong 2013; Pinget/Bosker/Quené/de Jong 2014; Rose 2015; Préfontaine/Kormos/Johnson 2016). Sie ermöglichen in ihrem methodischen Vorgehen nicht nur Erkenntnisse zur Relevanz einzelner Merkmale in den Sprechbeiträgen für die Bewertung, sondern auch zur Bedeutsamkeit von Merkmalsbündeln. Sie werden damit sowohl der Komplexität von Wahrneh-

mungskonstrukten wie Flüssigkeit als auch einem holistischen Verständnis von Wahrnehmung in ihrer Operationalisierung besser gerecht.

Betrachtet man die Ergebnisse dieser quantitativen Untersuchungen, so erreichen sie in mehreren Fällen beachtlich hohe Werte in Bezug auf die Relevanz einzelner Merkmale oder Merkmalsbündel für die Flüssigkeitsbewertung (zu einer ausführlicheren Zusammenfassung vgl. Reitbrecht 2017, 118–121): In der Studie Bosker/Pinget/Quené/Sanders/de Jong (2013, 169) konnten durch ein Modell bestehend aus sechs Flüssigkeitsparametern 83,8 Prozent der Varianz in den Flüssigkeitsbewertungen erklärt werden. Der Zusammenhang zwischen Sprechrate und perzipierter Flüssigkeit liegt in mehreren der oben genannten Studien der ersten Phase über $r=0,8$. Allerdings belegen keinesfalls alle Studien die Relevanz der Sprechrate in diesem Ausmaß. In der Untersuchung Derwing/Munro/Thomson/Rossiter (2009), in der die Ergebnisse nach zwei Lernenden-gruppen und mehreren Untersuchungszeitpunkten unterschieden werden, schwanken die Korrelationskoeffizienten zum besagten Zusammenhang zwischen Sprechrate und Flüssigkeitsbewertung beispielsweise von $r=0,261$ bis $r=0,850$. In der Studie Préfontaine/Kormos/Johnson (2016) ging im Gegensatz zu anderen Studien eine längere mittlere Pausendauer mit einer positiveren Flüssigkeitsbewertung einher.

Diese Divergenzen zwischen den Forschungsergebnissen der einzelnen Studien führten innerhalb der Forschungslinie zu der Frage nach möglichen weiteren Einflussfaktoren in Bewertungssituationen (vgl. Préfontaine/Kormos/Johnson 2016, 66). Préfontaine/Kormos (2016) greifen diese Frage explizit auf und vollziehen damit den Wechsel zu einer dritten Phase, nämlich den Wechsel vom quantitativen zum qualitativen Forschungsparadigma, indem sie an der Stelle von numerischen Bewertungsdaten Kommentare von drei Hörer*innen zu 120 L2-Sprechbeiträgen elizitieren.¹ Die Analyse dieser Kommentare zeigt deutlich, dass die Hörer*innen neben den in den quantitativen Designs berücksichtigten Parametern und Merkmalen auch weitere Aspekte thematisieren, welche die Autorinnen in folgende Gruppen (Häufigkeit der Thematisierungen in Klammern) zusammenfassen und dabei in den weiteren Ausführungen ihres Artikels die hier kursiv gesetzten Kategorien näher beleuchten und mit Beispielen aus dem Datenmaterial belegen: *Speed* (81), *Pause Phenomena* (172), *Lexical Retrieval* (60), *Self-correction* (48), *Efficiency/Effortlessness* (406), *Rhythm* (201), *Expressivity/Psychological state* (190), *Grammatical competence* (140), *Native-like oral discourse features* (136) (Préfontaine/Kormos 2016, 156). In der weiteren Diskussion ihrer Ergebnisse stützen die Autorinnen ihre Argumentation für qualitative

¹ Vergleichbar einzuordnen ist die Studie Hayes-Harb/Hacking (2015). Diese widmet sich der Bewertung von Aussprache, kombiniert in einem Mixed-Methods-Ansatz numerischer Bewertungsdaten mit Interviewdaten zur Begründung der Bewertungen und belegt, dass auch reliablen numerischen Bewertungsdaten individuell sehr unterschiedliche Bewertungsbegründungen zugrunde liegen können.

Untersuchungen vor allem auf zwei Aspekte, nämlich einerseits die Komplexität von Wahrnehmung und das komplexe Interagieren qualitativer und quantitativer Merkmale und andererseits auf die Bedeutung des Rhythmus im Vergleich zu einer rein geschwindigkeitsbezogenen Operationalisierung von Sprechflüssigkeit im Parameter der Sprechrate (Préfontaine/Kormos 2016, 165 f.; siehe dazu auch die Häufigkeiten der Nennungen zu *Speed* und *Rhythm* oben). Weitere interessante Punkte bleiben durch diese Argumentation und thematische Fokussierung allerdings verdeckt: So gehen die Autorinnen nicht näher darauf ein, dass (1) die häufigen Nennungen grammatikalischer Phänomene auf ein durchaus noch breiteres Flüssigkeitsverständnis der Hörer*innen hindeuten und (2) die Hörer*innen von den Sprechdaten ausgehend auch Schlüsse auf die kognitiven Aktivitäten der Sprecher*innen beim Sprechen in der Fremdsprache (siehe explizit *Lexical retrieval*) sowie auch auf die Sprecher*innen selbst (siehe *Expressivity/Psychological state*) ziehen, und dies angesichts der 60 beziehungsweise 190 ermittelten Nennungen nicht gerade selten.

Dieser zweite Punkt, der in Bezug auf Situationen des Testens und Prüfens mündlicher Leistungen von Relevanz ist, für den der Beitrag von Préfontaine/Kormos (2016) aber keine näheren Aussagen zulässt, wird im Folgenden anhand eines eigenen Datensatzes zu Flüssigkeitsbewertungen im Deutschen als Fremdsprache näher beleuchtet.

2. Studie zu Flüssigkeitsbewertungen im Kontext von Deutsch als Fremdsprache

2.1. Aufbau der Studie

Der hier gezeigte Datensatz von Flüssigkeitsbewertungen bezog sich auf insgesamt neun Sprechbeiträge (Bildergeschichtenerzählungen, Audiomaterial) von französischen beziehungsweise tschechischen Deutschstudierenden mit einer gesamten Deutschlerndauer von 6 bis 14 Jahren. Die an der Studie teilnehmenden Hörer*innen waren Studierende mit Erstsprache Deutsch in Lehrveranstaltungen zu Deutsch als Fremd- und Zweitsprache. Insgesamt 128 Studierende bewerteten einen oder zwei Sprechbeiträge, sodass in Summe 229 Bewertungen vorlagen.

Diese Bewertungen umfassten sowohl eine Einschätzung zur Flüssigkeit der gehörten Beiträge auf einer sechs-stufigen Likert-Skala als auch verbale Begründungen, die zusätzlich zur numerischen Bewertung mittels Fragebogen elizitiert wurden. Eingebettet war dieser Bewertungsschritt in eine umfassendere Sprechwirkungsstudie (Reitbrecht 2017), im Zuge derer die Hörer*innen die Sprechbeiträge mehrfach hörten und unterschiedliche Bewertungsaufgaben erhielten. In dem Fall der hier gezeigten Daten lautete die konkrete Frage auf dem Fragebogen: *Wie wirkt der Redebeitrag auf Sie?*. Neben *flüssig* wurden in der

nachfolgenden Likert-Skala auch die Kategorien *kohärent, schnell, grammatikalisch korrekt, inhaltlich verständlich und nachvollziehbar, variantenreich im Ausdruck, einfach im Satzbau, gut vorbereitet und monoton* zwischen den Polen *trifft vollkommen zu* und *trifft überhaupt nicht zu* bewertet. Zu *kohärent* und *flüssig* wurden mittels der Zusatzfrage *warum?* zudem die oben genannten verbalen Begründungen elizitiert. In der folgenden Auswertung werden nur die Kommentare zur Begründung für das Bewertungskriterium *flüssig* näher ausgewertet, die numerischen Bewertungen sowie die anderen Kategorien angesichts der hier konkret verfolgten Ziele nicht näher dargestellt.

2.2. Ergebnisse der Studie

Betrachtet man die Verbalkommentare in ihrer allgemeinen Qualität, so wird erkennbar, dass trotz der konkret zum Sprechbeitrag gestellten Fragestellung die Antworten in vielen Fällen auf die Sprecher*innen und ihr Sprechhandeln selbst referieren, z. B. [1] *spricht schnell* / [2] *macht viele »em« Pausen*. Interessant sind für die Ergebnisdiskussion auch Aussagen zu einer als weitgehend flüssig eingestuften Sprecherin, die Negationen beinhalten, z. B. [3] *kennt viele Füllwörter, stockt nicht, macht keine zu langen Pausen* / [4] *überlegt nicht lange vor jedem Wort*.

Zudem zeigen sich in mehreren Fällen nicht nur Aussagen über Merkmale und Phänomene, die an der sprachlichen Oberfläche erkennbar werden, sondern auch Rückschlüsse auf die zugrunde liegenden kognitiven Aktivitäten der L2-Sprecher*innen und ihr strategisches Vorgehen im Umgang mit der Aufgabenbewältigung und Herausforderungen beim Sprechen in der Fremdsprache (zu einer Auswahl siehe Abbildung 2).

- [5] *spricht langsam, stockend, überlegt viel*
- [6] *kaum Pausen, lässt sich von Fehlern nicht aufhalten*
- [7] *tw. stockend (überlegt nach gramm. Strukturen, diese dann meist richtig)*
- [8] *hat sich trotz Fehler nicht wirklich von der Geschichte beirren lassen*
- [9] *viele ähs, Sprecherin muss viel über das nächste Wort nachdenken*
- [10] *weil sie ihre Differenzen zwischen 1. & 2. Sprache gut kaschiert*
- [11] *gedankl. Fluss unbedingt bemerkbar, aber das Suchen nach den richtigen Worten verzögert den Redefluss (äh...) > Hesitationspartikel*
- [12] *macht ziemlich viele Pausen (weil sie keine Fehler machen will?)*

Abb. 2: Belege für Rückschlüsse auf kognitive Aktivitäten und strategisches Vorgehen der L2-Sprecher*innen

Darüber hinaus erfolgen in acht Kommentaren (von sieben verschiedenen Hörer*innen) Zuschreibungen, die über diese kognitive Dimension hinausgehen und auf die Sprecherpersönlichkeit oder die emotionale Befindlichkeit der Sprecher*innen abzielen (siehe Abbildung 3), wobei zu [14] anzumerken ist, dass sich

zu *hektisch* auch auf die Sprechweise beziehen kann, und das Beispiel [15] klar in Nähe zu den oben genannten Aussagen über die kognitiven Aktivitäten steht.

- [13] *wirkt bedacht, macht Pausen zwischen Wörtern*
[14] *flüssig, aber viele »ah's«, zu hektisch; wirkt abgehackt*
[15] *vor allem zu Beginn sehr flüssig, »wissend, wo sie mit dem Satz hinwill«*
[16] *sie spricht in einem durch (bemüht sich)*
[17] *viele Lückenfüller (ähm), nervös oder unsicher?*
[18] *Sie wirkt nervös und nicht so versiert im freien Sprechen, dadurch kommt auch viel von der eigentlichen Kompetenz abhanden.*
[19] *trotzdem d. Sprecherin hörbar angespannt / konzentriert ist*
[20] *geübte Sprecherin auch wenn sie Fehler macht*

Abb. 3: Zuschreibungen an die Sprecher*innen

2.3. Einordnung der Ergebnisse im Sinne eines konstruktivistischen Verständnisses von Wahrnehmung

Konstruktivistische Modelle und Auffassungen von Wahrnehmung gehen davon aus, dass Wahrnehmung nicht ausschließlich bottom-up gesteuert ist, also nach einem einheitlichen Reiz-Reaktions-Schema abläuft, sondern dass das wahrnehmende Individuum den Wahrnehmungsprozess und dessen Ergebnisse top-down mitprägt. Exemplarisch wird eine solche konstruktivistische Auffassung von Wahrnehmung im Folgenden anhand des Rezeptionsverständnisses der Sprechwirkungsforschung (vgl. dazu Hirschfeld/Neuber/Stock 2008, 2010) vorgestellt, welches darüber hinaus auch die Wahrnehmung der Sprechenden berücksichtigt und damit für die Diskussion der hier gezeigten Ergebnisse einen weiteren relevanten Aspekt beinhaltet.

Annahmen und Erwartungshaltungen auf Hörer*innenseite

Die Vertreter*innen der Sprechwirkungsforschung lehnen »[e]ine mechanische Ursache-Wirkungs-Relation im Sinne des behavioristischen Reiz-Reaktions-Schemas« (Stock 1982, 8 f) ab und gehen für die interpersonelle Kommunikation davon aus, dass »die bewusst reflektierte Beurteilung des Partners und seiner Äußerung auf Grund von Werten, Normen und Theorien« (Hirschfeld/Neuber/Stock 2010, 63) erfolgt. Wahrnehmung wird demnach als (re-)aktiver Prozess verstanden, den Hörer*innen aktiv mitgestalten und bei dem ihre Vorerfahrungen und Erwartungshaltungen eine zentrale Rolle einnehmen. Als Belege dafür können die Beispiele [3] und [4] genannt werden, in denen die Hörer*innen negative Formulierungen verwenden, in denen eben genau solche Annahmen gegenüber L2-Sprecher*innen zum Ausdruck kommen, sich aber in der Konfrontation mit dem konkreten Sprechmaterial nicht erfüllen. Ebenso sind in diesem Zusammenhang die Ergebnisse von Préfontaine/Kormos (2016) zu nennen. Diese

verdeutlichen, dass Hörer*innen auch andere Begriffskonzepte von Flüssigkeit als in der Fachliteratur angenommen in die Bewertungssituation mitbringen. Diese umfassen beispielsweise auch grammatikalische Korrektheit. Sie nehmen ihre Bewertungen also auch auf Basis anderer Normen, Begriffskonzepte und Bezugsgrößen vor.

Interpretationen im Sinne des Symptomcharakters sprachlicher Zeichen

Des Weiteren beruft sich die Sprechwirkungsforschung in ihrem Verständnis von Kommunikation auf das Bühlersche Organon-Modell (Bühler 1934/1999) und schreibt dabei dem Symptomcharakter des sprachlichen Zeichens eine wichtige Bedeutung in der interpersonellen Kommunikation zu. Wahrnehmung fokussiert nicht nur auf die symbolische Dimension, sondern immer auch auf den Sprechausdruck und die Sprecherpersönlichkeit, also auf das *Wie* und das *Wer* (vgl. Hirschfeld/Neuber/Stock 2008, 778 f), wofür das sprachliche Zeichen aufgrund seines Symptomcharakters ebenfalls Repräsentant sein kann. Es wird also davon ausgegangen, dass das sprachliche Zeichen bei jeder Sprachwahrnehmung auch in Bezug auf den/die Sprechende/n interpretiert werden kann.

Explizit belegen diese Annahme jene Aussagen, in denen die Bewerter*innen auf Basis des Gehörten in ihren begründenden Aussagen Zuschreibungen an die Sprecher*innen vornehmen (siehe Beispiele [13] bis [20]). Aber auch die in Auswahl gezeigten Beispiele [5] bis [12] sind in diesem Zusammenhang zu diskutieren. So ist zwar die Interpretation von Merkmalen der sogenannten Äußerungsflüssigkeit in Bezug auf die kognitive Flüssigkeit, also auf die zugrundeliegenden sprechplanerischen und strategischen Prozesse, in der Wahrnehmung von Flüssigkeit (perzipierte Flüssigkeit) ein nachvollziehbarer und naheliegender Akt, die Zusammenhänge zwischen den Flüssigkeitsdimensionen müssen aber ebenfalls in ihrer Komplexität gedacht werden (siehe zu dieser Diskussion im Fachdiskurs näher Reitbrecht 2017, 22–25). So schreibt Segalowitz beispielsweise einschränkend für das Zusammenspiel von kognitiver, perzipierter und Äußerungsflüssigkeit: »Normally, all three meanings will more or less coincide and be mutually supportive or, at least, people will normally assume this to be the case. Of course, this assumption may not always be true« (2010, 47). Raupach vermerkt, dass Häitationsphänomene nur zeigen, »that the learner was having some problems in his verbal planning, but they do not give evidence of the kind of problem he had« (1983, 201). Retrospektive Daten von Sprecher*innen zu ihren eigenen Sprechbeiträgen belegen zudem, dass bestimmte Reparaturstrategien/-prozesse an der sprachlichen Oberfläche verdeckt bleiben (Kormos 2000) und dass sich Häitationsphänomene auch auf zeitlich weiter entfernte Schwierigkeiten in der Sprachproduktion beziehen können (Reitbrecht 2015).

Betrachtet man ausgehend von dieser Forschungslage die in Abbildung 2 gezeigten Aussagen, so erweisen sich auch diese in mehreren Punkten als durch die Annahmen der Hörer*innen geprägte Interpretationen des Sprachmaterials im Sinne der Symptomhypothese. Ob den unflüssigen Sprechweisen der bewerteten

Sprecher*innen nämlich tatsächlich die Planung grammatikalischer Strukturen [7] oder das Suchen nach dem nächsten Wort [9] zugrunde liegen oder ob Differenzen zwischen Erst- und Zweitsprache gut kaschiert oder aufgrund des Sprachkompetenzniveaus im Sprechplanungsprozess ganz einfach nicht mehr relevant sind [10], kann anhand der Äußerungsflüssigkeit, also der sprachlichen Oberfläche, nicht gesichert erschlossen werden.

3. Diskussion in Bezug auf das Anwendungsfeld standardisierter Sprachtests

Während quantitative Untersuchungen zum Zusammenhang von Äußerungsflüssigkeit und Flüssigkeitsbewertungen für das Anwendungsfeld des standardisierten Testens und Prüfens einen wichtigen Beitrag dahingehend leisten, bestimmte Merkmale und Merkmalsbündel in ihrer Relevanz für Bewertungen gegeneinander abzuwägen, ist der Beitrag der hier gezeigten qualitativen Studien ein ambivalenter:

Einerseits nehmen sie klar eine kritische Perspektive in Bezug auf das Erreichen vollkommener Objektivität ein, obwohl diese in einem Zuge mit Reliabilität und Validität, aber auch weiteren Kriterien wie Praktikabilität, Transparenz, Fairness, Authentizität oder Nützlichkeit als ein Gütekriterium für (standardisierte) Sprachtests genannt wird (vgl. Grotjahn 2000, 308; Glaboniat 2010, 1291; Rossa 2016, 400). Versteht man Objektivität als »die Unabhängigkeit der Testergebnisse von irrelevanten Einflussfaktoren, z.B. von Personen, die den Test durchführen, auswerten und interpretieren« (Rossa 2016, 400), so verdeutlichen die oben gezeigten Verbalkommentare nämlich, dass die Bewertungsergebnisse sehr wohl subjektiv von den Hörer*innen und ihren jeweiligen Ausgangssituationen und Vorgehensweisen in der Auseinandersetzung mit dem Sprachmaterial mitkonstruiert werden – auch wenn diese wie im Fall der Untersuchung Hayes-Harb/Hacking (2015) zu Aussprache einen hohen Grad an Interrater-Reliabilität aufweisen.

Andererseits stoßen qualitative Studien aber auch auf ein Bewusstsein um die Problematik der Objektivität im Kontext standardisierter kommunikativ ausgerichteter Sprachprüfungen. So verweist Glaboniat beispielsweise bereits 1998 auf die Diskrepanz zwischen dem Kriterium der Objektivität und den Zielsetzungen kommunikativ ausgerichteter Sprachtests. Tschirner schreibt, dass »sich mündliche Prüfungen nicht im strengen Sinne objektiv bewerten« (2001, 109) lassen. Betrachtet man aktuelle Testprodukte der internationalen Anbieter *Goethe Institut* (online) und *Österreichisches Sprachdiplom Deutsch* (online) (siehe aber auch Tschirner 2001), so werden zahlreiche Bestrebungen und (Standardisierungs-)Maßnahmen erkennbar, die auf das Erreichen des Gütekriteriums der Objektivität abzielen, ohne mit der folgenden Auflistung einen Anspruch auf Vollständigkeit erheben zu wollen: die Standardisierung in den Aufgabenstel-

lungen, dem Ablaufplan und den Zeitvorgaben bei der Testdurchführung, dem ebenso standardisierten Auswertungssystem mit vorgegebenen Rastern, Kriterien, Deskriptoren, Punkteskalierungen und -verteilungen sowie Regeln für die Punkteermittlung bei den rezeptiven Fertigkeiten. Weitere Maßnahmen, um die Subjektivität der Prüfenden als Einflussfaktor zu reduzieren, stellen Prüfendenschulungen, die regelmäßige Abgabe und Kontrolle von Probeprüfungen, der Einsatz von zwei Prüfenden sowie das Publizieren von Beispielbewertungen dar. Auch das Einräumen von einem Ermessensspielraum bei der Bewertung mit Punkten zeugt von dem Bewusstsein, dass vollkommene Objektivität nicht erreicht werden kann.

In diesem Zusammenhang bieten die oben gezeigten qualitativen Forschungsergebnisse trotz ihres grundsätzlich hinterfragenden, kritischen Charakters Schnittstellen für Maßnahmen im Sinne einer kontinuierlichen Qualitätssicherung, z.B. bei der Durchführung von Prüfendenschulungen oder der Entwicklung begleitender Materialien. Ihr besonderer Mehrwert im Vergleich zu rein numerischen Bewertungen erwächst dabei aus der Tatsache, dass sie Einblick in die Gedankengänge der Bewerter*innen ermöglichen und damit auch konkrete inhaltliche Anknüpfungspunkte offen legen. In Bezug auf das Bewertungskriterium der Flüssigkeit zeigte sich so konkret, dass auch weite Begriffsauffassungen von Flüssigkeit existieren sowie wenig Trennschärfe zwischen den Dimensionen der Äußerungs- und der kognitiven Flüssigkeit gegeben ist. Es lassen sich dadurch also Problemfelder ermitteln, auf die bei der Erstellung von Begleitmaterialien für Prüfende explizit Bezug genommen werden kann. In dem hier konkret gezeigten Fall ist dies vor allem angesichts der Tatsache, dass auch der *Gemeinsame europäische Referenzrahmen für Sprachen* in den Niveaubeschreibungen zu Flüssigkeit (Europarat 2001, 37 f.) nicht konsequent zwischen Äußerungs- und kognitiver Flüssigkeit unterscheidet, von besonderer Wichtigkeit.

Zuletzt muss in diesem Beitrag aber in Bezug auf diese anwendungsorientierte Diskussion der Ergebnisse aus der Forschungslinie zu Flüssigkeitsbewertungen für die Prüfungspraxis – selbstkritisch – die Frage eingeräumt werden, ob sich die Ergebnisse aus den oben referierten Untersuchungen auf tatsächliche Sprachtestsituationen zur Fertigkeit Sprechen umlegen lassen. Die Bewerter*innen bewegten sich in den Studien nämlich klar in einem Raum, in dem ihren Bewertungen keinerlei Konsequenzen für die Sprecher*innen folgten. Auch waren Bewertungsvorgang und Rahmenbedingungen anders gestaltet. Beantworten können diese Frage nur Analysen und Auswertungen tatsächlicher Prüfer*innengespräche, wie sie im Anschluss an mündliche Prüfungen erfolgen. Somit ist diese abschließende Überlegung des Beitrags nicht nur als Reflexionsimpuls innerhalb der oben referierten Forschungslinie zu verstehen, sondern zugleich auch als expliziter Impuls für Sprachprüfungsanbieter, ihr Tun kontinuierlich forschend zu begleiten und Forschungs Kooperationen, in denen entsprechende Fragestellungen aufgegriffen werden können, proaktiv zu initiieren und zu fördern.

Literatur

- Applied Linguistics 30/4 (2009). *Special Issue: Complexity, Accuracy, Fluency (CAF) in Second Language Acquisition Research*.
- Bosker, Hans Rutger/Pinget, Anne-France/Quené, Hugo/Sanders, Ted/de Jong, Nivja H. (2013): »What makes speech sound fluent? The contributions of pauses, speed and repairs«. In: *Language Testing* 30/2, 159–175.
- Bühler, Karl (1934/1999): *Sprachtheorie. Die Darstellungsfunktion der Sprache*. 3. Aufl. Stuttgart: Lucius & Lucius (=UTB für Wissenschaft. Uni-Taschenbücher, Bd. 1159).
- De Jong, Nel/Perfetti, Charles A. (2011): »Fluency Training in the ESL Classroom: An Experimental Study of Fluency Development and Proceduralization«. In: *Language Learning* 61/2, 533–568.
- Derwing, Tracey M./Munro, Murray J./Thomson, Ronald I./Rossiter, Marian J. (2009): »The relationship between L1 fluency and L2 fluency development«. In: *SSLA* 31, 533–557.
- Derwing, Tracey M./Rossiter, Marian J./Munro, Murray J./Thomson, Ron I. (2004): »Second Language Fluency: Judgments on Different Tasks«. In: *Language Learning* 54/4, 655–679.
- Europarat, Hg. (2001): *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Übersetzt von Jürgen Quetz. Berlin u. a.: Langenscheidt.
- Glaboniat, Manuela (1998): »Kommunikatives Testen: Ein Widerspruch? – Merkmale und Grenzen kommunikativer Tests«. In: *ÖDaF-Mitteilungen* 2/1998, 27–38.
- Glaboniat, Manuela (2010): »Sprachprüfungen für Deutsch als Fremdsprache«. In: Hans-Jürgen Krumm, Christian Fandrych, Britta Hufeisen, Claudia Riemer, Hg.: *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. 2. Band*. Berlin ect.: Walter de Gruyter (=Handbücher zur Sprach- und Kommunikationswissenschaft, Bd. 35.2), 1288–1298.
- Goethe Institut. *Deutschprüfungen*. <<https://www.goethe.de/de/spr/kup/prf.html>> [Stand: 4.11.2017].
- Grotjahn, Rüdiger (2000): »Testtheorie: Grundzüge und Anwendungen in der Praxis«. In: Armin Wolff, Harald Tanzer, Hg.: *Sprache – Kultur – Politik. Beiträge der 27. Jahrestagung Deutsch als Fremdsprache vom 3.–5. Juni 1999*. Universität Regensburg: Fachverband Deutsch als Fremdsprache (=Materialien Deutsch als Fremdsprache, Bd. 53), 304–341.
- Hayes-Harb, Rachel/Hacking, Jane (2015): »Beyond rating data. What do listeners believe underlies their accentedness judgements?«. In: *Journal of Second Language Pronunciation* 1/1, 43–64.
- Hirschfeld, Ursula/Neuber, Baldur/Stock, Eberhard (2010): »Phonetische Sprechwirkungsforschung im Bereich der interkulturellen Kommunikation. In: Ursula Hirschfeld, Eberhard Stock, Hg.: *Sprechwissenschaftlich-phonetische Untersuchungen zur interkulturellen Kommunikation Russisch-Deutsch*. Frankfurt am Main: Peter Lang (=Hallesche Schriften zur Sprechwissenschaft und Phonetik, Bd. 33), 43–67.
- Hirschfeld, Ursula/Neuber, Baldur/Stock, Eberhard (2008): »Sprach- und Sprechwirkungsforschung«. In: Ulla Fix, Andreas Gardt, Joachim Knepe, Hg.: *Rhetorik und Stilistik. Ein internationales Handbuch historischer und systematischer Forschung*.

1. Band. Berlin ect.: de Gruyter (=Handbücher zur Sprach- und Kommunikationswissenschaft, Bd. 31.1), 772–786.
- Kormos, Judit (2000): »The Timing of self-repairs in second language speech production«. In: *SSLA* 22, 143–167.
- Kormos, Judit/Dénes, Mariann (2004): »Exploring measures and perceptions of fluency in the speech of second language learners«. In: *System* 32, 145–164.
- Osborne, John (2011): »Oral learner corpora and the assessment of fluency in the Common European Framework«. In: Ana Frankenberg-Garcia, Lynne Flowerdew, Guy Aston, Hg.: *New Trends in Corpora and Language Learning*. London: Bloomsbury Publishing, 181–197.
- Österreichisches Sprachdiplom Deutsch. <www.osd.at> [Stand 4.11.2017].
- Pinget, Anne-France/Bosker, Hans Rutger/Quené, Hugo/de Jong, Nivja H. (2014): »Native speakers' perceptions of fluency and accent in L2 speech«. In: *Language Testing* 31/3, 349–365.
- Préfontaine, Yvonne/Kormos, Judit (2016): »A qualitative analysis of perceptions of fluency in second language French«. In: *IRAL* 54/2, 151–169.
- Préfontaine, Yvonne/Kormos, Judit/Johnson, Daniel Ezra (2016): »How do utterance measures predict raters' perceptions of fluency in French as a foreign language?«. In: *Language Testing* 33/1, 53–73.
- Raupach, Manfred (1983): »Analysis and evaluation of communication strategies«. In: Claus Færch, Gabriele Kasper, Hg.: *Strategies in Interlanguage Communication*. London und New York: Longman, 199–209.
- Reitbrecht, Sandra (2017): *Häsitationsphänomene in der Fremdsprache Deutsch und ihre Bedeutung für die Sprechwirkung*. (Schriften zur Sprechwissenschaft und Phonetik 10). Berlin: Frank & Timme.
- Reitbrecht, Sandra (2015): »Es war eigentlich mehr so was wie ein totaler Stau im Kopf«. Retrospektive Daten zum Sprechen in der L3 Französisch. Vortrag bei der ÖGSD-Tagung. Salzburg, 20.11.2015.
- Rose, Ralph (2015): »Temporal variables in first and second language language speech and perception of fluency«. In: The Scottish Consortium for ICPHS 2015, Hg.: *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. URL: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0405.pdf> [Stand: 05.11.2017].
- Rossa, Henning (2016): »Testen und Prüfen«. In: Eva Burwitz-Melzer, Grit Mehlhorn, Claudia Riemer, Karl-Richard Bausch, Hans-Jürgen Krumm, Hg.: *Handbuch Fremdsprachenunterricht*. 6., völlig überarbeitete und erweiterte Auflage. Tübingen: A. Francke Verlag, 399–403.
- Riazaantseva, Anastasia (2001): »Second language proficiency and pausing«. In: *SSLA* 23, 497–526.
- Rossiter, Marian J. (2009): »Perceptions of L2 Fluency by Native and Non-native Speakers of English«. In: *The Canadian Modern Language Review* 65/3, 395–412.
- Segalowitz, Norman (2010): *Cognitive Bases of Second Language Fluency*. New York und London: Routledge.
- Segalowitz, Norman/Freed Barbara F. (2004): »Context, Contact, and Cognition in Oral Fluency Acquisition. Learning Spanish At Home and Study Abroad Contexts«. In: *SSLA* 26, 173–199.

Stock, Eberhard (1982): »Probleme und Ergebnisse der Sprechwirkungsforschung«. In: Eberhard Stock, Hg.: *Sprechwirkungsforschung, Sprecherziehung, Phonetik und Phonetikunterricht*. Halle (Saale) (=Martin-Luther-Universität Halle-Wittenberg. Wissenschaftliche Beiträge, Bd. 1982/55), 3–20.

Tschirner, Erwin (2001): »Die Evaluation fremdsprachlicher mündlicher Handlungskompetenz: Ein Problemaufriss«. In: *FLuL* 30, 87–115.

Sandra Reitbrecht lehrt und forscht am Didaktikzentrum für Text- und Informationskompetenz der Pädagogischen Hochschule Wien. In ihrer Dissertation beschäftigte sie sich mit der Bedeutung von Häsitationsphänomenen für die Sprechwirkung in der Fremdsprache Deutsch, zudem blickt sie auf Lehr- und Fortbildungstätigkeiten für verschiedene Institutionen im In- und Ausland zurück. Ihre Arbeitsschwerpunkte sind Phonetik und Aussprachedidaktik sowie Sprech- und Schreibdidaktik im mehrsprachigen Kontext.

sandra.reitbrecht@phwien.ac.at