

Subjektive Theorien als Faktor bei der Beurteilung fremdsprachlicher Kompetenzen

Ulrike Arras

Subjektive Theorien in einem weitgefassten Sinne, d. h. unter Einschluss von Wissensbeständen, Erfahrungen etc., spielen eine zentrale Rolle bei der Wahrnehmung, Einordnung und Beurteilung fremdsprachlicher Kompetenzen: Dies zeigen die qualitativen Daten aus einer empirischen Untersuchung, die anhand von introspektiven Verfahren während der Beurteilung von schriftlichen Prüfungsleistungen erhoben wurden. Der Beitrag schlägt vor, die Funktion der subjektiven Theorien bei der Beurteilung fremdsprachlicher Kompetenzen systematisch in die Aus- und Weiterbildung von Fremdsprachenlehrkräften und BeurteilerInnen von Prüfungsleistungen zu integrieren, um selbstreflexiv ein kritisches Bewusstsein für ihren Einfluss und ihre Rolle auszubilden und damit zur Qualitätssicherung bei der Leistungsbeurteilung beizutragen.

1. Problemaufriss

Die Erforschung subjektiver Theorien im Kontext Fremdsprachenunterricht ist relativ neu. In den letzten beiden Jahrzehnten sind jedoch in Fremdsprachendidaktik und Sprachlehrforschung etliche empirische Arbeiten auf diesem Gebiet entstanden, die zeigen, dass subjektive Theorien eine zentrale Rolle spielen beim Lehren und Lernen fremder Sprachen – seitens der Lehrkräfte ebenso wie seitens der Lernenden. Aus der deutschsprachigen Forschung seien exemplarisch genannt die Arbeiten zu subjektiven Theorien zum Fremdsprachenlernen (Kallenbach 1996; Schmelter 2004) sowie zum Selbstkonzept von Lehrkräften (Duxa 2001; Caspari 2003). Dabei wird deutlich, dass subjektive Theorien einen bedeutenden Einfluss, ja eine handlungsleitende Funktion haben und damit einen wichtigen Faktor im Kontext Fremdsprachenunterricht darstellen, denn didaktische Entscheidungen wie die Erstellung oder Auswahl von Lehrmaterialien, das *timing* und die Gestaltung des Unterrichts sind eng verknüpft mit zugrunde liegenden Wissensbeständen (Erfahrungswissen, deklaratives Wissen), mit Überzeugungen und subjektiven Theorien. Es ist naheliegend, dass auch die Beurteilung fremdsprachlicher Leistungen – das tägliche Brot aller Lehrkräfte – von solchen Faktoren gesteuert wird. Die Erforschung dieses Referenzsystems ist also von Bedeutung und sollte im Hinblick auf das Lehren und Lernen fremder Sprachen eine besondere Rolle nicht zuletzt im Kontext der Ausbildung spielen (Grotjahn 1998).

Die Studie, die dem vorliegenden Beitrag zugrunde liegt (Arras 2007), weist in diese Richtung: Sie widmet sich den Strategien und Prozessen bei der Beurteilung von Prüfungsleistungen, also der Frage, welche Faktoren die Beurteilung(sarbeit) beeinflussen. Dabei erweisen sich unsere, das Testkonstrukt konstituierenden, Erwartungen, unsere Erfahrungen und subjektiven Theorien als ausschlaggebend. Wie sich diese konstituieren können, will ich im Folgenden darstellen. Ich werde zunächst eine Annäherung an den Begriff subjektive Theorie vornehmen, um sodann aus den mannigfaltigen Daten der erwähnten Studie Hinweise für die handlungsleitende Funktion subjektiver Theorien bei der Beurteilung fremdsprachlicher (schriftlicher) Prüfungsleistungen anzuführen. Abschließend sollen Perspektiven für die weitergehende Forschung auf diesem Gebiet sowie für die Ausbildung und Schulung von Lehrkräften und BeurteilerInnen aufgezeigt werden.

2. Was sind subjektive Theorien?

Subjektive Theorien sind individuell geprägte kognitive Strukturen, die Überzeugungen, Einstellungen, Theoriewissen, aber insbesondere auch Erfahrungswissen, beinhalten. Der Begriff basiert vor allem auf der in den 70er und 80er Jahren von Groeben und anderen (Scheele & Groeben 1998) im Rahmen des so genannten Forschungsprogramms Subjektive Theorien entwickelten Prämisse, dass der Mensch mit der Fähigkeit des Theoretisierens begabt ist. Er ist in der Lage, rationale Entscheidungen zu treffen, sein Handeln zu reflektieren und dies auch zu kommunizieren. Das Forschungsprogramm Subjektive Theorien gilt im deutschsprachigen Raum als der „theoretisch und methodisch am weitesten ausgearbeitete Ansatz zur Rekonstruktion der Sicht des Subjekts“ (Steinke 1999: 53). Groeben & Erb (1997: 33) verstehen im Kontext dieses Forschungsprogramms Subjektive Theorien als „hochkomplexe Denkstrukturen“.

die vor allem auch Hypothesen(-hierarchien) mit zumindest impliziter Argumentationsstruktur aufweisen und nicht zuletzt das intentionale Fundament von Handlungen (als komplexe Begründung von Zielsetzungen, Überzeugungen über Handlungswirkungen et cetera) darstellen. Das Forschungsprogramm Subjektive Theorien ist daher auch als eine höchst differenzierte Ausarbeitung der *handlungstheoretischen* Perspektive anzusehen, die von den individuellen Bedeutungsdimensionen des jeweiligen Handelns ausgeht, die nicht mehr (sicher) durch Beobachtung von außen abbildbar sind, sondern nur durch Kommunikation mit den Handelnden selbst verstanden werden können (Hervorhebung im Original).

Dieser Ansatz basiert auf einem Subjektmodell des „reflexions-, kommunikations-, rationalitäts- und handlungsfähigen Menschen“ (Groeben & Erb 1997: 33).¹

Prinzipiell wird von einer Strukturparallelität zu wissenschaftlichen Theorien ausgegangen. Demnach definieren subjektive ähnlich wie wissenschaftliche Theorien eine Situation, machen Aussagen zur Vorhersehbarkeit von Sachverhalten, erklären diese und konstruieren und rechtfertigen auf dieser Basis schließlich Handlungsentwürfe: In einer bestimmten Situation X geschieht Y aus diesem und jenem Grund und deshalb handle ich so und nicht anders. Konstruieren wir ein Beispiel: Wenn ich als Deutschlehrerin in China in einer Gruppe erwachsener DeutschlerInnen auf dem Niveau A2 jeden sprachlichen Fehler korrigiere, dann sagen die Lernenden bald gar nichts mehr im Unterricht, weil sie Angst haben, Fehler zu machen und Fehler als etwas Negatives empfinden, das es zu vermeiden gilt. Diese Hypothese stütze ich durch Erfahrungswissen, das ich mir durch Aufenthalte und Lehrpraktika in dieser Kultur erworben habe sowie durch deklaratives Wissen über die chinesische Kultur. Deshalb korrigiere ich nur in bestimmten Fällen und präferiere, wenn möglich, Initiation zur Selbstkorrektur.

Zusammenfassend können folgende Merkmale und Qualitäten subjektiver Theorien genannt werden:

- Sie sind (zumindest z. T.) bewusstseinsfähig und damit verbalisierbar.
- Sie weisen in gewisser Hinsicht Strukturparallelität mit wissenschaftlichen Theorien auf, vor allem haben sie ähnlich wie wissenschaftliche Theorien die Funktionen, Prognosen zu entwickeln, Handlungen zu planen, diese zu erklären und zu rechtfertigen.
- Ihre potenziell verhaltens- und handlungssteuernde Funktion spielt zusammen mit anderen Faktoren wie Persönlichkeitsmerkmalen.
- Sie beziehen sich damit auf subjektiv erhobene Daten und subjektiv entwickelte Hypothesen.
- Daraus ergibt sich schließlich auch ihre Stabilität, d. h. ihre Gültigkeit wird gestützt durch subjektive Erfahrungen.
- Unter anderem deshalb sind sie unter Umständen bedeutender für unser Handeln als theoretisches Wissen, mit dem Lehrkräfte durch Studium, Lektüre und Fortbildungsmaßnahmen konfrontiert werden.

Auffällig ist, dass in der Fachliteratur der Begriff subjektive Theorie unterschiedlich verwendet wird, teils wird er im Sinne des genannten Forschungsprogramms gebraucht, teils aber wird eine eher weit gefasste Begriffsdefinition zugrunde gelegt, die sich auf subjektive Daten bzw.

¹ Aus diesem Ansatz resultiert die Entwicklung so genannter Dialog-Konsens-Methoden, mit denen versucht wird, eine adäquate – d. h. verstehende, valide – Beschreibung zu ermöglichen und zwar eine „Beschreibung dieser zum Teil eben höchst komplexen Intentionalität des individuellen Handelnden.“ (Groeben & Erb 1997: 33).

subjektive Hypothesen bezieht – so auch üblich in der Sprachlehrforschung. Ich schließe mich im Folgenden dem eher weit gefassten Begriffsverständnis an. Zudem zeigt die Diskussion in der Fachwelt die folgenden Problemfelder, die ich hier nur knapp skizzieren möchte:

- Inwieweit besteht tatsächlich Strukturparallelität zu wissenschaftlichen Theorien? Wie kann sie nachgewiesen werden?
- Wie kann die Validität der rekonstruierten subjektiven Theorien sichergestellt werden? Ist die kommunikative Validierung ausschlaggebend? Oder bedarf es darüber hinaus einer Handlungsvalidierung wie das genannte Forschungsprogramm vorschlägt? Und inwiefern ist dies methodisch überhaupt möglich und notwendig?

Einig ist man sich darin, dass subjektive Theorien eine handlungsleitende Funktion haben. Dies zeigen gerade die Studien mit explorativ-qualitativen Designs zur Erforschung subjektiver Theorien im Bereich Fremdsprachenlehren und -lernen, die seit den 90er Jahren in Deutschland entstehen. Sie weisen nach, dass didaktische Entscheidungen wie die Erstellung oder Auswahl von Lehrmaterialien oder die Gestaltung von Instruktion eng verknüpft sind mit zugrunde liegenden Wissensbeständen, mit Überzeugungen, aber auch mit Emotionen.

3. Subjektive Theorien bei der Beurteilung fremdsprachlicher Leistungen: Erkenntnisse aus einer empirischen Studie

Um spezifische Strategien und Prozesse bei der Beurteilung von Prüfungsleistungen im Kontext der Prüfung TestDaF (Test Deutsch als Fremdsprache) zu eruieren, wurden vier standardmäßig geschulte und erfahrene TestDaF-Beurteilerinnen in Fallstudien bei der Bewertung schriftlicher Leistungen aus der Prüfung TestDaF beobachtet. Dabei wurde deutlich, dass subjektive Theorien, subjektive, aber auch institutionalisierte Konzepte des Testkonstrukts, Erfahrungen, Erwartungen etc. die Beurteilungsarbeit maßgeblich prägen. Im Folgenden werde ich zunächst kurz den TestDaF sowie das Beurteilungsverfahren skizzieren und das Untersuchungsdesign erläutern.

3.1 Der TestDaF

Der Test Deutsch als Fremdsprache ist eine seit 2001 administrierte standardisierte Prüfung, mit der ausländische Studierende nachweisen können, dass sie über die erforderlichen sprachlichen Kompetenzen zum Studium an einer deutschsprachigen Hochschule verfügen. Die Überprüfung erfolgt anhand von vier Subtests zur Messung der Teilkompetenzen Leseverstehen, Hörverstehen, Schriftlicher Ausdruck und Mündlicher Ausdruck. Der TestDaF wird weltweit in nahezu 80 Ländern an eigens lizenzierten Testzentren dezentral durch-

geführt.² Die Aufgaben werden zentral vom TestDaF-Institut in Deutschland entwickelt, die Leistungen ebenfalls zentral in Deutschland bewertet, wobei die Leistungen aus den produktiven Prüfungsteilen von geschulten BewerterInnen beurteilt werden.³ Sie gehen kriterienorientiert vor, d. h. die individuelle Leistung wird im Hinblick auf vorgegebene Bewertungskriterien in Form skalierter Deskriptoren, die das Testkonstrukt widerspiegeln, beurteilt. Das Kriterienraster sieht eine Kombination aus holistischen und analytischen Kriterien vor, um sowohl die inhaltlichen und sprachlichen Aspekte des Textes als auch die kommunikative Qualität und „Hochschultauglichkeit“ der schriftlichen Leistung erfassen zu können.⁴ Die insgesamt neun Beurteilungskriterien sind gleich gewichtet. Das Raster besteht aus skalierten Deskriptoren, die bei der Beurteilung der individuellen schriftlichen Leistungen anzulegen sind.⁵

Das TestDaF-Institut stellt den BeurteilerInnen verschiedene Instrumente für die Beurteilungsarbeit bereit, um ihnen dabei zu helfen, ihre Maßstäbe anzugleichen und so sicherzustellen, dass die Beurteilungen weitgehend reliabel sind. Neben dem definierten Testkonstrukt, den Schulungen und den Bewertungskriterien werden testsatzspezifisch Kalibrierungen vorgenommen, um die BeurteilerInnen auf ihre Arbeit vorzubereiten. Außerdem erhalten die BeurteilerInnen regelmäßig Rückmeldung zu ihrem individuellen Strengemaß und ihrer Konsistenz. Insofern sind die BeurteilerInnen befähigt, ihre Bewertungen rational und begründet auf die zur Verfügung gestellten Instrumente zu gründen. Dennoch zeigen die Daten aus der genannten Untersuchung, dass auch emotionale und affektive Faktoren die Beurteilungsarbeit und Entscheidungen zu einem nicht unerheblichen Teil prägen. Nicht zuletzt ist es die subjektive Prägung von Erfahrungen und Erwartungen, die der Beurteilungsarbeit unterliegen, was ggfs. zu schwachen Interrater-Reliabilitäten führt.

Der Prüfungsteil Schriftlicher Ausdruck besteht aus einer komplexen materialgeleiteten Texterstellungsaufgabe, bei der die Prüflinge unter Beweis stellen sollen, inwiefern sie im Hochschulkontext wichtige Schreibhandlungen – Beschreiben und Argumentieren – angemessen umsetzen können. Der zu

2 Im Jahre 2007 zählte das TestDaF-Institut 378 lizenzierte Testzentren in 79 Ländern, s. TestDaF-Institut (2008: 47).

3 Eine überblicksartige Beschreibung des TestDaF geben beispielsweise Althaus (2004) und Arras (2006).

4 Zur Begriffsbestimmung holistisch versus analytisch s. die Diskussion in Arras (2007: 80ff).

5 Das Beurteilungsverfahren sowie die Schulungskonzeption des TestDaF-Instituts werden ausführlich dargestellt in Arras (2007). Die testmethodische Unterstützung durch Multifacetten-Rasch-Analysen zur Ermittlung des so genannten fairen Durchschnitts erläutert Eckes (2003, 2004, 2005). Basisinformationen zu Konzept, Inhalten, Aufgabenstruktur und Beurteilungsprinzipien sind zudem auf der Internetseite des TestDaF-Instituts abrufbar: www.testdaf.de.

erstellende Text soll eine rezeptionsfreundliche und logische Struktur aufweisen.⁶

3.2 Das Untersuchungsdesign

Die Hauptuntersuchung bestand aus vier Fallstudien. Ihr waren zwei kleinere Vorstudien zur Eingrenzung des Gegenstandes und zur Erprobung geeigneter Methoden vorgeschaltet. Die Hauptuntersuchung selbst bediente sich introspektiver Verfahren: Die vier Versuchspersonen (geschulte, erfahrene TestDaF-Beurteilerinnen) beurteilten je acht schriftliche Leistungen aus einer TestDaF-Prüfung unter Laut-Denken-Bedingungen. Am Tag darauf wurden in retrospektiven Interviews Einzelheiten der Beurteilungsarbeit gezielt nachgefragt, um so aus der Rückschau weitere Daten zum Vorgehen und zu den Beweggründen zu ermitteln.

Die Laut-Denken-Protokolle wurden transkribiert, segmentiert und kodiert, um die Fülle der Daten handhaben zu können. Bei den Daten der retrospektiven Interviews genügte eine einfache Transkription für eine Inhaltsanalyse.⁷

3.3 Befunde

Die Verbaldaten der Studie (Laut-Denken-Protokolle sowie retrospektive Interviews) geben etliche Hinweise darauf, dass subjektive Theorien im weiteren Sinn, also auch Annahmen, Einstellungen, Erwartungen, persönliche Vorlieben und Erfahrungen die Beurteilung und den Beurteilungsprozess maßgeblich steuern. Sie sind Teil diverser Beurteilungsstrategien, beispielsweise wenn eine Einschätzung damit begründet wird, dass ähnliche „Fälle“ aus der Berufspraxis bekannt sind. Hier werden strategisch Bezüge und Vergleiche hergestellt zwischen den zu beurteilenden Leistungen und konkreten Erfahrungen und Beobachtungen aus anderen, ähnlichen Zusammenhängen. Freilich sind es gerade diese subjektiven Zugänge, die zu unterschiedlichen Einschätzungen derselben Leistung führen können. Dies soll an einem Beispiel gezeigt werden. Grundsätzlich sind alle neun Bewertungskriterien, wie betont, gleich gewichtet. Dennoch legen die Versuchspersonen einen je unterschiedlichen Wert auf bestimmte Aspekte der Leistung. Dies geschieht nicht arbiträr, sondern wird wohl begründet, wie den Verbaldaten zu entnehmen ist. Beispielsweise gibt Versuchsperson K6 während der Beurteilung von Text 6 an „... die Behandlung der Aufgabe ist ja das Wichtigste, wenn man das liest, ne?“ (Seg. 276). In Text 6 wird aus ihrer Sicht die Aufgabe nicht angemessen behandelt (Seg. 283: „ist aber einfach einem wissenschaftlichen Text nicht angemessen, was hier, was

hier geschrieben wird, ne?“). K6s subjektive Theorie enthält als subjektives Konstrukt eine konkrete Vorstellung von der Angemessenheit bzw. Hochschultauglichkeit eines schriftlichen Textes. Diese Vorstellung basiert auf dem vom TestDaF-Institut vorgegebenen Testkonstrukt, bezieht aber auch individuelle, berufliche Erfahrungen im Kontext Deutsch als Fremdsprache sowie Erfahrungen aus der eigenen akademischen Sozialisation („disciplinary enculturation“ Pula & Huot 1993) ein. Für das subjektive Konstrukt werden subjektive Belege angeführt, also Beobachtungen, die als Belege für oder gegen die Erfüllung des Testkonstrukts sprechen. Grundsätzlich werden verschiedene Anhaltspunkte gesucht, die die Wahrnehmung und Hypothese verifizieren, etwa Hinweise auf das Alter wie im zitierten Beispiel: Die Person hinter dem Text wird als sehr jung und damit intellektuell als wenig reif wahrgenommen, sie wohnt evtl. noch bei den Eltern, vermutet K6, was ein weiterer Hinweis auf ihr Alter sein könnte. Zudem werden Anhaltspunkte für die sozio-kulturelle Herkunft des Prüflings ausgemacht. Hierbei wird vermutlich auf konkrete Erfahrungen im Umgang mit Studierenden aus bestimmten Kulturen (im vorliegenden Fall Bulgarien, Lateinamerika) rekurriert. Dies führt, das Testkonstrukt beachtend, schlussendlich zur eigentlichen Frage, die mit der Einstufung der Leistung zu beantworten ist: (Seg. 284): „Jo, Frage ist, kann so jemand studieren?“ K6 kommt zu dem negativen Schluss (Seg. 285): „Der muss noch was anderes lernen außer Deutsch. [...]“.

Versuchsperson K8 hingegen gibt im retrospektiven Interview (RIK8 ab 9:30) an: „Lesefluss ist für mich, sag ich mal, ne, ne relativ wichtige Sache.“ Denn,

das hat vielleicht einfach nur mit einem ganz subjektiven Wohlgefühl zu tun. Ich fühl mich einfach wohl, wenn ich einen Text durchgängig lesen kann und kann dem Schreiber folgen, gedanklich auch folgen, das ist für mich einfach ein angenehmeres Gefühl. Als wenn ich dauernd hängen bleibe. Ach, jetzt schon wieder ein Punkt. Ah was will der Dichter damit sagen, ne? Das ist ein subjektives Wohlgefühl einmal. Und dann denk ich ist Lesefluss sicher auch ein Indiz für, ja auch sprachliches Ausdrucksvermögen, denk ich mal, wenn einer sprachlich und gedanklich klar hat und das auch sprachlich rüberbringen kann, dann liest sich das auch automatisch gleich flüssiger. Egal ob das jetzt sprachlich anspruchsvoll ist in der Ausdrucksfähigkeit oder nicht. Es kann auch ganz einfach sein mit, nur mit Nebensätzen oder so was. Ohne große Sachen. Also es ist einfach so ein Wohlgefühl.

K8 hierarchisiert also die Beurteilungskriterien. Der Lesefluss scheint für sie ein den anderen Bewertungsmaßstäben übergeordnetes Kriterium zu sein: Wenn der Lesefluss beeinträchtigt ist, liegt das an gedanklich-logischen oder auch an sprachlichen Mängeln.

⁶ Der Aufgabentyp sowie die Anforderungen sind anhand von Modellprüfungen und Hinweisen zu den Beurteilungsmaßstäben über die Webseite des TestDaF-Instituts einsehbar: www.testdaf.de

⁷ Eine ausführliche Beschreibung der einzelnen Schritte und Analyseverfahren sowie eine Darstellung des Kodier- und Transkriptionssystems finden sich in Arras (2007).

Hingegen erachtet Versuchsperson K7 gerade den Lesefluss als weniger bedeutsam, wie sie im retrospektiven Interview erläutert (RIK7 ab 24:40),

denn das ist ja auch wirklich, [...] ja, ich denk das ist schon subjektiv sehr verschieden, ne? Ob eh, was man als flüssig empfindet und was nicht. Also ich denke, ich hab jetzt viel eh Erfahrung und ich mein, so viel Jahre Deutsch als Fremdsprache unterrichtet, da ist man schon manches gewöhnt [lacht]. Da les ich vielleicht schon flüssig über manche Sachen, wo jemand anders stockt. Aber ich denke, na ja, ein gewisses Stocken kann man auch verkraften, ne? Für die Kommunikation ist das nicht so schlimm. Aber wo ich eigentlich schon ein bisschen streng bin ist mit dem Gedankengang, also wenn ich finde, man kann den Gedankengang nicht verstehen, eh das notier ich schon negativ. Weil das ist ja das Wichtigste, ne? Also das ist für mich ein absolut wichtiges Kriterium, eigentlich, der Gedankengang kann man den, eh versteht man das, oder nicht? Bei der Struktur ist es auch so, ja also ich meine, wir sollen drauf achten. [...] Ich mein, es ist ja auch schön, wenn ein Text ne Einleitung hat und einen Schluss und Übergänge hat und das ist, das ist schon ne Störung beim Lesen, wenn auf einmal was anderes kommt. Also das ist sogar ne enorm, also das stört enorm, ne? Weil man muss ja erst gucken, was kommt denn jetzt. Und man erwartet was ganz anderes, ne? Und dann also das kostet einem schon Zeit, bis man sich dann wieder gefunden hat. Aber, also ich finde es nicht so kritisch wie mit dem Gedankengang. Obwohl ich mein natürlich, je nach dem eh, wenn es ein richtiger Bruch ist, je nach dem wie der Bruch ist, eh wirkt es sich auf den Gedankengang auch aus.

Diese wenigen Zitate zeigen zum einen, dass auf Vorlieben, Erfahrungen und Wissensbeständen gründende subjektive Theorien sowie das durch persönliche Berufserfahrung geprägte Testkonstrukt die Wahrnehmung und Beurteilung fremdsprachlicher Prüfungsleistungen steuern und dass diese Faktoren zu unterschiedlichen Einschätzungen führen können. Die Daten zeigen zudem, dass die Versuchspersonen, gerade wenn sie über intensive einschlägige Erfahrungen als Lehrkräfte, evtl. auch als Lernende fremder Sprachen, verfügen, in der Lage sind, ihre subjektiv geprägten Wahrnehmungen zu reflektieren, sie zu begründen, sie sich also bewusst zu machen. Sie zeigen also prinzipiell die Fähigkeit, mit Hilfe von Selbstreflexion zugrundeliegende Konstrukte und subjektive Theorien zu rekonstruieren, was für eine Professionalisierung gewinnbringend eingesetzt werden könnte, wie im Folgenden erläutert werden soll.

4. Schlussfolgerungen

Die Rekonstruktion subjektiver Theorien bei der Beurteilung fremdsprachlicher Leistungen und die Fähigkeit, diese Rekonstruktion selbstreflexiv zu bewerkstelligen, ist von hoher praktischer Relevanz und sollte für die Aus- und Fortbildung von Fremdsprachenlehrkräften nutzbar gemacht werden. Dies ist möglich, indem Lehrkräfte (und damit BeurteilerInnen fremdsprachlicher Leistungen) zur selbstreflexiven Beobachtung ihrer eigenen Beurteilungsarbeit, ihrer Vorgehensweisen und Beurteilungsstrategien animiert werden (s. hierzu Vorschläge in Arras 2007: 460ff.). Die Erforschung dieses Referenzsystems ist

erforderlich – nicht allein für die Beurteilung von Leistungen, die im Rahmen von standardisierten *High Stakes-Tests* wie dem TestDaF erbracht werden, sondern auch für die Evaluation von Leistungen im Unterrichtsalltag und in informellen Tests (*classroom embedded testing*). Ziel ist dabei nicht nur die Erhöhung der Interrater-Reliabilität und damit die Verbesserung der Fairness bei der Leistungsbeurteilung. Vielmehr erachte ich dies als logische Konsequenz aus den fremdsprachendidaktischen Neuorientierungen der letzten Jahre: Die Implementierung des Gemeinsamen europäischen Referenzrahmens für Sprachen (Europarat 2001) und die Kompetenzorientierung im Fremdsprachenunterricht haben u. a. zur Folge, dass Lernende (deren Leistungen ja stets auch beurteilt werden) mehr Transparenz seitens der Lehrkräfte und BeurteilerInnen erfahren. Sie haben ein Anrecht darauf zu verstehen, wie die Lernziele konzipiert sind und anhand welcher Maßstäbe ihre Leistungen eingestuft werden. Es genügt dabei offensichtlich nicht, anzugeben, welche ggfs. bereits vorliegenden Skalen und Kriterienraster wir bei der Beurteilung fremdsprachlicher Leistungen verwenden. Vielmehr bedarf es seitens der Lehrkräfte einer Bewusstmachung eigener indizierter Maßstäbe.

Da Beurteilung stets diverse subjektiv geprägte Wissen-, Erfahrungs- sowie Wertebestände aktiviert, sollte die Aus- und Weiterbildung von Lehrkräften um eine soziale Dimension erweitert werden. Fremdsprachenlehrkräfte sind demnach als soziale und „reflexive, (potentiell) rationale und kommunikationsfähige Aktanten zu betrachten, die selbst am besten über ihre Kognitionsinhalte (unter Einschluss von Intentionen und Gründen) Auskunft geben können“ (Grotjahn 1998: 50). Insgesamt betrachtet kann festgehalten werden, dass die Beschäftigung mit subjektiven Theorien hierbei folgende Ziele verfolgen sollte:

- Die Rekonstruktion subjektiver Theorien z. B. zum eigenen Lernprozess, zu Lernstrategien in gesteuerten und weniger gesteuerten Kontexten, zu eigenen, durch Erfahrung und Wissen geprägten Lehr- und Lernkonzepten, zur Wirkung von Instruktion und Lehrmaterial, zu den Faktoren, die unsere Wahrnehmung bei der Beurteilung fremdsprachlicher Leistungen bestimmen etc.
- Die praktische Auseinandersetzung mit geeigneten z. B. introspektiven Methoden, um über die Ausbildungsphase hinaus in der Lage zu sein, reflexiv, das sich sicherlich verändernde subjektive Theoriegebäude zu beobachten, zu überprüfen und ggf. Handlungen begründet zu optimieren.
- Die Befähigung zu (metakognitiven) Reflexionen über das eigene Handeln, über die eigenen Lern- bzw. Lehrstrategien bzw. über diejenigen Faktoren, die dieses Verhalten lenken, steuern und determinieren. Für den Bereich Beurteilung fremdsprachlicher Leistungen könnte als Zielvorgabe die Sensibilisierung für Faktoren stehen, die die individuelle Rezeption und Evaluation schriftlicher oder mündlicher Leistungen in der Fremdsprache determinieren; Mit anderen Worten: Die Ausbildung einer *assessment awareness*, also einer „Bewusstheit der eigenen Beurteilungsarbeit, der verwendeten Strategien, der

durchlaufenden Phasen und der Prozesshaftigkeit des Vorgehens auf individueller Ebene“ (Arras 2007: 449).

Die Entwicklung einer Reflexionsfähigkeit, um sich die eigenen – sich auch verändernden – Handlungsmuster, die zugrunde liegenden Wissensbestände, Überzeugungen und Einstellungen bewusst zu machen, sie zu hinterfragen und zu evaluieren und auf dieser Basis schließlich begründet eine Revision und Optimierung der Handlung herbeizuführen, sollte integraler Bestandteil der Professionalisierung von BeurteilerInnen, aber auch von Lehrkräften sein. Zumal die Fähigkeit, das eigene Handeln zu reflektieren, ein wichtiges Instrument darstellt bei der lebenslangen oder zumindest für die Dauer der Berufsausübung währenden Professionalisierung. Auf die besondere Rolle dieses Referenzsystems in der Ausbildung von Fremdsprachenlehrkräften hat Rüdiger Grotjahn allerdings bereits Anfang der 90er Jahre (Grotjahn 1991) hingewiesen.

Literaturverzeichnis

- Althaus, Hans-Joachim (2004), Der TestDaF. In: DAAD (Deutscher Akademischer Austauschdienst) (Hrsg.) (2004): *Die internationale Hochschule: Ein Handbuch für Politik und Praxis, Band 8*. Bielefeld: Bertelsmann, 80-87.
- Arras, Ulrike (2006), Der TestDaF. Konzept und Prinzipien des standardisierten Tests Deutsch als Fremdsprache. In: *Fòrum – Anuari de l'Associació de Germanistes de Catalunya*. Akten des sechsten Kongresses des Katalanischen Deutschlehrer- und Germanistenverbandes (A.G.C.), Tarragona, April 2005, 39-52. [http://www.tinet.org/~asgc2/Forum_2005/Autors/Arras/arras04.html (30.04.08)].
- Arras, Ulrike (2007), *Wie beurteilen wir Leistung in der Fremdsprache? Strategien und Prozesse bei der Beurteilung schriftlicher Leistungen in der Fremdsprache am Beispiel der Prüfung Test Deutsch als Fremdsprache (TestDaF)*. Giessener Beiträge zur Fremdsprachendidaktik. Tübingen: Narr.
- Arras, Ulrike (erscheint), Wie es zu einer Beurteilung kommt. Ein Forschungsbericht zu Strategien bei der Beurteilung schriftlicher Leistungen im Kontext der Prüfung TestDaF. In: Wolff, Armin; Hunstiger, Agnieszka & Koreik, Uwe (Hrsg.) (erscheint), *Chance Deutsch: Schule – Studium – Arbeitswelt*. Beiträge der 34. Jahrestagung DaF 2006. Regensburg: FaDaF.
- Arras, Ulrike (erscheint), What's on a rater's mind? Die Erforschung von Beurteilungsstrategien und ihre Bewusstmachung durch Schulungsmaßnahmen als Voraussetzungen für die Testvalidität, erscheint in den Tagungsakten der AILA-Tagung 2008 in Essen.
- Caspari, Daniela (2003), *Fremdsprachenlehrerinnen und Fremdsprachenlehrer. Studien zu ihrem beruflichen Selbstverständnis*. Tübingen: Narr.
- Duxa, Susanne (2001), *Fortbildungsveranstaltungen für DaZ-Kursleiter in der Weiterbildung und ihre Wirkungen auf das professionelle Selbst der Lehrenden*. Materialien Deutsch als Fremdsprache. Heft 57. Regensburg.
- Eckes, Thomas (2003), Qualitätssicherung beim TestDaF: Konzepte, Methoden, Ergebnisse. *Fremdsprachen und Hochschule* 69, 43-68.

- Eckes, Thomas (2004), Facetten des Sprachtestens: Strenge und Konsistenz in der Beurteilung sprachlicher Leistungen. In: Wolff, Armin; Ostermann, Torsten & Chlosta, Christoph (Hrsg.) (2004): *Integration durch Sprache*. Regensburg: FaDaF, 485-518.
- Eckes, Thomas (2005), Analyse und Evaluation sprachproduktiver Prüfungen beim TestDaF. In: Kühn, Ingrid; Lehker, Marianne & Timmermann, Waltraud (Hrsg.) (2005), *Sprachtests in der Diskussion*. Frankfurt am Main: Lang, 60-93.
- Europarat (2001): *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen*. Berlin et al.: Langenscheidt.
- Groeben, Norbert & Erb, Egon (1997), Menschenbilder. In: Straub, Jürgen; Kempf, Wilhelm & Werbik, Hans (Hrsg.) (1997), *Psychologie. Eine Einführung. Grundlagen, Methoden, Perspektiven*. München: dtv, 17-41.
- Grotjahn, Rüdiger (1991), The Research Programme Subjective Theories: A new approach in second language research. *Studies in Second Language Acquisition* 13, 187-214.
- Grotjahn, Rüdiger (1998), Subjektive Theorien in der Fremdsprachenforschung: Methodologische Grundlagen und Perspektiven. *Fremdsprachen Lehren und Lernen* 27, 33-59.
- Kallenbach, Christiane (1996), *Subjektive Theorien. Was Schüler und Schülerinnen über Fremdsprachenlernen denken*. Tübingen: Narr.
- Pula, Judith J. & Huot, Brian A. (1993), A model of background influences on holistic raters. In: Williamson, Michael & Huot, Brian A. (eds.) (1993), *Validating Holistic Scoring for Writing Assessment: theoretical and empirical foundations*. Cresskill, NJ: Hampton Press, 237-265.
- Scheele, Brigitte & Groeben, Norbert (1998), Das Forschungsprogramm Subjektive Theorien, Theoretische und methodologische Grundzüge in ihrer Relevanz für den Fremdsprachenunterricht. *Fremdsprachen lehren und lernen* 27, 12-32.
- Schmelter, Lars (2004), *Selbstgesteuertes oder potenziell expansives Fremdsprachenlernen im Tandem*. Tübingen: Narr.
- Schocker-von Dittfurth, Marita (2000), *Forschendes Lernen in der fremdsprachlichen Lehrerbildung. Grundlagen, Erfahrungen, Perspektiven*. Tübingen: Narr.
- Steinke, Ines (1999), *Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung*. Weinheim/München: Juventa.
- TestDaF-Institut (Hrsg.) (2008), Jahresbericht 2006/07 [http://www.testdaf.de/institut/pdf/Jahresbericht06_07.pdf (30.04.09)].