

*Ute Feldmann/Rüdiger Grotjahn/Brigitte Stemmer (Ruhr-Universität Bochum)*

## Was messen Sprachtests eigentlich?

### Überlegungen zur introspektiven Validierung von C-Tests

#### 1. Einleitung

In unserem Beitrag möchten wir anhand der Darstellung eines zur Zeit laufenden Forschungsprojekts<sup>1</sup> auf einige grundlegende methodologische Probleme bei der Sprachtestvalidierung eingehen. Zentral für das Projekt ist die Untersuchung der Frage, was der vor kurzem entwickelte und mittlerweile so erfolgreiche C-Test nun eigentlich genau mißt. Wie im folgenden hoffentlich deutlich werden wird, ist eine erschöpfende Beantwortung dieser Frage auf der Basis des herrschenden, psychometrisch-statistisch ausgerichteten Forschungsparadigmas allein nicht möglich.

#### 2. Der C-Test

##### 2.1 Grundlagen und Testformat

Lange Zeit galt bei nicht wenigen Sprachtestern der Cloze Test als besonders geeignetes Instrument zur Sprachstandsmessung. Sieht man vom Einsatz des Cloze Tests zur Diagnose spezifischer Fertigkeitsbereiche einmal ab, dann sind Cloze Tests sog. integrative oder holistische Tests und messen nach fast übereinstimmender Auffassung mehr oder minder global Kompetenz in einer Erst-, Zweit- oder Fremdsprache. Der Cloze Test ist außerdem ein sog. *reduced redundancy test* (vgl. hierzu Klein-Braley 1985; Gaies 1986). Die Häufigkeit richtiger Lösungen gilt als Maß für die Leistungsfähigkeit einer globalen Sprachverarbeitungskompetenz, von Spolsky (1973) "general language proficiency" genannt, bzw. eines Sprachverarbeitungsmechanismus, von Oller (1973) als "pragmatic expectancy grammar" bezeichnet. Die internalisierte Erwartungsgrammatik bildet die Grundlage für die konstruktive, antizipatorische Verarbeitung von Sprache. Der Cloze Test erfaßt nach Oller den Ausprägungsgrad dieser Erwartungsgrammatik und damit auch gleichzeitig die Gesamtkompetenz in einer Sprache (vgl. auch Oller 1976; 1979 sowie die Kritik bei Vollmer/Sang 1980).

In jüngerer Zeit wurde allerdings zunehmend Kritik am Cloze-Verfahren geäußert (vgl. insbesondere Alderson 1978; 1979; Klein-Braley 1981). Es werden u.a. folgende Kritikpunkte genannt:

- 1) Um eine ausreichende Anzahl von Items zu erhalten, muß ein Cloze Test relativ lang sein.

<sup>1</sup> Die Anfänge des Projekts gehen auf das Jahr 1983 zurück. Seit April 1986 wird das Projekt (Leitung: Friedrich Denig, Rüdiger Grotjahn; Mitarbeiterinnen: Ute Feldmann, Brigitte Stemmer) durch die Deutsche Forschungsgemeinschaft finanziell unterstützt.

- 2) Aufgrund des Tilgungsprinzips bestehen Cloze Tests häufig nur aus einem längeren Text; dies kann wegen der Textspezifität zu einer Beeinträchtigung der Testfairneß führen.
- 3) Die Faktoren 'Text', 'Tilgungsrate' und 'Startpunkt' beeinflussen die Reliabilitäts- und Validitätskoeffizienten.
- 4) Bei der exakten Bewertungsmethode sind Cloze Tests sogar für kompetente Muttersprachler häufig zu schwierig. Die Bewertung nach der Akzeptabilität der Lösung ist dagegen in der Regel weit zeitaufwendiger und weniger reliabel.
- 5) Die Schwierigkeit von Cloze Tests hängt vom Häufigkeitsverhältnis der gelöschten Inhalts- und Strukturwörter ab.
- 6) Wie insbesondere die Sekundäranalysen von Klein-Braley (1981) ergeben haben, sind Cloze Tests weniger reliabel als weithin angenommen.

Als Folgerung aus der Kritik stellten Christine Klein-Braley und Ulrich Raatz 1981 eine Modifikation des Cloze Verfahrens vor, den C-Test (vgl. Raatz/Klein-Braley 1982). Der C-Test basiert auf denselben test- und sprachtheoretischen Grundlagen wie der Cloze Test, unterscheidet sich jedoch im Testformat. Er gilt ebenso wie der Cloze Test als Operationalisierung des Konzepts des *reduced redundancy testing* und beruht damit auf folgender Annahme: je kompetenter die Lerner in einer Fremdsprache sind, desto besser können sie Gebrauch von ihrer pragmatischen Erwartungsgrammatik und der natürlichen Redundanz eines Textes machen und desto besser können sie den C-Test lösen. Des Weiteren gilt der C-Test ebenso wie der Cloze Test als integratives Meßinstrument zur Erfassung globaler Sprachkompetenz in Erst-, Zweit- und Fremdsprache.

Um ein höheres Maß an Testfairneß zu erreichen, d.h. um Prüflinge mit Spezialwissen nicht zu bevorzugen, werden beim C-Test im Gegensatz zum Cloze Test mehrere kurze (aus etwa 60 bis 70 Wörtern bestehende) Texte unterschiedlicher Thematik gewählt. Die gewählten Texte sollen jeweils eine Sinneinheit bilden, inhaltsneutral und interessant sein, kein Spezialvokabular enthalten, kein Spezialwissen verlangen, zielgruppenadäquat und so weit wie möglich authentisch sein (vgl. Raatz/Klein-Braley 1985). Um die übrigen in der Kritik am Cloze Test genannten Probleme zu vermeiden, wird folgendes Tilgungsprinzip gewählt: Beginnend mit dem zweiten Wort des zweiten Satzes wird in jedem Text bei jedem zweiten Wort die zweite Hälfte getilgt. Wörter mit einem Buchstaben und Eigennamen bleiben unberücksichtigt. Wortverbindungen mit Apostroph oder Bindestrich wie französisch *l'un, qu'il, faut-il* oder *celui-ci* werden als ein Wort gezählt. Hat ein Wort  $n$  Buchstaben und ist  $n$  ungerade, dann beträgt die Anzahl der zu tilgenden Buchstaben  $(n+1)/2$ , d.h. bei Wörtern mit einer ungeraden Anzahl von Buchstaben ist die Zahl der getilgten Buchstaben um eins höher als die Zahl der nicht getilgten Buchstaben. In jedem Text sollte die gleiche Zahl von Tilgungen vorgenommen werden. Am Textende sollte ein kurzes unversehrtes Textstück als Kontext für die Lösungen stehen bleiben. Die ausgewählten Texte werden in aufsteigender Schwierigkeit angeordnet. Es wird ein Punkt für jede exakte (oder bisweilen auch für jede akzeptable) Rekonstruktion des Originalworts gegeben. Entsprechend dem geschilderten Prinzip besteht ein C-Test meist aus vier bis sechs Texten mit 20 bis 25 Items (Lücken). Ein Beispiel

für einen spanischen C-Test-Text einschließlich der Testinstruktionen findet sich in Abbildung 1.

Nachfolgend finden Sie 6 verschiedene kurze Texte, in denen bei jedem zweiten Wort ein Teil des Wortes fehlt. Bitte versuchen Sie, den fehlenden Teil zu ergänzen.

La vida evoluciona. La sociedad, también. Hoy día existen bienes de consumo que la comunidad entendiendos como imprescindibles para la vida cotidiana doméstica; un frigo\_\_\_\_\_, el ag\_\_\_\_\_, corriente, l\_\_\_\_\_, energía eléc\_\_\_\_\_, no só\_\_\_\_\_, son d\_\_\_\_\_, gran util\_\_\_\_\_, sino q\_\_\_\_\_, su inexis\_\_\_\_\_, impediría e\_\_\_\_\_, desarrollo nor\_\_\_\_\_, de l\_\_\_\_\_, acontecimientos e\_\_\_\_\_, una soci\_\_\_\_\_, moderna. D\_\_\_\_\_, mismo mo\_\_\_\_\_, existen cie\_\_\_\_\_, productos alimen\_\_\_\_\_, de l\_\_\_\_\_, que sie\_\_\_\_\_, hay u\_\_\_\_\_, reserva e\_\_\_\_\_, el ho\_\_\_\_\_, en to\_\_\_\_\_, los hog\_\_\_\_\_. Dentro de este "stock" doméstico a que nos referimos, ocupan lugar preponderante los vinos, aguardientes y licores en proporción variable según el gusto del interesado.

Abb. 1. C-Test Spanisch: Instruktionen und Textbeispiel

## 2.2 Korrelationstatistische Untersuchungen

In bisherigen Untersuchungen zum C-Test wurde der Sprachstand drei verschiedener Gruppen gemessen: die Muttersprache von Kindern, Fremdsprachenkenntnisse von Schülern/Studenten/Sprachkursteilnehmern und die Zweitsprachenkenntnisse von z.B. ausländischen Arbeitnehmern in der Bundesrepublik Deutschland. Es wurden u.a. folgende Sprachen berücksichtigt: Deutsch, Englisch, Französisch, Spanisch, Hebräisch, Türkisch und Finnisch. In allen genannten Sprachen erwies sich das C-Prinzip als anwendbar. In vielen Fällen ergaben sich hohe Reliabilitäten (zum Teil betrug Cronbach's Alpha mehr als 0,9). Dies ist umso erstaunlicher, als viele der untersuchten C-Test-Versionen zum ersten Mal eingesetzt wurden. Auch die Korrelationen der C-Tests mit verschiedenen Außenkriterien, wie z.B. Schulnoten, Lehrerurteilen über den Sprachstand der Schüler sowie mit standardisierten Sprachtests sind zum Teil erstaunlich hoch (vgl. z.B. die Übersicht in Klein-Braley/Raatz 1984). Auch die faktorielle Validität wurde mehrfach untersucht (vgl. z.B. Raatz/Klein-Braley 1983; Raatz 1984; Grotjahn/Stemmer 1985; Grotjahn 1986c).

Der gegenwärtige Forschungsstand läßt sich folgendermaßen zusammenfassen: Der C-Test ist ein leicht zu entwickelndes, ökonomisches und vor allem ein hochreliabels Meßinstrument. Was für eine zugrundeliegende Fähigkeit der C-Test genau mißt, d.h. seine Konstruktvalidität, ist allerdings bisher u.E. nicht ausreichend geklärt. Eine Klärung dieser Frage erscheint uns umso dringlicher, als der C-Test in zunehmendem Maße für praktische Entscheidungen (z.B. zur Leistungsdifferenzierung) eingesetzt wird.

## 3. Einige kritische Bemerkungen zur korrelationsstatistischen Testvalidierung

Die bisherigen Untersuchungen zur Validität von C-Tests beschränken sich - sieht man einmal von Cohen/Segal/Weiss (1984; 1985) und unseren eigenen Untersuchungen ab - auf die Analyse der Korrelationen von C-Test-Werten mit z.B. anderen

Sprachtests, mit Intelligenztests oder auch mit Lehrerurteilen über den Leistungsstand der Schüler in sprachlichen Fächern. Obwohl wir solche Untersuchungen prinzipiell für wichtig halten, kann u.E. mit Hilfe korrelationsstatistischer Untersuchungen allein die Validitätsfrage nicht zufriedenstellend beantwortet werden. Es lassen sich zumindest drei Gründe anführen, warum der korrelationsstatistische Ansatz vor allem bei der Konstruktvalidierung unzureichend ist (vgl. Grotjahn 1986b; c):

- 1) Die Konstruktvalidierung eines Sprachtests an z.B. anderen Sprachtests setzt wiederum die Konstruktvalidität dieser Tests voraus. Diese ist jedoch in den meisten Fällen höchstens ansatzweise nachgewiesen und wenn, dann in der Regel wiederum auf korrelationsstatistischem Wege. Die potentielle Zirkularität dieses Vorgehens dürfte offensichtlich sein.
- 2) Die korrelationsstatistische Validierung sagt uns nichts über die der Lösung eines Tests zugrundeliegenden individuellen Verarbeitungsmechanismen. So kann z.B. die falsche Lösung eines Items auf einer richtigen Hypothese und die richtige Lösung eines Items auf einer falschen Hypothese beruhen. Außerdem kann die Lösung eines Items das Ergebnis ganz unterschiedlicher mentaler Prozesse sein (vgl. Cohen 1984a und allgemein Ericsson/Simon 1984; 1986). Das üblicherweise praktizierte Verfahren, im Fall eines im statistischen Sinne, d.h. z.B. hinsichtlich Reliabilität und Trennschärfe zufriedenstellenden Items, aufgrund der Lösung des Items quasi automatisch eine Zuordnung zu bestimmten Bereichen einer zugrundeliegenden Fähigkeitsdimension vorzunehmen, ist deshalb äußerst problematisch.
- 3) Die bei der korrelationsstatistischen Validierung üblicherweise verwendete (explorative) Faktorenanalyse ist aus mehreren Gründen nicht unproblematisch. So hängen z.B. die Ergebnisse faktoranalytischer Untersuchungen im starken Maße von Anzahl und Typ der verwendeten Variablen, von der jeweiligen Personenstichprobe sowie von der spezifischen faktoranalytischen Technik ab (vgl. die Hinweise in Grotjahn 1986b).

Es bedarf deshalb zur Validierung von Tests einer Analyse der mentalen Prozesse, die beim Bearbeiten des Tests in den Probanden ablaufen. Eine entsprechende Forderung nach Änderung der üblichen Forschungspraxis hat u.a. Vollmer (1981; 1983) erhoben, der u.E. sehr richtig feststellt:

"What we suggest is the application of alternative research strategies: ... We will never really understand what correlations between tests of different skills mean, what they signify, and why some are higher than others - unless we better understand and are able to model more precisely the cognitive potentials and task specific operations on which performance in the various language tests depends." (1981:115)

In Anbetracht der geschilderten Situation haben wir uns entschlossen, mit Hilfe der Methoden des Lauten Denkens und des retrospektiven Interviews Einblick in die beim Bearbeiten eines C-Tests ablaufenden mentalen Prozesse zu gewinnen. Bevor wir den gewählten Versuchsaufbau genauer darstellen, möchten wir zumindest kurz auf die methodologischen Grundlagen des gewählten Forschungsansatzes eingehen. Eine weitergehende methodologische Diskussion mit entsprechenden Literaturhinweisen findet sich in Grotjahn (1986a; b).

## 4. Lautes Denken und Retrospektion bei der C-Test-Validierung

### 4.1 Methodologische Vorbemerkungen

Kennzeichnend für einen Großteil der empirischen Forschung im Bereich der Sozialwissenschaften ist eine mehr oder minder explizite Orientierung am methodologischen Paradigma der empirisch-analytischen Wissenschaftstheorie. Um den Kriterien der Objektivität und der Reliabilität zu genügen, wird danach getrachtet, die Reaktivität des verwendeten Forschungsinstruments möglichst weitgehend zu eliminieren. Außerdem wird versucht, das Ausmaß an notwendiger Inferenz auf Seiten von Forscher und Proband möglichst gering zu halten. Dies führt in vielen Fällen dazu, daß die Quantität und Qualität der möglichen Informationen drastisch eingeschränkt wird. Außerdem zieht das Streben nach statistischer Repräsentativität und Quantifizierung meist eine Vernachlässigung des Einzelfalles nach sich; der einzelne Proband wird dann nur noch in Hinblick auf seinen Beitrag zur statistischen Gesamttendenz gesehen.

Für Wissenschaftskonzeptionen wie verstehende Soziologie, explorative Sozialforschung, Ethnomethodologie oder reflexive Psychologie ist dagegen in Abgrenzung zur empirisch-analytischen Methodologie die möglichst unvoreingenommene Exploration des zu untersuchenden Phänomens von entscheidender Bedeutung. Im Vordergrund steht die Sicherung der Validität, wobei allerdings dieser Begriff eine partielle Uminterpretation erfährt.

Grundvoraussetzung für die Gewinnung valider Daten ist die Beachtung des Kriteriums der *Offenheit* (vgl. z.B. Hoffmann-Riem 1980:343f.). Gemäß diesem Kriterium wird in vielen Fällen auf eine theoretische Strukturierung des Forschungsgegenstandes so lange verzichtet, bis die Strukturierung durch die Forschungsobjekte selbst erfolgt. Das Kriterium der Offenheit impliziert somit den Verzicht auf eine Hypothesenbildung *ex ante*.

In enger Beziehung zum Kriterium der Offenheit steht das Kriterium der *Kommunikation* (vgl. z.B. Hoffmann-Riem 1980:346f.). Während die empirisch-analytische Methodologie Kommunikations- und Interaktionsbeziehungen zwischen Forscher und Forschungsobjekt als Störfaktor betrachtet, gehen die genannten Wissenschaftskonzeptionen davon aus, daß erst das Eingehen entsprechender Beziehungen die Gewinnung valider Daten ermöglicht. Damit wird der 'Proband' nicht mehr wie in der empirisch-analytischen Methodologie als Erkenntnisobjekt betrachtet, sondern als prinzipiell gleichberechtigtes Erkenntnisobjekt, dessen Kognitionen und subjektive Theorien von zentraler Bedeutung für den Theoriebildungsprozeß sind (vgl. auch die entsprechenden Forderungen des Koordinierungsgremiums im DFG-Schwerpunkt 'Sprachlehrforschung' 1983:32f.).

Indem der Forscher bei Erhebungsverfahren wie z.B. Lautem Denken oder nicht-standardisiertem Interview als Interpret selbst zum Forschungsinstrument wird, laufen entsprechende Verfahren allerdings Gefahr, nicht nur die Kriterien der Objektivität und der Reliabilität, sondern letztendlich auch das der Validität zu verletzen.

Es gibt eine Reihe von Versuchen, den subjektiven Ermessensspielraum von Interpretationen einzuschränken und *Methoden des kontrollierten Fremdverstehens* zu

entwickeln. Unter diesen Versuchen verdient u.E. die sogenannte kommunikative Validierung besondere Aufmerksamkeit (vgl. zum folgenden Treiber/Groeben 1981; 1983; Lechler 1982; Grotjahn 1986a).

Die *kommunikative Validierung* zielt darauf ab, die Validität einer Interpretation dadurch zu sichern, daß überprüft wird, inwieweit der Proband der Interpretation seiner Äußerungen zustimmt. Kern der kommunikativen Validierung ist damit ein dialogkonsenstheoretisches Wahrheitskriterium. Folgt man den u.E. zutreffenden Argumenten von Groeben/Scheele (1977) oder auch von Treiber/Groeben (1981; 1983), dann hat die kommunikative Validierung allerdings nur eine beschränkte Reichweite. Anhand der Zustimmung des Probanden kann nämlich nur über die Rekonstruktionsadäquanz der Interpretation des Forschers hinsichtlich der Kognitionen des Probanden entschieden werden. Kognitionen haben jedoch zumindest potentiell Erklärungskraft auch für das Handeln der Probanden und sind damit zumindest potentiell auch explikative Konstrukte, deren Realitätsadäquanz es zu überprüfen gilt. Dies kann allerdings nicht mehr konsenstheoretisch geschehen, sondern dazu bedarf es einer falsifikationstheoretischen Überprüfung anhand externer Beobachtungsdaten im Sinne der empirisch-analytischen Methodologie.

In Abgrenzung zur kommunikativen Validierung wird die falsifikationstheoretische Überprüfung rekonstruierter Kognitionen auf der Basis der Handlungen der Subjekte auch als *Handlungsvalidierung* bezeichnet (vgl. Wahl 1982; Treiber/Groeben 1981; 1983). Kennzeichnend für die Handlungsvalidierung ist die Berechnung von Korrelationen zwischen Kognitionen und beobachtbarem Verhalten. Die Handlungsvalidierung ist der kommunikativen Validierung zeitlich nachgeordnet, vom Anspruch hinsichtlich der Reichweite jedoch übergeordnet.

#### 4.2 Das Forschungsprojekt

Unter Berücksichtigung von Vorschlägen von z.B. Ericsson/Simon (1980), Huber/Mandl (1982), Treiber/Groeben (1981) und Wahl (1982) haben wir erstmals 1983 versucht, mit Hilfe der Methoden des Lauten Denkens und des retrospektiven Interviews Einblick in die beim Bearbeiten eines C-Tests ablaufenden mentalen Prozesse zu gewinnen. Introspektive Datenerhebungsverfahren wurden bisher vor allem zur Untersuchung von Übersetzungs- und Kommunikationsstrategien in einer Fremdsprache eingesetzt. Dagegen gibt es erst relativ wenige Versuche, introspektive Methoden<sup>2</sup> wie Lautes Denken oder retrospektives Interview bei der Validierung von Sprachtests zu verwenden. Ein erster Überblick findet sich bei Cohen (1984a). In Bezug auf den C-Test gibt es erst vier Untersuchungen, nämlich Cohen/Segal/Weiss (1984; 1985), Feldmann/Stemmer (1986) und Grotjahn/Stemmer (1985). Der von uns gewählte Versuchsaufbau läßt sich folgendermaßen charakterisieren (vgl. auch Grotjahn 1986a):

<sup>2</sup> Wir verwenden den Begriff 'introspektive Methoden' als Sammelbegriff für z.B. Lautes Denken, Lautes Sprechen, Retrospektion, Introspektion (im engeren Sinne). Wir unterscheiden hier auch nicht zwischen 'Lautem Denken' (*thinking aloud*) und 'Lautem Sprechen' (*talking aloud*). Eine terminologische Diskussion verschiedener Typen introspektiver Methoden findet sich z.B. in Cohen (1984b; 1986) und Ericsson/Simon (1984; 1986).

Den Versuchspersonen (Vpn) – es handelte sich vorwiegend um Studentinnen/Studenten – wurden jeweils in einem Einzelexperiment drei verschiedene französische bzw. vier verschiedene spanische, in aufsteigender Schwierigkeit angeordnete C-Test-Texte zur Bearbeitung vorgelegt. Um sicherzustellen, daß die Vp die zu lösende Aufgabe vollständig verstand, wurde sie/er relativ ausführlich über die Zielsetzung unserer Untersuchung informiert. Gleichzeitig sollte damit erreicht werden, daß die Vp sich als 'kompetenter Mitarbeiter' im Sinne von Huber/Mandl (1982) fühlte. Daran anschließend erhielt die Vp eine schriftliche Beschreibung der Aufgabe. Hierin wurde sie/er z.B. angewiesen, alles zu sagen, was sie/er beim Ausfüllen der C-Test-Lücken dachte oder auch fühlte. Weiterhin wurde eine kurze Tonbandaufnahme zur Illustration von Lautem Denken bei der Bearbeitung eines C-Tests vorgespielt. Schließlich wurde der Vp ein kurzer C-Test-Text zur Übung vorgelegt. Nach dieser Instruktionsphase wurde die Vp zur Bearbeitung der C-Test-Texte allein gelassen. Ihr/sein Lautes Denken wurde auf Tonträger aufgenommen.

Danach hörten sich Vp und Versuchsleiterin gemeinsam die Tonbandaufnahme der Phase des Lauten Denkens an. In dieser retrospektiven Phase konnten die Vpn spontan zu ihren Äußerungen Stellung nehmen und Fragen der Versuchsleiterin bezüglich ihrer Äußerungen beantworten. Da dies die erste Untersuchung dieser Art in bezug auf den C-Test war, haben wir, um nicht das Kriterium der Offenheit zu verletzen, die Fragen nicht im voraus festgelegt. Diese Phase, die u.a. der kommunikativen Validierung der Interpretationen der Versuchsleiterin dienen sollte, wurde ebenfalls auf Tonträger aufgezeichnet. Am Ende des Versuchs füllten die Vpn noch einen Fragebogen zu einigen für uns wichtigen Hintergrundinformationen aus. Insgesamt betrug die Zeitdauer für jeden einzelnen Versuch durchschnittlich etwa drei (Französisch) bzw. vier Stunden (Spanisch). Bisher haben wir Daten von 30 Probanden für Französisch und von 40 Probanden für Spanisch erhoben und z.T. schriftlich transkribiert.

#### 5. Datenanalyse

Da das von uns gewählte Datenerhebungsverfahren bisher noch nicht auf C-Tests angewandt worden ist, können wir nicht auf bereits erprobte Analyseverfahren zurückgreifen. Es liegt daher nahe, Forschungsarbeiten auf Gebieten wie z.B. der kognitiven Psychologie, der Sprachlehr- und -lernforschung und der Sprachwissenschaften in bezug auf das Ziel, mentale Problemlösungsprozesse unserer Lernerinnen und Lerner zu erfassen, auf ihre Applizierbarkeit zu überprüfen. Dies bedeutet, daß wir bei der Erstellung eines Analysemodells verschiedene Komponenten und deren Zusammenwirken berücksichtigen müssen. Dazu gehören zum Beispiel:

*Aufgabenanalyse.* Durch sie erhalten wir beispielsweise Hinweise, welche Art von kognitiven Prozessen ablaufen könnten, wenn die Vpn die Aufgabe bearbeiten.

*Remanalyse.* Ausgehend vom (richtig oder falsch) vervollständigten Item lassen sich z.B. die Beziehung dieses Items zu anderen Sprachen (gibt es ein ähnliches Wort in einer anderen Sprache?), zu anderen Textelementen (welche Art von kohäsiver und kohärenter Bindung besteht zwischen Item und Text?) usw. untersuchen.

*Analyse des Problemlösungsverhaltens der Vpn.* Wie gehen die Vpn an die Aufgabe heran? Welche Operationen und Strategien setzen die Vpn ein, um die Aufgabe

zu lösen? Inwiefern spiegelt bestimmtes Problemlösungsverhalten bestimmte Verarbeitungsprozesse wider?

In unserer bisherigen Arbeit haben wir uns vor allem mit der Analyse des Problemlösungsverhaltens und der Möglichkeit der Identifizierung bestimmter Verarbeitungsprozesse beschäftigt. Eine detaillierte Aufgaben- und Itemanalyse steht noch aus.

### 5.1 Verarbeitungsprozesse

Die Kognitionspsychologie betrachtet Kognition als Informationsverarbeitung. In dem Informationsverarbeitungsmodell der sogenannten Schema-Theorie wird Information als in Form von Schemata gespeichert betrachtet. Rumelhart (1980:34) beschreibt den schema-theoretischen Ansatz wie folgt:

"A schema theory is basically a theory about knowledge. It is a theory about how knowledge is represented and about how that representation facilitates the use of the knowledge in particular ways. According to schema theories, all knowledge is packaged into units. These units are schemata. Embedded in these packets of knowledge is, in addition to the knowledge itself, information about how this knowledge is to be used."

Auf die Bearbeitung des C-Tests bezogen heißt das: jeder Input, den die Lerner bei der Bearbeitung des C-Tests erhalten, wird auf ein solches Schema abgebildet. Es werden folgende zwei grundlegende Arten der Informationsverarbeitung unterschieden: *bottom-up processing* und *top-down processing* (vgl. Bobrow/Norman 1975; Adams/Collins 1979). Nach Rumelhart (1980) können diese auf jeder Analyseebene auftreten.

Beim *bottom-up processing* wird der Input auf Schemata mit aufsteigender Komplexität abgebildet. So identifiziert die Lernerin/der Lerner z.B. beim Lesen des ersten Wortes eines C-Test-Textes auf der untersten Ebene visuell zunächst kleine Einheiten wie Grapheme oder Graphemfolgen, die anschließend in Phoneme oder Phonemfolgen kodiert werden. Diese wiederum werden zu größeren Einheiten wie Silben und Wörter zusammengesetzt. Auf der nächst höheren Ebene wird eine semantisch-syntaktische Analyse durchgeführt, die möglicherweise zur Identifizierung von Satzkonstituenten führt und schließlich eine erste semantische Interpretation eines größeren Textsegmentes ermöglicht. Auf der dritten Ebene wird schließlich eine Text- und Pragmatikanalyse durchgeführt, die letztendlich zum Textverständnis führen kann.

*Top-down processing* findet in umgekehrter Richtung statt. Als Ausgangspunkt dienen Schemata auf höherer Ebene wie z.B. die allgemeine Bedeutung eines Textes. Von dieser Ebene aus geht die Lernerin/der Lerner durch die tiefer liegenden Ebenen – wie z.B. Satzebene oder Wortebene – und sucht Informationen, die sie/er in die Schemata einordnen kann. Adams und Collins (1979:5) bemerken hierzu:

"Bottom-up processing ensures that the reader will be sensitive to information that is novel or that does not fit her or his ongoing hypotheses about the content of the text; top-down processing helps the reader to resolve ambiguities or to select between alternative possible interpretation of the incoming data."

*Top-down* und *bottom-up processing* werden beim kompetenten Native Speaker simultan aktiviert. Hinsichtlich der Aufgabenstellung beim C-Test erschien es uns daher interessant zu untersuchen, inwieweit das Problemlösungsverhalten der Lerner Hinweise auf diese Formen der Verarbeitung gibt.

Bei der C-Test-Bearbeitung muß Information aus dem Langzeitgedächtnis abgerufen werden. Das erfolgreiche Abrufen ist von zwei Bedingungen abhängig. Zunächst muß das Verarbeitungssystem sich in einem bestimmten Abrufungsmodus befinden. Auf den C-Test bezogen heißt das, daß die Lernerin/der Lerner 'in Bereitschaft' sein muß. Des weiteren muß ein angemessener Abrufstimulus präsent sein. Der offensichtlichste Abrufstimulus ist im C-Test der Anfang des unvollständigen Wortes. Daneben gibt es aber sicherlich noch weitere Stimuli. So dienen z.B. auch der Kontext und Kontext dieser unvollständigen Wörter als Stimuli. Gelingt der Lernerin/dem Lerner keine erfolgreiche Verarbeitung, dann kann das mehrere Gründe haben. Der Abbildungsprozeß kann durch die Art des Inputs gestört werden wie z.B. die unvollständigen Wörter im C-Test. Norman und Bobrow (1975) bezeichnen diesen Sachverhalt als *signal data-limits*.

Weiterhin kann eine erfolglose Bearbeitung auch in der mangelnden Beherrschung der jeweiligen Fremdsprache durch die Lernerin/den Lerner begründet sein. So können z.B. vergessene Elemente der Fremdsprache zu sog. *memory data-limits* führen.

Konzentriert sich die Lernerin/der Lerner darauf, ein bestimmtes Wort zu vervollständigen, kann die Bedeutung dieses Wortes innerhalb des Textzusammenhangs völlig in den Hintergrund treten. Dies ist z.B. der Fall, wenn die Lerner versuchen, ein Wort zu vervollständigen, indem sie einfach Buchstaben oder Silben an den vorgegebenen Wortanfang hängen, ohne in irgendeiner Weise auf den Kontext zu achten. Manche Lerner fühlten sich bei der Aufgabenbearbeitung auch einfach dadurch gestört, daß sie gleichzeitig 'laut denken' mußten. Dieser Sachverhalt ist ein Beispiel für *resource-limited processing*.

### 5.2 Bisherige Analysen

Von den bisher transkribierten Daten haben wir 10 verbale Protokolle der französischen und 10 verbale Protokolle der spanischen C-Tests einer exemplarischen Analyse unterzogen (vgl. Feldmann/Stemmer 1986). Dabei sind wir u.a. von Untersuchungen ausgegangen, die sich mit Strategien von Lernern in fremdsprachlichen Interaktionen beschäftigen (vgl. Færch/Kasper 1983). Mit Færch/Kasper (1980:60) verstehen wir unter Strategien "a potentially conscious plan for solving what to the individual presents itself as a problem in reaching a particular goal". Ausgehend von diesem Ansatz haben wir anhand unserer Daten ein *vorläufiges* Modell vom Problemlösungsverhalten der Lerner sowie eine Liste der von den Lernern verwendeten Strategien bei der C-Test Bearbeitung erstellt. Abschließend haben wir dann diese Strategien auf ein *top-down* und *bottom-up processing*-Kontinuum abgebildet. Aufgrund dieses ersten exemplarischen Analyseversuchs erscheint uns dieser Ansatz geeignet, das Problemlösungsverhalten und die (fremd-)sprachlichen Verarbeitungsprozesse der Lerner bei der C-Test-Bearbeitung zu beschreiben und zu erklären. Auch hoffen wir, durch

die noch ausstehende detaillierte Aufgaben- und Itemanalyse weitere Unterstützung für unser Erklärungsmodell zu erhalten.

Dennoch hat dieser Ansatz auch seine Grenzen (vgl. hierzu Grotjahn 1986b). Zum einen wird die wechselseitige Abhängigkeit von kognitiven Prozessen nicht berücksichtigt. Des weiteren haben wir auf der Produktebene bisher nicht systematisch die Anzahl und Art der Items, die der Lerner bereits vervollständigt hat, und die möglicherweise hierdurch zunehmende Textkohäsion und -kohärenz berücksichtigt. Hinzu kommt drittens, daß unterschiedliche Testergebnisse sicherlich nur teilweise aufgrund von Unterschieden im Problemlösungsverhalten erklärt werden können. Viertens sind möglicherweise bestimmte mentale Prozesse mit unserem Datenerhebungsverfahren nicht oder nur sehr bedingt erfassbar. Fünftens wissen wir bisher nicht, inwieweit die Aufgabenstellung selbst – nämlich das Laute Denken – die Verarbeitungsprozesse beim C-Test-Lösen beeinflussen.

## 6. Ausblick

Obwohl die introspektive Analyse der mentalen Prozesse in unserem Projekt im Vordergrund steht, versuchen wir das Problem, was der C-Test nun eigentlich mißt, letztendlich von mehreren unterschiedlichen Seiten anzugehen. Methodologisch gesehen kann unser Ansatz als sog. Triangulierungsstrategie charakterisiert werden (zum Konzept der Triangulierung vgl. Denzin 1978; Patton 1980). Wir hoffen, auf diese Weise ein weniger 'methodologisch verzerrtes' und letztendlich auch vollständigeres Bild dessen zu erhalten, was der C-Test mißt. Es werden von uns folgende drei grundlegende Triangulierungstypen verwendet (vgl. Grotjahn 1986a):

- 1) *Daten-Triangulierung*, d.h., wir benutzen verschiedene Datenquellen, um ein und dasselbe Problem zu untersuchen;
- 2) *Theorie-Triangulierung*, d.h., wir interpretieren ein und denselben Datensatz aus unterschiedlichen theoretischen Perspektiven; und
- 3) *methodische Triangulierung*, d.h., wir verwenden unterschiedliche Methoden bei der Analyse ein und desselben Problems.

Zusätzlich zur interpretativen Analyse der Verbalprotokolle des Lauten Denkens und des retrospektiven Interviews sind folgende Analysen entweder bereits begonnen oder für die nahe Zukunft geplant (vgl. Grotjahn 1986b):

- 1) *Eine theoretische (logische) Analyse der Aufgabe, die die Probanden zu lösen haben.* Im Hinblick auf diese Aufgabenanalyse<sup>3</sup> prüfen wir im Augenblick, inwieweit allgemeine Textverarbeitungsmodelle, wie z.B. das Modell von Van Dijk/Kintsch (1983), Modelle des erst- und/oder fremdsprachlichen Leseprozesses (z.B. Adams/Collins 1979; Just/Carpenter 1980; Karcher 1985) sowie informationsstatistische Ansätze (vgl. die Hinweise bei Klein-Braley 1985) Verwendung finden können.
- 2) *Handlungsvalidierung der Protokollaten.* Entsprechend dem Paradigma der bereits beschriebenen Handlungsvalidierung wird die Validität der Verbalprotokolle und unserer interpretativen Analysen überprüft durch

<sup>3</sup> Zur Bedeutung von Aufgabenanalysen für kognitionspsychologische Untersuchungen vgl. z.B. Bromme (1985) oder Ericsson/Simon (1986).

- a) Berechnung von Korrelationen zwischen identifizierten Strategien und beobachtbarem Verhalten (z.B. Leistung im C-Test gemessen anhand des Gesamtpunktwertes);
- b) Vorhersage zukünftigen Verhaltens (z.B. Lösung eines spezifischen Items) auf der Basis der identifizierten Strategien; und
- c) Unterrichtung von weniger erfolgreichen C-Test-Bearbeitern in Strategien, die erfolgreichere C-Test-Bearbeiter verwenden und Überprüfung, ob sich der C-Test-Punktwert der weniger erfolgreichen C-Test-Bearbeiter aufgrund der Verwendung der unterrichteten Strategien erhöht.

- 3) *Eine Frequenzanalyse der identifizierten Strategien.* In diesem Zusammenhang werden Unterschiede in den Verwendungshäufigkeiten der einzelnen Strategien durch Probanden- und Textvariablen erklärt.
- 4) *Eine deskriptive Analyse der Lösungen der C-Test-Items.* Hier analysieren wir die Items z.B. im Hinblick auf die Frequenz von akzeptablen, inakzeptablen und fehlenden Lösungen sowie im Hinblick auf die linguistischen Eigenschaften der versuchten Lösungen.
- 5) *Eine explanatorische Analyse der Lösungen der C-Test-Items auf der Basis unserer introspektiven und retrospektiven Daten.* In diesem Zusammenhang werden die Lösungen als Indikatoren der Interimsprache der Lerner angesehen und mit Hilfe verschiedener psycholinguistischer Prozesse und Strategien erklärt. Damit folgen wir nicht mehr länger dem begrenzten Paradigma der 'klassischen' Interimsprachenanalyse, sondern verwenden einen u.E. der Komplexität des Phänomens 'Interimsprache' adäquateren Ansatz.

Obwohl eine Reihe von Analysen noch aussteht und trotz mancher Einschränkungen bezüglich des gewählten Ansatzes (vgl. 5.2), glauben wir aufgrund unserer bisherigen Erfahrungen einen vielversprechenden Weg gefunden zu haben, der es uns ermöglicht, ein besseres Verständnis dafür zu bekommen, was der C-Test und möglicherweise auch andere Sprachtests tatsächlich messen.

## Bibliographie

- Adams, M.J./Collins, A.: "A schema-theoretic view of reading." Freedle, R.O. (ed.): *New directions in discourse processing*. Norwood, N.J.: Ablex 1979: 1-22.
- Alderson, J.C.: *A study of cloze procedure with native and non-native speakers of English*. Unpublished Ph. D. dissertation. Edinburgh: University of Edinburgh 1978.
- Alderson, J.C.: "The cloze procedure and proficiency in English as a foreign language." *TESOL Quarterly* 13 (1979): 219-227.
- Bobrow, D./Norman, D.A.: "Some principles of memory schemata." Bobrow, D./Collins, A. (eds.): *Representation and understanding: Studies in cognitive science*. New York: Academic Press 1975: 131-149.
- Bromme, R.: "Aufgaben- und Problemanalyse bei der Untersuchung des problemlösenden Denkens." Jüttemann, G. (ed.): *Qualitative Forschung in der Psychologie: Grundfragen, Verfahrensweisen, Anwendungsfelder*. Weinheim/Basel: Beltz 1985: 259-281.

- Cohen, A.D.: "On taking language tests: what the students report." *Language Testing* 1 (1984): 70-81. (a).
- Cohen, A.D.: "Studying second-language learning strategies: how do we get the information." *Applied Linguistics* 2 (1984): 101-112. (b).
- Cohen, A.D.: "Using verbal reports in research on language learning." Færch, C./Kasper, G. (eds.). 1986 (im Druck).
- Cohen, A.D./Segal, M./Weiss, R.: "The C-Test in Hebrew." *Language Testing* 1 (1984): 221-225.
- Cohen, A.D./Segal, M./Weiss, R.: "The C-Test in Hebrew." Klein-Braley, C./Raatz, U. (eds.). 1985: 121-127.
- Denzin, N.K.: *The research act*. New York: McGraw-Hill 1978.
- Ericsson K.A./Simon, H.A.: "Verbal reports as data." *Psychological Review* 87 (1980): 215-251.
- Ericsson, K.A./Simon, H.A.: *Protocol analysis. Verbal reports as data*. Cambridge, Mass.: MIT Press 1984.
- Ericsson, K.A./Simon, H.A.: "Verbal reports on thinking." Færch, C./Kasper, G. (eds.). 1986 (im Druck).
- Færch, C./Kasper, G.: "Processes and strategies in foreign language learning and communication." *Interlanguage Studies Bulletin* 5 (1980): 47-118.
- Færch, C./Kasper, G. (eds.): *Strategies in interlanguage communication*. London/New York: Longman 1983.
- Færch, C./Kasper, G. (eds.): *Introspection in second language research*. Clevedon: Multilingual Matters 1986 (im Druck).
- Feldmann, U./Stemmer, B.: "Thin\_\_\_\_\_ aloud a\_\_\_\_\_ retrospective da\_\_\_\_\_ in C-te\_\_\_\_\_ taking: diff\_\_\_\_\_ languages - diff\_\_\_\_\_ learners - sa\_\_\_\_\_ approaches?" Færch, C./Kasper, G. (eds.). 1986 (im Druck).
- Gaies, S.J.: "Validation of the noise test." Grotjahn, R./Klein-Braley, C./Stevenson, D. (eds.): *Taking their measure: The validity and validation of language tests*. Bochum: Brockmeyer 1986 (im Druck).
- Groeben, N./Scheele, B.: *Argumente für eine Psychologie des reflexiven Subjekts*. Darmstadt: Steinkopff 1977.
- Grotjahn, R.: "On the methodological basis of introspective methods." Færch, C./Kasper, G. (eds.). 1986. (im Druck). (a).
- Grotjahn, R.: "Test validation and cognitive psychology: Some methodological considerations." *Language Testing*. 1986 (im Druck). (b).
- Grotjahn, R.: "Ist der C-Test ein Lesetest?" Addison, A./Vogel, K. (eds.): *Fremdsprachenausbildung an der Hochschule*. Bd. 2. Bochum: AKS 1986 (im Druck). (c).
- Grotjahn, R./Stemmer, B.: "On the development and evaluation of a C-test for French." Klein-Braley, C./Raatz, U. (eds.). 1985: 101-120.
- Hoffmann-Riem, C.: "Die Sozialforschung einer interpretativen Soziologie - Der Datengewinn." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 32 (1980): 339-372.
- Huber, G.L./Mandl, H.: "Gedankenstichproben." Huber, G.L./Mandl, H. (eds.): *Verbale Daten*. Weinheim/Basel: Beltz 1982: 104-118.
- Just, M.A./Carpenter, P.A.: "A theory of reading: From eye fixations to comprehension." *Psychological Review* 87 (1980): 329-354.
- Karcher, G.L.: "Aspekte einer Fremdsprachenlegetik. Zur Differenzierung von erst- und fremdsprachlichem Lesen." *Jahrbuch Deutsch als Fremdsprache* 11 (1985): 14-35.
- Klein-Braley, C.: *Empirical investigations of cloze tests: an examination of the validity of cloze tests as tests of general language proficiency in English for German university students*. Unpublished Ph. D. dissertation. Duisburg: Universität Duisburg 1981.
- Klein-Braley, C.: "Reduced redundancy as an approach to language testing." Klein-Braley, C./Raatz, U. (eds.). 1985: 1-13.
- Klein-Braley, C./Raatz, U.: "A survey of research on the C-Test." *Language Testing* 1 (1984): 134-146.
- Klein-Braley, C./Raatz, U. (eds.): *C-Tests in der Praxis. (Fremdsprachen und Hochschule 13/14)*. Bochum: Arbeitskreis der Sprachenzentren, Sprachlehrinstitute und Fremdspracheninstitute (AKS) 1985.
- Koordinierungsgremium im DFG-Schwerpunkt 'Sprachlehrforschung' (ed.): *Sprachlehr- und Sprachlernforschung: Begründung einer Disziplin*. Tübingen: Narr 1983.
- Lechler, P.: "Kommunikative Validierung." Huber, G.L./Mandl, H. (eds.): *Verbale Daten*. Weinheim/Basel: Beltz 1982: 243-258.
- Norman, D./Bobrow, D.: "On data-limited and resource-limited processes." *Cognitive Psychology* 7 (1975): 44-64.
- Oller, J.W.J.: "Cloze tests of second language proficiency and what they measure." *Language Learning* 23 (1973): 105-118.
- Oller, J.W.J.: "Evidence for a general language proficiency factor: an expectancy grammar." *Die Neueren Sprachen* 76 (1976): 165-174.
- Oller, J.W.J.: *Language tests at school*. London: Longman 1979.
- Patton, M.Q.: *Qualitative evaluation methods*. Beverly Hills: Sage 1980.
- Raatz, U.: "The factorial validity of C-Tests." Culhane, T./Klein-Braley, C./Stevenson, D.K. (eds.): *Practice and problems in language testing 7*. Colchester: University of Essex 1984: 124-139.
- Raatz, U./Klein-Braley, C.: "The C-test - a modification of the cloze procedure." Culhane, T./Klein-Braley, C./Stevenson, D.K. (eds.): *Practice and problems in language testing IV*. Colchester: University of Essex 1982: 113-138.
- Raatz, U./Klein-Braley, C.: "Ein neuer Ansatz zur Messung der Sprachleistung. Der C-Test: Theorie und Praxis." Horn, R./Ingenkamp, K./Jäger, R.S. (eds.): *Tests und Trends 3. Jahrbuch der Pädagogischen Diagnostik*. Weinheim/Basel: Beltz 1983: 107-138.
- Raatz, U./Klein-Braley, C.: "How to develop a C-Test." Klein-Braley, C./Raatz, U. (eds.). 1985: 20-22.
- Rumelhart, D.E.: "Schemata: The building blocks of cognition." Spiro, R./Bruce,

Helmut Brammerts/Karin Kleppin/Federico Zarco de la Hoz  
(Ruhr-Universität Bochum)

## Konzept eines Tandemkurses

### 1. Probleme eines Feriensprachkurses im Zielland

Im Rahmen der Universitäts- und Städtepartnerschaft Bochum/Oviedo findet seit 1981 jeweils im September ein Ferienkurs Spanisch mit ca. 35 Bochumer Studenten, begleitet und unterrichtet von zwei Bochumer Spanischlehrern, in Oviedo statt.

Der Entschluß, ab 1984 zusätzlich einen Deutschkurs für Angehörige der Universität Oviedo anzubieten, wurde nicht nur deshalb gefaßt, weil die Universität in Oviedo entsprechende Wünsche geäußert hatte; wir erhofften uns vielmehr davon auch eine Verbesserung der Lernbedingungen für die deutschen Spanischkursteilnehmer, insbesondere eine Erleichterung der Kontaktaufnahme mit Spaniern.

Die Bochumer Studenten waren in einem Studentenwohnheim untergebracht, wo sich zur Nachprüfungszeit auch spanische Studenten aufhielten. Im Gegensatz zu vielen Sprachkursen in typischen Ferienorten mit ausländischer Touristenmehrheit boten sich also Möglichkeiten zu Kontakten mit gleichaltrigen Kommilitonen. Dennoch wurde diese Möglichkeit nur von einem Teil der Deutschen ausgenutzt: die Kontakte blieben auf verhältnismäßig wenige Kommunikationssituationen (Umgang im Wohnheim, Essen, Bar, Fiestas) beschränkt.

Dies war zum einen auf die begrenzten Spanischkenntnisse zurückzuführen, jedoch auch auf die starken Bindungen innerhalb der Lerngruppe, die ja gerade bei einem Auslandsaufenthalt außerhalb der gewohnten Umgebung dem einzelnen Rückhalt bietet. Verstärkt wurden diese Bindungen noch durch gemeinsame Unternehmungen wie Exkursionen usw.: die Kommunikation fand im wesentlichen auf Deutsch statt.

Ähnliches läßt sich bei fast allen Feriensprachkursen im Zielland beobachten. Auch bei sprachlich heterogenen Lernergruppen ist häufig nicht die Zielsprache wichtigstes Kommunikationsmittel, sondern eine Fremdsprache, die alle schon besser beherrschen, wie z.B. Englisch.

Kontaktaufnahmen zu *native speakers* erfordern zudem auch ein großes Maß an Eigeninitiative, Mut und Selbstbewußtsein bei dem Lerner. Er merkt sehr schnell, daß Gespräche mit ihm auch für die *native speakers* verhältnismäßig anstrengend sind, und daß diese eine Menge Geduld aufbringen und in der Regel die Hauptlast des Gesprächs zu tragen haben. Seine Hemmungen werden dadurch verstärkt. Der Zwang zur Reduktion der inhaltlichen Komplexität dessen, was man äußern will, wird beiden bewußt und erschwert zusätzlich die Kommunikation. Weibliche Kursteilnehmer können durch Probleme mit ihrer Rolle als Frau – besonders im südeuropäischen Ausland – noch weiter verunsichert werden.

Um den oben genannten Problemen entgegenzuwirken, haben wir von Anfang an versucht, den zusätzlich angebotenen Deutschkurs mit dem Spanischkurs zu verbinden. Vor allem wurden *Tandems* aus je einem Deutschen und einem Spanier gebildet,

- B./Brewer, W. (eds.): *Theoretical issues in reading comprehension: Perspectives and cognitive psychology, linguistics, artificial intelligence, and education.* Hillsdale, N.J.: Erlbaum 1980: 33-47.
- Spolsky, B.: "What does it mean to know a language or how do you get someone to perform his competence?" Oller, J.W./Richards, J.C. (eds.): *Focus on the learner.* Rowley, Mass.: Newbury House 1973: 164-176.
- Treiber, B./Groeben, N.: "Handlungsforschung und epistemologisches Subjektmodell." *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie* 1 (1981): 117-138.
- Treiber, B./Groeben, N.: "Vorarbeiten zu einer Reflexiven Sozialtechnologie – Die Integration von dialog-konsens-theoretischem Wahrheits- sowie Falsifikationskriterium am Beispiel subjektiver Theorien von Lehrern." Zedler, P./Moser, H. (eds.): *Aspekte qualitativer Sozialforschung.* Opladen: Leske & Budrich 1983: 163-208.
- Van Dijk, T.A./Kintsch, W.: *Strategies of discourse comprehension.* New York: Academic Press 1983.
- Vollmer, H.: "Why are we interested in 'General Language Proficiency'?" Klein-Braley, C./Stevenson, D.K. (eds.): *Practice and problems in language testing* 1. Bern/Frankfurt/M.: Lang 1981: 96-123.
- Vollmer, H.: "The structure of foreign language competence." Hughes, A./Porter, D. (eds.): *Current developments in language testing.* London: Academic Press 1983: 3-29.
- Vollmer, H./Sang, F.: "Zum psycholinguistischen Konstrukt einer internalisierten Erwartungsgrammatik." *Linguistik und Didaktik* 42 (1980): 122-148.
- Wahl, D.: "Handlungsvalidierung." Huber, G.L./Mandl, H. (eds.): *Verbale Daten.* Weinheim/Basel: Beltz 1982: 259-274.