

Preibusch, Wolfgang; Hagemeister, Volker; Schuricht, Klaus; Seyhan, Hayrettin
**Cloze-Tests als Instrumente zur Kontrolle der Deutsch-Kenntnisse
türkischer Schüler in der Sekundarstufe I**

Unterrichtswissenschaft 11 (1983) 2, S. 180-193



Quellenangabe/ Reference:

Preibusch, Wolfgang; Hagemeister, Volker; Schuricht, Klaus; Seyhan, Hayrettin: Cloze-Tests als Instrumente zur Kontrolle der Deutsch-Kenntnisse türkischer Schüler in der Sekundarstufe I - In: Unterrichtswissenschaft 11 (1983) 2, S. 180-193 - URN: urn:nbn:de:0111-pedocs-295132 - DOI: 10.25656/01:29513

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-295132>

<https://doi.org/10.25656/01:29513>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Allgemeiner Teil

Wolfgang Preibusch, Volker Hagemester, Klaus Schuricht, Hayrettin Seyhan

Cloze-Tests als Instrumente zur Kontrolle der Deutsch-Kenntnisse türkischer Schüler in der Sekundarstufe I

Zur Kontrolle der Wirkungen von Unterrichtsmaterial (Einführung in die Elektrizitätslehre) für türkische Schüler auf der Sekundarstufe I wurden Cloze-Tests entwickelt. Diese wurden mit Vierfachauswahl-Vorgaben administriert. Beobachtet werden sollten etwaige Verbesserungen in der Beherrschung des Deutschen (a) im unterrichtsthematischen Sprachfeld, (b) außerhalb dieses Feldes, aber an einem ebenfalls schultypischen Textrepräsentanten sowie (c) an einem Repräsentanten literarisch überformter Umgangssprache. Zur Unterscheidung zwischen Versprachlichung des Unterrichtsthemas und Kenntniszuwachs unterhalb solcher Versprachlichung wurde zusätzlich ein deutsch-türkischer, stark illustrierter Wissenstest administriert. Es wurden verschiedene Integrationsindizes zur Bewertung der Validität der Instrumente erhoben. Die Zusammenhänge zwischen diesen Indizes und den hochreliablen Cloze-Tests erweisen sich als komplizierter, denn zuvor vermutet.

Cloze tests for observing the knowledge of German of Turkish pupils

In order to investigate the effects of teaching material – introduction to the science of electricity – for Turkish pupils at secondary level (Sekundarstufe I) cloze tests were developed. The correct answer had to be selected from a choice of four possible answers. The aim of the tests was to observe improvements – if any – in the knowledge of German (a) in linguistic areas relating to subjects dealt with in lessons, (b) in other areas, but using a text of a type also frequently employed at school, (c) in other areas, using a basically colloquial text which had been given a certain amount of literary polish. In order to distinguish to what extent the pupils had learnt to verbalize the subject dealt with in class and to what extent they had increased their knowledge of the subject independently of such verbal ability, they underwent an additional test in German and Turkish with plentiful illustrations which was designed solely to test their knowledge of the subject. Various integration indices were worked out to assess the validity of the instruments. The connexions between these indices and the highly reliable cloze tests reveal themselves to be more complicated than previously suspected.

1. Problemlage

Betrachtet man das Instrumentenangebot zur Kontrolle von Deutsch als Fremdsprache in der Hoffnung, ein taugliches Verfahren für Forschungs- und Entwicklungsaufgaben zu finden, dann stößt man auf überwiegend recht unhandliche, zum Teil nur in Individualadministration einzusetzende Instrumente, die überdies von ihren Autoren aus Sorge um ihren professionellen Ruf vorsichtig auf die Ebene von „Erhebungsinstrumenten“ (Portz und Pfaff 1981) oder „Handreichungen“ zur „Sprachstandsmessung“ (Fliegner und Gogolin 1982) zurückgestuft werden, ungeachtet der Frage, als was sie – als Test oder

Handreichung – in der real existierenden Schule fungieren. Die Vorsicht ist, nach wissenschaftlichen Standards, korrekt und angezeigt, erweist sich aber gleichwohl fallweise als ein bloßes „Seine Hände in Unschuld waschen“, denkt man etwa an sich anbahnende Versuche, wenigstens Teilen der ausländischen Bevölkerung qua Sprachstandseinstufung die Integrationsprobleme für ihre Kinder als Motiv für die Erwägung der Rückkehr in ihre Heimatländer „objektiv“ vor Augen zur führen. Selbst im Rahmen angewandter Sprachwissenschaft scheinen die Kriterien alles andere als eindeutig oder doch wenigstens konsensual zu sein: Die Pädagogische Arbeitsstelle des Deutschen Volkshochschul-Verbandes nennt ihr Instrument definitiv einen Einstufungs- und Diagnostiktest, obwohl sich die Art seiner Erprobung nicht essentiell von der der bereits genannten Verfahren unterscheidet und Informationen über Validierungsverfahren und Gütekriterien nach Testentwicklungsstandards als lediglich möglicherweise vorhanden vermutet werden können (*Einstufungs- und Diagnostiktest* 1981).

Das Projekt, in dessen Zusammenhang die hier zu beschreibenden Verfahren entwickelt und angewandt worden sind, hatte weder implizit noch explizit selektive Funktionen, eher das Gegenteil; es ging um die Entwicklung von Unterrichtsmaterialien für den Physikunterricht, die für türkische Schüler nicht nur bei der türkischen Hoch-Sprache anknüpfen sollten, sondern auch bei der Begriffsbildung einer Bevölkerung, die überwiegend aus ländlichen, handarbeitsorientierten Lebenszusammenhängen stammt. Diese Schüler sollten mittels des Materials sowohl Physik (Einführung in die Elektrizitätslehre) als auch das angemessene, begrifflich entwickelte Deutsch im Rahmen des thematischen Feldes erlernen (*Hagemeister, Preibusch, Schuricht und Seyhan* 1981, 1982). Ziele der Kontrolle der Deutsch-Kenntnisse in diesem Projekt mußten deshalb sein:

1. Erfassung von unterrichtlich beeinflussten Sprachfeldern (Thema des Physikunterrichts);
2. Erfassung von nicht unterrichtlich beeinflussten Sprachfeldern;
3. Erfassung von Unterrichtseffekten „unabhängig“ von Sprache;
4. Erfassung von Reifungsprozessen (vgl. *Campbell und Stanley* 1963, spez. S. 175);
5. Erfassung von Unterrichtseffekten getrennt nach Muttersprache und Zweitsprache (Türkisch und Deutsch).

Bereits 1968 hat *Spolsky* in einem Referat schon im Titel die Frage gestellt: „What does it mean to know a language, or how to get someone to perform his competence?“ Dabei trifft er eine begriffliche Unterscheidung, die – wie selten in der Fremdsprachendidaktik – eine eminente lernpsychologische Realität zu besitzen scheint; er unterscheidet zwischen „funktionalem Gebrauch“ der Sprache als Ausweis ihrer Beherrschung und dem „Kennen struktureller Elemente“ als Ausweis der Sprachbeherrschung, letztere so, wie sie sich in Gestalt von Listen struktureller Erscheinungen für die Bereiche Phonologie, Syntax, Morphologie usw. für jede Sprache zusammenstellen lassen. Die Beherrschung solchen Elementen-Repertoires garantiert keinen „funktionalen Gebrauch“ des Repertoires als Sprache, etwa eine fremdsprachige Zeitung lesen können, einen qualifizierten Arbeitsplatz im fremdsprachigen Land ausfüllen können oder einem Ausländer ein Spiel erklären und es mit ihm spielen können (vgl. *Preibusch* 1979). *Spolsky* (1968) sieht vor allem zwei Gründe für die Defizite des Sprachbeherrschungsbegriffs auf der Grundlage der (meist ja lehrgangsbezogenen) Strukturelemente-Listen:

1. Sprache sei nicht mechanistisch, sondern kreativ, wobei er an Chomsky, Humboldt und Descartes anknüpft;
2. Sprache sei redundant, wobei eminent wichtig sei, daß das Redundanzmerkmal statistischer und nicht linguistischer Natur sei und in einem langen Reifungs- und Teilhabeprozess an der Sprechergemeinde erworben werde;

und wir würden hinzufügen:

3. Sprache hat den Charakter komplexer Verlaufsgestalten, die aus mehr als nur den puren Wörtern bestehen, wozu Situation, Sprechererscheinung, Stimme, die nativen Fehler (d. h. die in der Sprechergemeinde erwartbaren Normverstöße) u. v. a. m. gehören (vgl. Preibusch 1979).

Spolsky (1968) gelangt, von diesen Überlegungen ausgehend, zu dem Schluß, daß für eine Kontrolle des funktionalen Gebrauchs einer (Fremd-)Sprache die redundanzeinschränkende Verfahren die angemessenen seien. In diesem Sinne seien Cloze-Tests¹ und Verfahren wie der von ihm und Mitarbeitern entwickelte Noise-Test (auch Indiana Test genannt) Instrumente, die den Probanden am ehesten gültige Reaktionen für funktionalen Sprachgebrauch abverlangen. Über diese Überlegungen hinaus sprachen für uns einige Gründe aus der Forschung zum Zweitspracherwerb wie auch technisch-praktische Gründe dafür, Cloze-Tests für den hier anstehenden Zweck zu verwenden:

1. Die angezielten Schüler der Sekundarstufe I können schreiben und lesen; eine Bewährung und Übung der Schriftsprache ist auf dieser Schulstufe, gegenüber der Grundschule etwa, deutlich vorrangig vor mündlichem Gebrauch.
2. Der fachliche Rahmen des Entwicklungsprojekts verweist auf die Beherrschung des erforderlichen Schrift-Deutschs aus fachimmanenten Gründen stärker als auf das Deutsch des alltäglichen, mündlichen Umgangs.
3. Die Schule sollte – folgt man etwa den Forschungsergebnissen von Swain (1981) – für L1-Minoritätskinder die Majoritätssprache nicht nur erst spät einführen, sondern das Schwergewicht vornehmlich auf die „cognitive academic linguistic proficiency“ legen.
4. Cloze-Tests gestatten durch Textsorten-Wahl und vor allem durch die sehr fein variierbare LösCHFrequenz eine außergewöhnlich sensible Anpassung an das Probandenniveau.

2. Konstruktive Aspekte

Erste Explorationen mit Cloze-Test-Entwürfen² zeigten, daß türkische Schüler in der Berliner Schule Schwierigkeiten mit dem Lückenausfüllen haben würden, die nicht notwendigerweise gültige Hinweise auf das Textverständnis und die Sprachbeherrschung liefern. Für die Untersucher entstünden daraus überdies kaum anders als willkürlich zu lösende Scoring-Probleme, etwa hinsichtlich der Akzeptanz einer Lösung bei zunehmend abweichender Rechtschreibung. Zudem wurde deut-

lich, daß Lösungsmodi, die Schreiben erforderten, zu einer systematischen Überschätzung der Antwortkategorie „Weiß nicht“ (= Auslassen von Items) führen würden. Zudem schien es uns erforderlich, Unterschiede zwischen den Schülern in Assoziationsflüssigkeit, Kreativität u. ä. als im gegebenen Kontext irrelevante Variablen zu neutralisieren. Wir entschieden uns deshalb für Mehrfachauswahlvorgaben anstelle bloßer Lücken; dabei ergab sich zudem die Möglichkeit, linguistisch hypothesenhaltige Distraktoren planmäßig einzuführen. Nach *Jonz* (1976) dürfte zudem eine hochgradige Äquivalenz dieses Testformats gegenüber der üblichen Lückenvorlage erwartet werden; er fand zwischen beiden Formaten eine Ergebniskorrelation von $r = 0.80$. (Siehe auch *Porter* 1976 hinsichtlich der Verwendung von Mehrfachauswahlvorgaben bei Cloze-Tests.)

Weitere Vorversuche mit diesem Format zeigten uns schließlich, daß trotz der Vorgaben hohe Diskriminationskoeffizienten für das einzelne Item und sehr hohe Werte für interne Konsistenz der Tests erwartbar sein würden.

In den ersten Vorversuchen wurden türkischen Schülern der Sekundarstufe I die Entwurfsfassungen der Cloze-Tests vorgelegt, in denen anstelle des gelöschten Wortes lediglich stets gleichgroße Lücken waren, die frei ausgefüllt werden mußten. Die auf diese Weise erzielten Antwortenlisten zu jedem Item zeigten in der Regel so viele, als Distraktoren geeignete, häufig auftretende Fehler, daß damit der Distraktorenbedarf weitgehend gedeckt werden konnte. (Neben dem zutreffenden Lösungswort sollten stets drei Distraktoren angeboten werden.) Die Fehlerliste der Vorversuche wurde dann klassifiziert, wodurch wir zu einem begrifflich tauglichen Suchschema für fehlende Distraktoren gelangten.

Dieses Schema umfaßt neun Kategorien von Fehlermerkmalen; es zeigte sich nämlich, daß Fehler oft nicht allein vermittels einer einzigen Kategorie faßbar waren, wengleich auch dies möglich war. Die neun Kategorien sind wie folgt beschrieben worden:

1. Der Fehler (bei uns: Distraktor) besitzt aufgrund des vorlaufenden Satzabschnitts (nach Form und Inhalt) eine hohe Übergangswahrscheinlichkeit, wobei der nach der Lücke stehende Teil des Satzes außer Betracht bleibt.

Beispiel:

„Meine Eltern können nur wenig (Geld / arbeiten / Deutsch / kommen) ...“
(Vorlesewettbewerb, Item 15; das Fehlermerkmal trifft auf „arbeiten“ zu.)

2. Der Fehler besitzt eine hohe Auftretenswahrscheinlichkeit in einem begrenzten Satzausschnitt, in dem er einen Sinn erzeugt, der mit dem vollständigen Satz unvereinbar ist.

Beispiel:

Laut und deutlich (sahen / riefen / war / herein) sie Tina.
(Vorlesewettbewerb, Item 11; das Fehlermerkmal trifft auf „sahen“ zu, wenn die beiden ersten Wörter des Satzes aus der Lösungsabwägung ausgeklammert werden.)

Die Fehlermerkmale 1 und (bedingt auch) 2 sind von *Potter* (1982) untersucht worden; Probanden mit schwachen Leseleistungen schneiden schlechter ab, wenn die kritische Information eines Satzes nach der Lücke steht; bei Probanden mit guten Leseleistungen wirkt sich diese Positionsdivergenz der kritischen Information nicht aus.

3. Der Fehler ergibt einen grammatisch korrekten, für sich sinnvollen Satz, der aber im Lichte des weiteren Textumfeldes oder des Gesamttextes nicht haltbar ist.

Beispiel:

(Da / Tina / Mutter / Lehrer) dachte: Das konnte nicht wahr sein!
(Vorlesewettbewerb, Item 3; das Fehlermerkmal trifft auf „Mutter“ zu.)

4. Der Fehler ergibt einen sinnvollen (Teil-)Satz, wenn man eine gewisse grammatische Laxheit toleriert, wobei sich der eigentliche Sinn des (Teil-)Satzes unter Umständen ändert.

Beispiel:

Manche Vögel bauen (sie / ins / neu / ihr) Nest auch auf dem Dach oder unter dem Dach.
(Vögel, Item 7; das Fehlermerkmal trifft auf „neu“ zu.)

5. Der Fehler ergibt eine sinnhaltige Wortfolge, wenn ein weiteres Wort der Mehrfachauswahl mitgelesen wird.

Beispiel:

Kleine Glühlampen braucht (viele / man / Kinder / weg) für Fahrradlampen und für Taschenlampen.
(Batterie und Glühlampe, Item 6; das Fehlermerkmal trifft auf „viele“ zu, wobei „Kinder“ mitgelesen wird und „braucht“ (statt „brauchen“) vernachlässigt wird.)

6. Der Fehler erscheint deshalb als naheliegende Lösung, weil ein im Satz enthaltenes Wort damit wiederholt wird.

Beispiel:

Das Nestinnere legen sie mit Federn und feinem Heu aus, damit es schön (wird / warm / Nest / weich) und warm wird.
(Vögel, Item 15; das Fehlermerkmal trifft auf „wird“ und „warm“ zu. Dieses Beispiel zeigt aber auch die Überschneidung mehrerer Kategorien: zwischen dem Vorlauf und beiden Distraktoren besteht auch eine hohe Übergangswahrscheinlichkeit (Kategorie 1), und es ergibt sich überdies aus Vorlauf und jedem der beiden Distraktoren ein sinnvoller Satz (Kategorie 4)).³

7. Der Fehler steht zu einem Wort (oder mehreren Wörtern) im Satz in einer engen assoziativen Nähe, die fallweise auch stark emotionaler Natur sein kann. Fehler dieser Kategorie sind nicht aus dem Thema des Gesamttextes erklärbar.

Beispiel:

„Wir wollen ja (eines / später / nachts / nicht) Tages wieder heim.“
(Vorlesewettbewerb, Item 18; das Fehlermerkmal trifft auf „nicht“ zu; das Eingeständnis, nicht (wieder) in das Herkunftsland der Eltern zu wollen, ist gerade unter Türken tabuisiert, im Gegensatz etwa zu Jugoslawen.)

8. Der Fehler scheint durch seine thematischen Merkmale dem Thema des Textes als die wahrscheinlichste Lösung diffus zugeordnet zu werden; es liegt ein so sehr oberflächliches Verständnis vom Text vor, so daß der thematische Reiz des Distraktors zum einzig verlässlichen Anhaltspunkt wird.

Beispiel:

Wenn eine Glühlampe herunterfällt, geht (Elektriker / weg / nicht / sie) meist kaputt.
(Batterie und Glühlampe, Item 15; das Fehlermerkmal trifft auf „Elektriker“ zu; es wurde von 37% von N = 489 Schülern angekreuzt (Key-Wahl 31%!))

9. Der Fehler stellt eine so grob abwegige, auch durch keine andere Kategorie gedeckte „Lösungsmöglichkeit“ dar, daß er für – meist schwache – Probanden nach der generalisierten Antwortneigung infrage kommt: „Die unwahrscheinlichste Vorgabe muß es sein.“

Beispiel:

Große Glühlampen verwendet man in (mir / Haus / Schrauben / Wohnungen) für Deckenlampen und Stehlampen.

(Batterie und Glühlampe, Item 5; das Fehlermerkmal trifft auf „mir“ zu. Schüler, die diesen Distraktor ankreuzen, erreichen einen Testscore von $T = 44,6$, d. h. sie liegen etwa eine halbe Standardabweichung unter dem Mittelwert ($M = 50$, $s = 10$) auf der T-Skala, also etwa beim 33. Prozentrangplatz. Die mittleren T-Werte der solche Distraktoren ankreuzenden Schüler liegen zwischen 39,3 und 49,5).

Bereits eingangs haben wir auf die Ziele des Einsatzes der zu entwickelnden Instrumente verwiesen und überdies dargelegt, welche sprachplanerischen bzw. sprachpolitischen Prämissen uns angezeigt erschienen; in diesem Sinne wurde folgendes begriffliches Kontinuum für die Textsortensuche und -auswahl aufgestellt:

1. ein literarischer Text, der Umgangssprache und Alltagssituation jenseits spezifischer Unterrichtsinhalte wiedergibt;
2. ein Repräsentant „didaktischer Texte“ (Lehrbuch, Arbeitsbogen u. ä.), dessen Thema allerdings wenig kompliziert sein sollte, im Herkunftsmilieu der Schüler, mindestens aber ihrer Eltern, zum Alltagswissen zählen sollte und mutmaßlich nicht geschlechtersegregierend wirken sollte;
3. ein „didaktischer Text“ zum Unterrichtsthema (Einführung in die Elektrizitätslehre), der das Unterrichtsprogramm sprachlich abdecken sollte.

Zur dritten Textsorte mußte es zur Kontrolle des erworbenen physikalischen Wissens – unabhängig von seinem Verständnis in Gestalt deutscher Sätze – eine Kontrollinstanz geben; zu diesem Zweck wurde hier ein zweisprachiger Text konstruiert, der unter Verwendung von wenig und sehr einfacher Sprache (auch Türkisch) sowie reichlicher Illustration tendenziell das „reine“ Physikwissen erfassen sollte. Dieser (ebenfalls mit Mehrfachauswahlvorgaben operierende) Test wird hier nicht dargestellt werden. (Siehe hierzu *Hagemeister, Preibusch, Schuricht* und *Seyhan* 1982.)

Mit diesem Kontinuum mußte es möglich sein, die unterrichtlich beeinflussten Sprachfelder (Physikunterricht) sowohl hinsichtlich Sprache (Deutsch) als auch hinsichtlich „reinen“ Wissens zu kontrollieren, die Auswirkungen eines bestimmten Unterrichtsthemas auf einen zwar thematisch abseitigen, vom Texttypus aber vergleichbaren Textes abzuschätzen sowie schließlich auch Veränderungen im allgemeinen umgangssprachlichen Alltagsdeutsch zu verzeichnen. Ausgewählt wurden deshalb als

- Repräsentant der Textsorte 1 „Vorlesewettbewerb“
- Repräsentant der Textsorte 2 „Die Vögel“
- Repräsentant der Textsorte 3 „Von Batterie und Glühlampe“ (Dieser Text

wurde von uns als eine Art Programmzusammenfassung geschrieben und bearbeitet.)

In diesen Texten (abzüglich Überschriften) wurde dann regelmäßig im Text „Der Vorlesewettbewerb“ (VLW) jedes 8. Wort gelöscht, im Text „Die Vögel“ (DV) jedes 7. Wort und im Text „Batterie und Glühlampe“ (BGX) jedes 8. Wort. Die höhere Löschfrequenz bei DV wurde notwendig, als sich der Text als zu leicht erwies, wenn nur jedes 8. Wort gelöscht würde; er erwies sich auch so in der Hauptuntersuchung immer noch als der leichteste Text. Damit ergaben sich drei Cloze-Tests mit 27 (VLW), 25 (DV) bzw. 38 Items (BGX).

3. Einsatz der Instrumente

Es wurde davon ausgegangen, daß die Lehrer mit dem Einsatz des Programms im Januar 1981 beginnen würden; der Vortest mit dem Instrumentarium sollte in diesen ersten Tagen des Unterrichtsbeginns zu dem Thema erfolgen. Tatsächlich zeigte sich später, daß einige Lehrer (insbesondere auch türkische Lehrer) das ihnen zuvor bereits ausgelieferte Material schon vor diesem Termin eingesetzt hatten, z.T. in Sorge, „sie könnten zu schlecht abschneiden“, obwohl der Zweck der Untersuchung ausführlich und plausibel vermittelt zu sein schien. Entsprechend streute auch der Abschluß der Arbeit mit dem Programm in den 10 beteiligten Klassen. Einige Klassen hatten tatsächlich schon vor dem Vortest, andere bis April/Mai 1981 das Programm absolviert, begannen teilweise mit Wiederholungen, andere brauchten bis in den Juni hinein die Zeit zur Bearbeitung. Im Juli wurden die Tests ein zweites Mal administriert, und zwar außer in den Programmklassen auch noch in 24 Klassen, die nicht mit dem Programm gearbeitet hatten. (Vgl. zur Arbeit mit dem Programm *Hagemeister* u. a. 1981, 1982.) Trotz schulaufsichtlicher Sicherstellung der Versuchsbedingungen und der entschiedenen Kooperationsbereitschaft aller Beteiligten waren nur für 106 von 193 + ? involvierten Schülern der Programmklassen (vermeintliche) Vor- und Nachprogramm-Daten zu erlangen. Entsprechend muß unsere Ergebnisdarstellung Shift-Aspekte weitgehend offen lassen.

Die Testbatterie und ein Fragebogen zur Person⁴ wurden in einer vormittäglichen Sitzung von bis zu 110 Minuten bearbeitet. Dieser Zeitrahmen gewährte den Schülern eine Bearbeitung der Aufgaben ohne Zeitdruck; sie wurden ausdrücklich aufgefordert, erst abzugeben, wenn sie meinten, fertig zu sein.

4. Ergebnisse

Die folgende Darstellung der Ergebnisse konzentriert sich auf die Daten aus der Erhebung II (Juli 1981).

Tabelle 1 gibt die wichtigsten Test-Kennwerte auf der Grundlage der Ergebnisse von N = 489 türkischen Schülern wieder. Die Reliabilitätswerte (interne Konsistenz-Indizes), die Tabelle 1 für die Erhebung II ausweist, sind praktisch exakte Reduplikationen der Werte aus Erhebung I.

Berücksichtigt man die formale Vorgehensregel bei der Bestimmung der Items (Lücken) bei Cloze-Tests, dann darf die durchgängige Größenordnung der Konsistenz-Indizes als erstaunlich gelten. Zweifellos spielt dabei die hier verwendete Art der Mehrfachauswahlen-Vorgabe eine erhebliche Rolle, doch war das nicht vorauszusehen. Für die Texte VLW und DV sind auch die Test-Retest-Korrelationen ausgezeichnet zu nennen, bedenkt man außer den üblichen Werten für diesen Reliabilitätstypus die enorme Instabilität des Feldes. Die wesentlich niedrigere Test-

Bezeichnung der Kennwerte	Cloze-Tests			Wissenstest BGS
	VLW	DV	BGX	
arithmetisches Mittel	12,03	12,54	16,82	17,22
Standardabweichung	6,61	6,63	10,28	5,96
Reliabilität:				
- $r_{tt_{korr}}$.89	.91	.94	.90
- $r_{tt_{KR\ 20}}$.89	.90	.94	.87
- Test (I)	(n = 126)	(n = 126)	(n = 106)	
Retest (II)	.70	.69	.33	—*

* Angabe nicht möglich, weil dieser Test nach der Erhebung I bearbeitet werden mußte.

Tabelle 1: Ausgewählte Test-Kennwerte für die drei Cloze-Tests und dem Wissenstest Physik

Retest-Reliabilität von $r = .33$ für BGX steht zweifellos mit der unterrichtlichen Intervention in Zusammenhang. Es gibt allerdings Gründe zu vermuten, daß die Mädchen die wesentliche Quelle der geringen Test-Retest-Reliabilität sind. Ursächlich ist hier wohl nicht nur eine mutmaßliche Thema x Geschlechtsrolle – Interaktion, sondern eine bereits vor Unterricht, grundsätzlich andere Vorwissens-, Einstellungs- und Motivationslage als bei den Jungen. Es ist denkbar, daß für türkische Mädchen *während* des Unterrichts das Thema, seine Möglichkeiten der Emanzipation, seine praktische Relevanz und seine Lehr-Lern-Struktur zu Gesichtspunkten werden, die den Determinus von Vorwissen auf Lernerfolg in besonderem Maße außer Kraft setzen. Wir haben Gründe zu der Annahme, daß dies nur einen kleinen Teil der Mädchen betrifft, diesen indes desto gründlicher.

Schwierigkeits- indizes	Cloze-Tests			Wissenstest BGS
	VLW	DV	BGX	
Minimum	.16	.20	.21	.33
Median	.44	.52	.435	.66
Maximum	.64	.73	.75	.86
p > .90	—	—	—	—
p < .10	—	—	—	—
Diskriminations- koeffizienten				
Minimum	.21	.28	.05	.19
Median	.61	.62	.725	.54
Maximum	.80	.78	.89	.79
$r_{bis} < .30$	2	1	1	1

Tabelle 2: Übersicht über die Verteilung der Schwierigkeitsindizes (p) und Diskriminationskoeffizienten (r_{bis}) in den drei Cloze-Tests und dem Wissenstest Physik

Tabelle 2 vermittelt einen Überblick über die beiden hauptsächlichen Itemselektionskennwerte. Schwierigkeitsindizes (Lösungsquoten) von weniger als 10 oder mehr als 90% treten, wie ersichtlich ist, nicht auf. Bei den Diskriminationskoeffizienten fallen für die Cloze-Tests einmal 2 Items und zweimal 1 Item unter die konventionelle $r_{bis} = .30$ -Marke, und zwar

- im Text VLW die Items 4 (.21) und 20 (.27),
- im Text DV das Item 3 (.28) und
- im Text BGX das Item 9 (.05).⁵

Zusammenfassend läßt sich sagen, daß das verhältnismäßig einfache Verfahren der Cloze-Test-Konstruktion, selbst bei Mehrfachauswahl-Vorgaben, zu so befriedigenden Instrumenten führt, daß mehr herkömmliche Sprachtest-Item-Formate dahinter im Schnitt erheblich zurückbleiben. Entscheidend aber scheint uns, damit nochmals auf *Spolsky* zurückgreifend, daß die herkömmlichen Itemformate mit Cloze-Test-Items hinsichtlich Inhaltsvalidität nicht entfernt konkurrieren können. Dies gilt mindestens für vergleichbare Forschungszusammenhänge und ein Sprachverständnis, bei dem der instrumentelle Charakter von Sprache als soziales Medium im Vordergrund steht.

Zum Zeitpunkt II wurden auch die 95 *deutschen Schüler* getestet, die die Majorität der Schüler in den von uns untersuchten gemischten Klassen stellten. Naturgemäß liegen die Lösungswerte dieser Schüler sehr extrem, so daß sich eine Berechnung und Wertung der Item-Kennwerte verbietet. Ihre Scores sind gleichwohl nicht durchgängig perfekt, ein Ergebnis, das für andere fremdsprachliche Tests, die von Nativen absolviert wurden, bekannt ist (vgl. etwa die Untersuchung von *Clark* 1977 für den *TOEFL* oder *Alderson* 1981 für Cloze-Tests sowie *Darnell* 1968, der sein Scoring-Verfahren für Cloze-Tests auf dieser Tatsache aufbaut). Auffallend ist indes eine über die drei Cloze-Tests hin bis zum Wissenstest prägnant werdende Quote von fehlerhaft gelösten Items ($p < 80\%$; dies ist auch die von *Clark* gewählte Grenze); je stärker über Sprachbeherrschung hinaus Sachwissen ins Spiel kommt, desto häufiger werden Items fehlerhaft gelöst (VLW 3,7% aller Items, DV 12,0%, BGX 18,4% und BGS 48,2%). Dabei bleibt zu berücksichtigen, daß nach deutschen Maßstäben sämtliche Texte (und Aufgaben des Wissenstests) niveaumäßig dem Grundschulbereich entstammen. Diese Daten stellen rein sprachimmanentes Lösungsverhalten mindestens partiell infrage.

Wie bereits dargestellt, wurden die türkischen Schüler auch nach dem Umfang ihrer Kontakte zu gleichaltrigen Deutschen und zu deutscher Lektüre befragt. Die Ergebnisse dieser Befragung liegen für $N = 487$ Schüler vor. Wir haben den Zusammenhang zwischen der Zahl der Kontakte/Lektüren und den Ergebnissen in den Cloze-Tests grafisch untersucht.⁶ Danach läßt sich nicht die Plausibilitätsformel verallgemeinern: Je mehr, desto besser. Sowohl die Art des Kontaktes als auch sein Umfang können mit sehr unterschiedlichem Erfolg in den Cloze-Tests einhergehen. Das Hausaufgabenmachen mit Deutschen hat einen neutralen bis drastisch abträglichen Effekt auf die Scores. Möglicherweise werden dabei Hausaufgaben der deutschen Schüler nur abgeschrieben. Der negative Effekt steigt von VLW über

DV bis BGX an. *Schulweg mit Deutschen und Deutsche als beste Freunde* sind mit den Cloze-Test-Scores eindeutig vorteilhaft vergesellschaftet, solange die Kontakte-Anzahl 2 nicht überschritten wird; höhere Kontakte-Anzahlen können für eine „Integration“ in sozial auffällige deutsche Cliques stehen, besonders dann, wenn man die soziale Struktur der deutschen Restbevölkerung in den Bezirken Kreuzberg und Teilen der Bezirke Wedding und Tiergarten im Auge hat. Gleichwohl sind zwei solche Kontakte noch vorteilhafter als nur einer.

In noch diffizilerem Verhältnis steht die Anzahl der *Nachmittagskontakte* mit den Testergebnissen; optimal scheinen hier zwei deutsche Kontakte und günstig noch drei zu sein. Über höhere Kontaktzahlen kann Verlässliches kaum gesagt werden. Ein einzelner Nachmittagskontakt mit einem deutschen Kind scheint indes neutral bis eher abträglich mit den Cloze-Testergebnissen vergesellschaftet zu sein. Es ist nicht auszuschließen, daß hier problematische Persönlichkeitszüge einzelgängerischer deutscher Kinder ins Spiel kommen, die mit der Wahl eines ausländischen, sozial also unterlegenen Spielgefährten ihren Niederschlag findet. Zudem ergibt sich zwischen *einem* deutschen Kind (möglicherweise auch Erwachsenen!) und *einem* türkischen Kind einerseits eine andere, das Sprachverhalten determinierende Sozialsituation als zwischen *mehr als einem* deutschen Kind und *einem* türkischen Kind andererseits. Während im ersten Fall das einzelne deutsche Kind sich eines Pidgin-Deutchs gegenüber dem türkischen Kind bedient, wird in einer Gruppensituation, in der deutsche Kinder in der Majorität sind, zumindest zwischen diesen (eine) Hochdeutsch (-Näherung) gesprochen; mindestens zu hören bekommt das türkische Kind ein besseres Deutsch. Die *Lektüre deutscher Bücher* in der Freizeit zeigt einen gleichmäßig positiven Zusammenhang mit allen drei Cloze-Test-Scores. Er ist geringer zu veranschlagen als Schulwegkontakte und Freundschaftsbeziehungen oder zwei Nachmittagskontakte; je mehr deutsche Titel allerdings genannt werden, desto besser fallen die Testergebnisse aus.

Am Schluß dieses Abschnitts soll Diagramm 1 verdeutlichen, auf welche Weise die Cloze-Test-Ergebnisse die sehr unterschiedliche *Dauer des Schulbesuchs in Deutschland* abbilden. (Zum Zwecke der besseren Vergleichbarkeit wurden die Rohscore-Mittelwerte jeder Aufenthaltsdauer-Gruppe in Prozentscores vom jeweiligen Test-Maximalscore transformiert.) Erstaunlich ist schon auf den ersten Blick der ziemlich gebündelte Verlauf aller drei Verläufe. Man wird sagen dürfen, daß sich die Fähigkeit zur Bewältigung aller drei Textrepräsentanten im wesentlichen gleichförmig entwickelt. Der besonders steile Anstieg der Verläufe am Anfang (von 0 auf 1 Jahr deutschen Schulbesuchs) liegt im Bereich des unmittelbar Einsichtigen. Indes zeigen die Verläufe zweimal so etwas wie ein Plateau, wo die Verbesserung zu stagnieren scheint, nämlich für 2- und 3jährigen deutschen Schulbesuch sowie für mehr als 6jährigen deutschen Schulbesuch. Hält man dagegen, daß für Schulbesuchsdauer in der Türkei die Gruppen „kein türkischer Schulbesuch“ (22,4%) und „5 Jahre türkischer Schulbesuch“ (26,9%) weit aus allen übrigen Häufigkeiten herausragen, dann haben wir es knapp zur Hälfte mit Schülern zu tun, die entweder nur hier zur Schule gegangen sind oder die 5jährige türkische Grundschule absolviert haben.

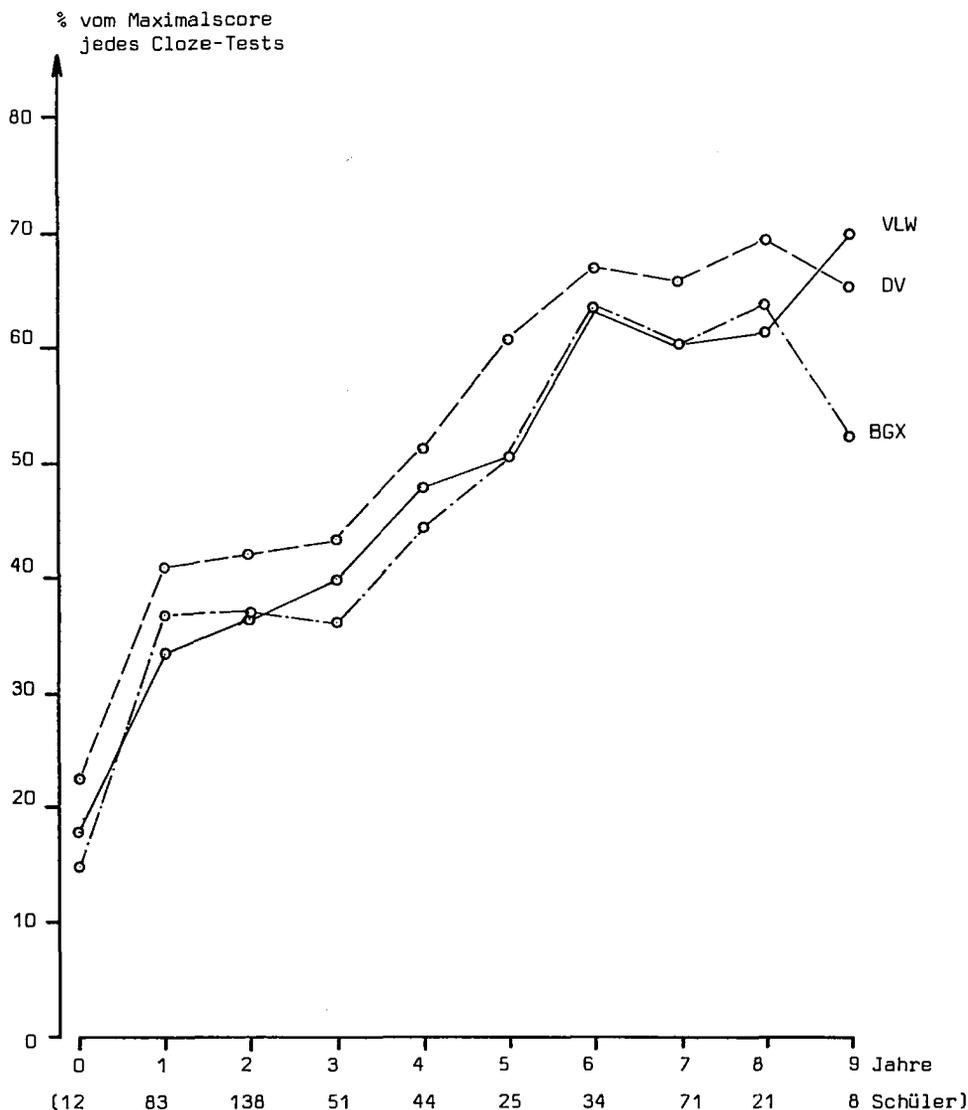


Diagramm 1: Mittelwerte in Prozent vom Maximalscore in Abhängigkeit von der Dauer des Schulbesuchs in Deutschland

Wer also als türkisches Kind die 6jährige Berliner Grundschule besucht hat, erreicht ein Fertigungsplateau, das augenscheinlich in weiteren 1, 2 oder 3 Jahren auf der Sekundarstufe I, d. h. fast immer auf der Hauptschule und dort meist noch in inferioren Gruppierungen, nicht mehr wesentlich zu verbessern ist. Dieses Faktum dürfte, da wir – mit Ausnahme einer einzigen 6. Klasse – nur Schüler der Sekundarstufe I untersucht haben, auf dem niedrigeren Plateau verschärft zum Ausdruck kommen: als Gerade-noch- oder Nicht-mehr-Grundschüler nach Deutschland gekommen, erfolgt in den Klassen, in die diese 12- bis 16jährigen Türken eingeschult

werden (also in den etwa noch bis zu 3 Jahren Schule) keine nennenswerte Verbesserung der Fertigkeiten, die die Cloze-Tests erfassen.

Wenngleich hier nicht verschwiegen werden soll, daß mit 12 oder 13 Jahren bei den untersuchten Schülern so etwas wie aktives Ausgrenzungsverhalten gegenüber der Kultur und Sprache des Gastlandes zu beginnen scheint, so kann doch nicht übersehen werden, daß die Zusammenfassung dieser Schüler in Ghettoschulen mit dem 7. Schuljahr dieses Ausgrenzungsmotiv als verbleibende Perspektive nicht nur stimuliert, sondern auch die Grundlagen zu weiterer Integration objektiv schwächt. Dagegen kann auch nicht eingewandt werden, daß nach der Grundschulzeit so etwas wie eine – ungerechtfertigte – Selbstzufriedenheit mit dem oder Selbstbeschränkung auf den erreichten Fertigungsstand im Deutschen einsetzt; die Schüler, die überhaupt erst auf Sekundarstufe I mit einer deutschen Schule in Berührung kommen, sind trotz ihrer alles andere als zureichend erlebbaren Deutschkenntnisse offenbar ebenfalls dem Nullwachstum in der Beherrschung der Landessprache unterworfen, wie es die Hauptschulen erzielen, und zwar wohl eher nolens als volens.

5. Zusammenfassung

Die Ergebnisse zeigen, daß mindestens für den hier gegebenen Zweck, für die hier angezielte Schulstufe und für Türkisch als Erstsprache der Schüler der untersuchte Texttypus eine reliable und valide Instrumentenklasse darstellt. Die Fragen der Sensitivität und der Auswirkungen unterrichtlicher Intervention sind trotz des dahingehend ausgelegten Untersuchungsdesigns hier nicht beantwortbar, und zwar aufgrund von Störungen im Feld. Auf jeden Fall scheint der hier zugrundegelegten Textsorten-Skala eine didaktisch-psychische Realität zu entsprechen, die weiterer, kontrollierter Erforschung bedarf.

Völlig verfehlt wäre es, in dem Instrumentarium den Repräsentanten einer Klasse von Selektionshilfsmitteln zu sehen. Wäre solcher Mißbrauch auch nicht zu verhindern, so erfolgte er doch in einer durch die Tatsachen nicht zu rechtfertigenden Weise: Es ist deutlich geworden – was hier nicht Gegenstand der Darstellung war –, daß es eine prädeterminierte Hierarchie in den Leistungsniveaus von Vorbereitungsklassen über besondere Klassen hin zu gemischten Klassen (mit maximal 25 % Ausländern) gibt. Es gibt, so meinen wir, keine Grundlage dafür, diese gesellschaftlich bedingten Umstände über den Weg der individuell zu verantwortenden Zuschreibungsweise von „Erfolg“ oder „Versagen“ in scheinbare Objektivität zu verwandeln⁷.

Anmerkungen

¹ Im Rahmen dieses Aufsatzes kann unmöglich ein Überblick über die Cloze-Test-Forschung gegeben werden. Es sei deshalb auf zwei Bibliografien verwiesen, und zwar *Fram* (1972) und *Oller* (1975).

² Mit Bedacht sei auf *Darnell* (1968) verwiesen, weil der interessierte Leser hier auch eine ausführliche Darstellung eines der meistdiskutierten Themen in Sachen Cloze-Tests findet, nämlich die Fürs und

- Widers unterschiedlicher Scoring-Methoden. (Das in der Bibliografie zitierte ERIC-Dokument ist im Pädagogischen Zentrum greifbar.) Ein auch noch eher kritischer Artikel von leichter Zugänglichkeit und jüngeren Datums ist der von *Alderson* (1981). Ein Abdruck der Endfassung der von uns verwendeten Cloze-Tests kann beim erstgenannten Autor angefordert werden.
- ³ In unzulässigem Transfer muttersprachlicher Erfahrung werden Wortwiederholungen zur Fehlerquelle; der Fehlertypus wird auch ohne Vorgabe (also nur bei Lücke) produziert. Wir haben in 11 der 90 Items der Cloze-Tests insgesamt 13 Wortwiederholungen als Distraktoren verwendet. Auf sie entfielen zwischen 6 und 22% der Antworten.
- ⁴ Der Fragebogen zur Person erhob Geschlecht, Alter, Nationalität, Aufenthaltsdauer in Deutschland, Schulbesuchsdauer in Deutschland und in der Türkei, verschiedene Integrationsindizes (deutsche Lektüre, Schulweg mit deutschen Kindern, Hausaufgaben mit deutschen Kindern, deutsche Kinder als beste Freunde) sowie die Vertrautheit mit verschiedenen Örtlichkeiten im Stadtgebiet als Index für räumliche Expansion über die Ausländer-Quartiere hinaus. Diese Örtlichkeiten (Kaufhäuser, Straßen, Parks, Freizeitziele) liegen bis zu über 25 km von den Ausländer-Quartieren entfernt. Der Fragebogen war in Deutsch und Türkisch abgefaßt.
- ⁵ Bei BGX Item 9 ($r_{\text{bis}} = .05$) ist ein spezielles Problem des Physik-Fachjargons berührt. Im Türkischen erfolgt die entsprechende Begriffsbildung exakt umgekehrt: „offen“ bedeutet fließenden Strom, „geschlossen“ abgeschalteten, nicht fließenden Strom. Schüler, denen der korrekte Sachverhalt unter Rekurs auf die Muttersprache vertraut war, konnten hier in eine Schere zwischen „Wissen“ und Vertrautheit mit der sehr spezifischen deutschen Definition durch die Physiker oder Elektrotechniker geraten. – Dieses Item ist auch nur von 23% der deutschen Schüler ($N = 95$) gelöst worden.
- ⁶ Eine korrelationsstatistische Analyse verbot sich wegen der extremen Verteilungen der Kontakte/Lektüren. Insgesamt muß die Zahl der Kontakte gering genannt werden, wobei Mädchen immer noch stärker ghettoisiert leben als Jungen. So machen 97,1% aller untersuchten türkischen Schüler nie Hausaufgaben mit deutschen Kindern zusammen, 92% gehen den Schulweg stets ohne deutsche Kinder, 91,8% nennen keinen Deutschen als besten Freund, 89,9% haben auch nachmittags beim Spielen keine deutschen Kontakte und immer noch 53,8% nennen keine deutsche Freizeitlektüre in Buchform.
- ⁷ Dabei beschreibt diese Schulklassen-Typen-Hierarchie nicht einmal alle systematischen Verzerrungen, die weit über individuelles Versagen oder Bemühen hinausreichen; wir werden zeigen (*Preibusch, Hagemester, Schuricht, Seyhan* 1982), wie sehr und wie tiefgreifend die gesamte Thematik durch ein nach mitteleuropäischen Maßstäben schwer vorstellbares Geschlechterproblem belastet ist. Ferner bleiben – auch unter den Türken – nicht unerhebliche sozio-ökonomische Unterschiede zwischen den familiären Hintergründen der Schüler u.v.a.m.

Literatur

- Alderson, J. C.*: Native and Nonnative Speaker Performance on Cloze Tests; in: *Language Learning*, 30, 1/1981, 59–76.
- Campbell, D. T., und Stanley, J. C. (1963)*: Experimental and Quasi-Experimental Designs for Research on Teaching; in: *Handbook of Research on Teaching* ed. by N. L. Gage. Chicago (Rand McNally) ⁵1965, 171–246.
- Clark, J. L. D.*: The Performance of Native Speakers of English on the Test of English as a Foreign Language; TOEFL Research Reports, Report 1; Princeton N. J. (ETS) 1977.
- Darnell, D. K.*: The Development of an English Language Proficiency Test of Foreign Students, Using a Clozentropy Procedure. Final Report. Washington D.C. (US Dept. of Health, Education, and Welfare, Office of Education, Bureau of Research) Oct. 1968.
- Einstufungs- und Diagnostiktest*: Das Zertifikat Deutsch als Fremdsprache. Einstufungs- und Diagnostiktest. Hinweise für den Kursleiter, Auswertungsschablone, Testheft. Bonn, Frankfurt a.M. (Deutscher Volkshochschulverband e.V.) 1981.
- Fliegner, J., und Gogolin, I.*: Sprachstandsmessung bei Schulanfängern. Düsseldorf (Pädagogisches Institut der Landeshauptstadt Düsseldorf) 1982.
- Fram, R. D.*: A Review of the Literature Related to the Cloze Procedure, 1972, 264 S.
- Hagemester, V., Preibusch, W., Schuricht, K., und Seyhan, H.*: Diskussion von Strategien für erfolgreichen Physikunterricht bei türkischen Schülern; in: *Zur Didaktik der Physik und Chemie. Probleme und Perspektiven*. Vorträge auf der Tagung für Didaktik der Physik/Chemie in Berlin, Sept. 1981; hrsg. v. H. Härtel. Alsbach (Bergstraße) (Leuchtturm-Verlag) 1981.

- Hagemeister, V., Preibusch, W., Schuricht, K., und Seyhan, H.:* Physikunterricht bei türkischen Schülern: Ergebnisse einer empirischen Untersuchung und Empfehlungen für den Unterricht mit ausländischen Schülern. Berlin (Pädagogisches Zentrum) erscheint demnächst.
- Jonz, J.:* Improving on the Basic Egg: The M-C Cloze; in: *Language Learning*, 26, 2/1976, 255–265.
- Oller, J. W. jr.:* Research with Cloze Procedure in Measuring the Proficiency of Non-Native Speakers of English; an Annotated Bibliography, 1975, 32 S. ERIC Clearinghouse on Languages and Linguistics, Arlington, Va.
- Porter, D.:* Modified Cloze Procedure: A More Valid Reading Comprehension Test; in: *English Language Teaching Journal*, 30, 2/1976, 151–155.
- Portz, R., und Pfaff, C. W.:* SES – Soziolinguistisches Erhebungsinstrument zur Sprachentwicklung. Berlin (Pädagogisches Zentrum) 1981.
- Potter, F.:* The Use of Linguistic Context: Do Good and Poor Readers Use Different Strategies?; in: *British Journal of Educational Psychology*, 52, 1982, 16–23.
- Preibusch, W.:* Frühbeginn des Englischunterrichts als didaktisches Problem. Empirische Untersuchung über die Beschaffenheit des erlernten Englisch in Abhängigkeit vom Curriculum. Weinheim, Basel (Beltz) 1979.
- Preibusch, W.:* Hörverstehen in Englisch als Fremdsprache und Sprechervielfalt im Unterricht. Grünwald (FWU) erscheint demnächst (Reihe AV-Forschungsberichte).
- Preibusch, W., Hagemeister, V., Schuricht, K., und Seyhan, H.:* Wer sind „die türkischen Schüler“? Ergebnisse einer Analyse von Daten über Deutschkenntnisse und ausgewählte Integrationsindizes. (In Vorbereitung)
- Spolsky, B.:* What does it mean to know a language, or how do you get someone to perform his competence? Paper presented on the Second Conference on Problems in Foreign Language Testing, University of Southern California, November 7–9, 1968.
- Swain, M.:* Time and Timing in Bilingual Education; in: *Language Learning*, 31, 1/1981, 1–14.

Verfasser:

Wolfgang Preibusch, Dr. Volker Hagemeister, Pädagogisches Zentrum, Uhlandstr. 96–97, 1000 Berlin 31; Klaus Schuricht, Muskauer Str. 35, 1000 Berlin 36; Hayrettin Seyhan, Freie Universität, Insitut für Islamwissenschaft, Boltzmannstr. 4, 1000 Berlin 33.