

# Empirie

## Experimentelle Überprüfung des Bogus-Pipeline-Paradigmas: Einstellungen gegenüber Türken, Deutschen und Holländern

HANS DIETER MUMMENDEY, HEINZ-GERD BOLTEN und MARGRET ISERMANN-GERKE

Universität Bielefeld

Mittels zweier Bogus-Pipeline-Versuchsanordnungen (simulierte elektromyographische Einstellungsmessung) erfolgt eine auf deutsche Verhältnisse zugeschnittene Replikation des Experimentes von SIGALL & PAGE (1971) zur Gültigkeit des Bogus-Pipeline-Paradigmas. 108 Personen, aufgeteilt auf drei experimentelle Bedingungen (BPL-Alt, BPL-Neu, Paper-Pencil) beurteilen drei ethnische Gruppen (Türken, Deutsche, Holländer) hinsichtlich elf in einer Voruntersuchung gewonnenen Adjektive. Die erwarteten Effekte (Wechselwirkung zwischen Einstellungsobjekt = ethnische Gruppen einerseits, und Meßmethode = experimentelle Bedingungen andererseits) zeigen sich sowohl bei globalen Einstellungswerten als auch bei Anwendung der multivariaten Varianzanalyse (MANOVA).

A German replication of SIGALL & PAGES (1971) experiment testing the validity of the bogus-pipeline-paradigm has been performed by comparing 108 subjects' judgments about three ethnic groups (Turks, Germans, Dutch) under either of three experimental conditions (two bogus-pipeline conditions, and a paper-pencil rating condition). As expected there were significant interaction effects between attitude objects (ethnic groups) and attitude measurement techniques (bogus-pipeline vs. paper-pencil) not only when composite scores of 11 adjective attitude judgments were computed but also when multivariate analyses of variance (MANOVA) were applied. The results thus give further support to the bogus-pipeline paradigm of attitude measurement.

### Problemstellung

Mit dem «Bogus-Pipeline-Paradigma» liegt seit Anfang der 70er Jahre ein experimentalpsychologischer Ansatz vor, um Reaktionstendenzen der «Sozialen Erwünschtheit» (social desirability) bei der Einstellungsmessung zu reduzieren.

Das grundlegende Experiment für das Bogus-Pipeline-Paradigma wurde 1971 von SIGALL & PAGE veröffentlicht. Die Autoren verteilten 60 weiße Amerikaner nach Zufall auf die Zellen eines  $2 \times 2$ -Versuchsplanes mit den unabhängigen Faktoren «Meßmethode» und «Einstellungsobjekt»; dabei erwarteten sie *Wechselwirkungen zwischen der Meßmethode und der Beurteilung des Einstellungsobjektes*. Eine von

den beiden Abstufungen des Faktors «Meßmethode» war die Einstellungsmessung gemäß dem Bogus-Pipeline-Paradigma: Die Versuchsperson wird im Experiment zunächst von der Fähigkeit einer Versuchsanordnung, subjektive Bewertungen von Einstellungsobjekten physiologisch zu messen, überzeugt. Anschließend soll sie zwecks Untersuchung, «in welchem Maße jemand seine tatsächlichen Empfindungen kennt», die Meßwerte des Gerätes bei der Darbietung einer Reihe von Einstellungsgegenständen vorhersagen; die Vorhersage dieser Apparate-Werte wird dann als relativ beschönigungsfreie Einstellungsäußerung aufgefaßt. Die zweite Abstufung des Faktors «Meßmethode» besteht in der Einstellungsmessung unter üblichen Papier-Bleistift-Bedingungen, d.h., die Versuchsperson gibt ihr Einstellungsurteil durch Ankreuzen auf einer Rating-Skala ab.

*Anmerkung:* Die vorliegende Arbeit wurde durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Mu-597/1-1) unterstützt.

Die beiden unabhängigen Bedingungen des zweiten Faktors «Einstellungsobjekt» waren die Urteilsobjekte «AMERICANS» und «NEGROES». Zur Beurteilung des Einstellungsobjektes hatte jede Versuchsperson auf siebenstufigen Ratingskalen anzugeben, in welchem Maße insgesamt 22 Eigenschaften auf die betreffende Personengruppe zuträfen. Dabei handelte es sich um eine Liste von Eigenschaften, die bereits KARLINS, COFFMAN & WALTERS (1969) von KATZ & BRALY (1933) übernommen und in Untersuchungen zur Stereotypisierung von farbigen Amerikanern angewendet hatten. SIGALL & PAGE verwendeten zur Messung von Einstellungen gegenüber weißen Amerikanern und Farbigen aber nicht nur diese 22 einfachen Ratings, sondern sie berechneten darüber hinaus ein Maß der globalen Bewertung des Einstellungsobjektes: Jedes der 22 Rating-Urteile, die eine Versuchsperson über ein Einstellungsobjekt abgab, wurde mit einem Maß der sozialen Erwünschtheit des betreffenden Adjektivs multipliziert, und es wurde der Durchschnitt dieser Produkte gebildet.

Neben einer großen Zahl signifikanter Haupteffekte, besonders für den Faktor «Einstellungsobjekt», fanden sich in sechs von 22 Fällen erwartungsgemäß signifikante Effekte der Wechselwirkung zwischen «Einstellungsobjekt» und «Meßmethode». Die signifikanten Wechselwirkungen besagen jeweils, daß Farbige unter BPL-Bedingungen im Unterschied zu Paper-Pencil-Bedingungen anders als weiße Amerikaner beurteilt wurden; dies betraf die Merkmale HONEST, SENSITIVE, PHYSICALLY DIRTY, STUPID, LAZY und INTELLIGENT. Betrachtet man nun den Bewertungs-Score, der summarisch den Grad der sozial positiven Bewertung der beiden Gruppen angibt, so ergab die entsprechende  $2 \times 2$ -Varianzanalyse eine signifikante Interaktion (vgl. Abb. 1).

Während der mittlere Grad der Bewertung beider Gruppen unter normalen Rating-Bedingungen, d.h. bei Einstellungsmessung unter Paper-Pencil-Bedingungen, fast identisch war (0.53 für AMERIKANER, 0.49 für NEGER), unterschieden sich die Bewertungen unter Bogus-Pipeline-Bedingungen beträchtlich (0.84 für AMERIKANER, –0.03 für NEGER).

Das berichtete Experiment wurde von Si-

FAVOURABILITY

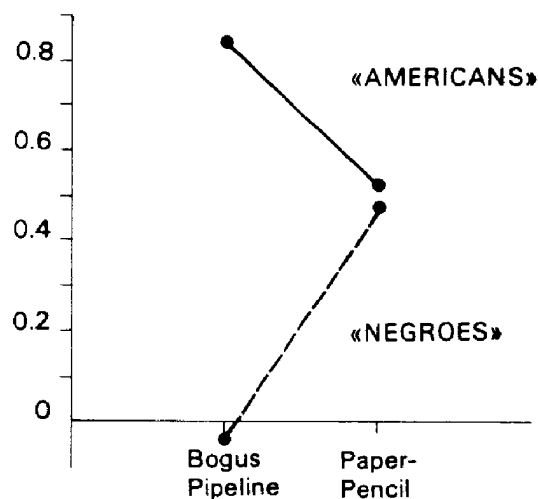


Abb. 1: Mittlere Bewertungen (favourability scores) von «AMERICANS» und «NEGROES». Im Experiment von SIGALL & PAGE (1971) (min./max. Ordinatenwerte –2.83/+2.83).

GALL & PAGE (1971) sowie von JONES & SIGALL (1971) als grundlegender Beleg für die Tauglichkeit des BPL-Paradigmas, und das heißt für die Reduktion des Social-Desirability-Effekts in der BPL-Situation angesehen. MUMMENDEY & BOLTEN (1981) sprechen in diesem Zusammenhang von einem «experimentum crucis» des BPL-Paradigmas. BRACKWEDE (1980) gibt eine Übersicht über 17 sozialpsychologische Experimente, von denen die große Mehrzahl die Gültigkeit des BPL-Paradigmas zu bestätigen scheinen.

Eine erste deutsche Replikation des Experiments von SIGALL & PAGE (1971) nahmen MUMMENDEY, SCHIEBEL, TROSKE, HESENER & BOLTEN (1979) vor, indem sie möglichst wortgetreue Übersetzungen der 22 Eigenschaften von KATZ & BRALY (1933) verwendeten und ihre Versuchspersonen – zur Hälfte unter BPL-, zur Hälfte unter Paper-Pencil-Bedingungen – die Einstellungsobjekte «SCHWARZAFRIKANER», «TÜRKISCHE GASTARBEITER», «HOLLÄNDER» und «DER DURCHSCHNITTLLICHE DEUTSCHE» beurteilen ließen. Die Auswahl der Einstellungsobjekte geschah unter folgenden Gesichtspunkten: Da zu vermuten war, daß studentische Versuchspersonen ein negativ getöntes Deutschenbild haben, wurde zusätzlich die Einstellung gegenüber «HOLLÄNDERN» gemessen; als deutsches Äquivalent für das Objekt «NEGROES»

wurden «TÜRKISCHE GASTARBEITER» als weiteres Objekt untersucht; zusätzlich wurde das Objekt «SCHWARZAFRIKANER» berücksichtigt, da es sich auf «NEGER» bezieht, die jedoch keine US-Bürger sind, so daß sie von einem vermutlich negativen US-Stereotyp verschont bleiben.

Neben einer Vielzahl signifikanter Haupteffekte, vor allem für den Faktor «Einstellungsobjekt», ergaben sich in diesem Experiment nur sehr selten signifikante Wechselwirkungen zwischen Einstellungsobjekt und Meßmethode. Während für die Vergleiche HOLLÄNDER/SCHWARZAFRIKANER, SCHWARZAFRIKANER/DEUTSCHE und SCHWARZAFRIKANER/TÜRKEN signifikante Wechselwirkungen vollständig fehlten, ergaben sich für den Vergleich HOLLÄNDER/TÜRKEN (der nach den Erwartungen der Autoren am ehesten einem Vergleich zwischen einer besonders positiv und besonders negativ bewerteten Gruppe entspricht) bei Annahme eines Signifikanzniveaus von zehn Prozent vier interpretierbare Unterschiede, nämlich für die Eigenschaften «FLEISSIG», «GROSSMÄULIG», «MUSIKALISCH» und «PROGRESSIV». Beim Vergleich von Deutschen und Türken ergab sich lediglich eine auf dem 5%-Niveau gesicherte Wechselwirkung für das Adjektiv «UNSAUBER»; während DEUTSCHE und TÜRKEN unter Papier-Bleistift-Bedingungen annähernd gleich beurteilt wurden, wurden unter Bogus-Pipeline-Bedingungen die Türken als wesentlich «unsauber» eingeschätzt.

Die nicht überzeugenden Resultate der deutschen Wiederholung des Experimentes von SIGALL & PAGE (1971) durch MUMMENDEY et al. (1979) werden von den Autoren bereits teilweise damit begründet, daß die aus dem Amerikanischen entnommenen Eigenschaftswörter nicht ohne weiteres für die Untersuchungen von Einstellungen gegenüber ethnischen Gruppen durch deutsche Vpn übertragbar seien. Ferner läßt sich gegen die Auswertungsprozedur von SIGALL & PAGE (1971), sofern sie sich auf eine Vielzahl einzelner  $2 \times 2$ -Varianzanalysen bezieht, Kritik äußern. Bei dieser Vorgehensweise, die von MUMMENDEY et al. (1979) wiederholt wurde, wächst mit der Anzahl einzelner Signifikanztests die Chance zufällig zustande gekommener signifikanter Resultate.

In der vorliegenden Untersuchung soll diesen beiden hauptsächlich Mängeln der früheren Experimente Rechnung getragen werden, indem einerseits in sorgfältiger Weise objektadäquate Eigenschaftswörter ausgewählt werden, und andererseits eine Auswertungsstrategie angewendet wird, mittels derer die Frage der Wechselwirkung zwischen Meßmethode und Einstellungsobjekt durch einen einzigen Test geprüft werden kann.

Neben den genannten methodischen Verbesserungen sollen in der vorliegenden Untersuchung auch einige weitere Modifikationen der Replikation des Experimentes von SIGALL & PAGE vorgenommen werden. So wies zum Beispiel OSTROM (1973) auf gewisse Unterschiede in der Darbietung der Einstellungsobjekte durch den Versuchsleiter unter der BPL- und der Paper-Pencil-Bedingung hin (vgl. auch QUIGLEY-FERNANDEZ & TEDESCHI, 1978). In der vorliegenden Untersuchung soll daher der Beurteilungsvorgang im Unterschied zu MUMMENDEY et al. (1979) so erfolgen, daß er unter beiden experimentellen Bedingungen möglichst ähnlich ist. Im Unterschied zu den älteren Arbeiten soll im vorliegenden Experiment jede Versuchsperson sämtliche Einstellungsobjekte beurteilen; dadurch entstehende sequentielle Effekte sollen durch den Versuchsplan ausgeglichen werden. Den Erfahrungen aus der früheren deutschen Untersuchung folgend sollen als Einstellungsobjekte lediglich TÜRKEN, DEUTSCHE und HOLLÄNDER verwendet werden.

Schließlich soll in dem folgenden Experiment der Nachweis erbracht werden, daß der Bogus-Pipeline-Effekt relativ unabhängig von der Art, speziell von der Größe und dem apparativen Aufwand der Versuchsanordnung ist. Dazu soll zusätzlich zu der bereits von MUMMENDEY et al. (1979) verwendeten, relativ aufwendigen Versuchsanordnung ein kleines, handliches Gerät entwickelt werden, das im wesentlichen die gleichen Funktionen erfüllt.

## Methoden

### Versuchsplan

Es ergibt sich ein  $3 \times 3$ -Versuchsplan mit den Faktoren «Meßmethode» und «Einstellungsobjekt». Der erste Fak-

	Vp	TÜRKEN	DEUTSCHE	HOLLÄNDER	
BPL Alt	1	1, 2, 3 ... k	1, 2, 3 ... k	1, 2, 3 ... k	Items
	2	⋮	⋮	⋮	
	3	⋮	⋮	⋮	
	⋮	⋮	⋮	⋮	
	n	⋮	⋮	⋮	
BPL Neu	1	1, 2, 3 ... k	1, 2, 3 ... k	1, 2, 3 ... k	Items
	2	⋮	⋮	⋮	
	3	⋮	⋮	⋮	
	⋮	⋮	⋮	⋮	
	n	⋮	⋮	⋮	
PP	1	1, 2, 3 ... k	1, 2, 3 ... k	1, 2, 3 ... k	Items
	2	⋮	⋮	⋮	
	3	⋮	⋮	⋮	
	⋮	⋮	⋮	⋮	
	n	⋮	⋮	⋮	

Abb. 2: Versuchsplan.

tor wird durch die Versuchsbedingungen «Bogus-Pipeline-Alt», «Bogus-Pipeline-Neu» und «Paper-Pencil» repräsentiert. Bei dem Faktor «Einstellungsobjekt» handelt es sich um einen Faktor mit Meßwiederholung; er wird durch die Objekte «TÜRKEN», «DEUTSCHE» und «HOLLÄNDER» repräsentiert (vgl. Abb. 2).

Wie aus Abbildung 2 ersichtlich wird, werden Einflüsse der genannten Faktoren sowohl auf die Beurteilung der Einstellungsobjekte mittels einzelner Adjektive (der Items) als auch auf einen aus den Items gebildeten Gesamtscore erwartet.

**Hypothesen**

Insgesamt werden drei Hypothesen formuliert; für den Nachweis des Bogus-Pipeline-Effekts ist jedoch allein Hypothese 3 entscheidend.

1. Es werden sich Unterschiede in der Beurteilung der ethnischen Gruppen (TÜRKEN, DEUTSCHE, HOLLÄNDER) ergeben.
2. Es werden sich Unterschiede zwischen den Versuchsbedingungen BPL-Alt, BPL-Neu und Paper-Pencil ergeben.
  - (2a) Zwischen den Versuchsbedingungen BPL-Alt und BPL-Neu werden sich keine interpretierbaren Unterschiede ergeben.
  - (2b) Es werden sich Unterschiede zwischen den beiden BPL-Bedingungen einerseits, und der PP-Bedingung andererseits ergeben.
3. Es wird eine signifikante Wechselwirkung zwischen den Faktoren «Einstellungsobjekt» und «Meßmethode» erwartet.

**Voruntersuchung: Auswahl der Adjektive**

Aufgrund des Studiums verschiedener Quellen fachlicher und populärer Literatur, mehreren Gesprächen mit Mitarbeitern des Ausländerreferates einer Wohltätigkeitsorganisation, ferner in Anlehnung an die Vorarbeiten von SIGALL & PAGE sowie aufgrund eigener Überlegungen und unter Berücksichtigung signifikanter Ergebnisse unserer früheren Untersuchung wurden insgesamt 44 Adjektive ausgewählt, mit denen sich vor allem «Türken» in der Bundesrepublik Deutschland im Sinne negativer Einstellungen bzw. Vorurteile beschreiben ließen.

Eine Liste mit diesen 44 Adjektiven wurde in mehreren Seminaren 56 Studienanfängern unterschiedlicher Fachbereiche zusammen mit Sieben-Punkten-Antwortskaleten und einer Instruktion vorgelegt, die darauf zielte, Tendenzen sozialer Erwünschtheit gegenüber den Urteilsobjekten («DEUTSCHER», «TÜRKE», «HOLLÄNDER») in systematisch variiert Reihenfolge) möglichst gegenzusteuern, indem die Probanden aufgefordert wurden, sich nicht in dem im akademischen Rahmen sozial erwünschten Sinne zu verhalten, sondern möglichst vorurteilshaltige Urteile abzugeben.

Für jedes Adjektiv wurde eine einfaktorielle Varianzanalyse für abhängige Stichproben mit dem Faktor «Ethnische Gruppe» und den Treatments DEUTSCHE/TÜRKEN/HOLLÄNDER gerechnet, um zu prüfen, welche Eigenschaftswörter zwischen den drei ethnischen Gruppen diskriminieren. Mit zwei Ausnahmen ergaben sich durchweg auf dem 5%-Niveau gesicherte Urteilsdifferenzen zwischen den drei Urteilsobjekten. Eine Inspektion der statistischen Kennwerte zeigte abermals, daß «Holländer» als Idealgruppe auftreten, «Deutsche» und «Türken» etwa gleich «schlecht» beurteilt werden, wobei jedoch «Türken» noch insgesamt etwas besser abschneiden als «Deutsche».

Sodann wurde getrennt für jede ethnische Gruppe eine Faktorenanalyse über die Items gerechnet, die zu einer recht ähnlichen Faktorenstruktur für die Urteilsobjekte «Türken» und «Holländer» führte, während für «Deutsche» eine abweichende Faktorenstruktur resultierte. Gleichartig war jedoch in allen drei Fällen zu beobachten, daß der jeweils erste Faktor eine große Zahl gemäß dem FÜRNRATT-Kriterium signifikanter Ladungen auf sich zieht, im Gegensatz zu den weiteren Faktoren. Da für die weitere Untersuchung vor allem Urteilsunterschiede zwischen «Türken» und «Holländern» zu erwarten waren und da sich die Faktorenstruktur der Adjektive für diese beiden ethnischen Gruppen als fast identisch in bezug auf den ersten extrahierten Faktor erwiesen hatte, wurden für das Experiment jene 11 Items ausgewählt, die in bezug auf «TÜRKEN» signifikant auf dem ersten Faktor luden.

In Tabelle 1 werden die elf endgültig ausgewählten Items zusammen mit den Kennwerten der Voruntersuchungen für die Urteilsobjekte «TÜRKEN», «DEUTSCHE» und «HOLLÄNDER» wiedergegeben.

Tabelle 1 enthält zugleich die Trennschärfe-Koeffizienten (Item-Test-Korrelationen jedes Items bzw. Adjektives), die für TÜRKEN sämtlich auf dem 1%-Niveau signifikant sind und zwischen .72 für VERTRAUENSWÜRDIG und .51 für GROSSMÄULIG variieren. Der alpha-Konsistenzkoeffizient beträgt .90 für TÜRKEN, .82 für DEUTSCHE und .78 für HOLLÄNDER. Außergewöhnliche Standardabweichungen bzw. Schwierigkeitsgrade sind bei den aus-

Tab. 1: Mittelwerte, Standardabweichungen und Item-Test-Korrelationen der 11 aufgrund der Voruntersuchung ausgewählten Adjektive.

(n = 56)	TÜRKEN			DEUTSCHE			HOLLÄNDER		
	$\bar{X}$	s	$r_{it}$	$\bar{X}$	s	$r_{it}$	$\bar{X}$	s	$r_{it}$
AUFDRINGLICH	3.78	1.78	.57	4.08	1.71	.30	3.02	1.81	.39
ZUVERLÄSSIG	4.56	1.37	.69	5.25	1.36	.20	5.32	0.94	.08
RECHTHABERISCH	4.12	1.41	.59	5.45	1.38	.65	3.28	1.38	.33
VERTRAUENSWÜRDIG	4.55	1.59	.72	4.19	1.62	.46	5.36	1.06	.15
TRIEBHAF	4.42	1.73	.65	3.71	1.65	.47	3.51	1.25	.51
UNGEBILDET	3.98	1.53	.61	2.89	1.33	.48	2.57	1.37	.56
HINTERHÄLTIG	2.77	1.44	.69	3.46	1.41	.63	2.49	1.27	.41
GROSSMÄULIG	3.91	1.55	.51	5.39	1.52	.69	3.51	1.68	.56
UNBEHERRSCHT	4.89	1.56	.70	4.00	1.81	.56	3.25	1.42	.68
SCHWATZHAFT	4.37	1.73	.63	4.37	1.68	.48	3.83	1.77	.50
UNSAUBER	3.48	1.71	.64	2.51	1.40	.41	2.47	1.34	.55

gewählten Adjektiven nirgends zu beobachten. Insgesamt erwiesen sich die elf Adjektive demnach als geeignet zur Bildung einer eindimensionalen Skala. (Der Adjektiv-Liste wird jeweils noch als «warming-up-Item» das Eigenschaftswort BESCHIEDEN vorangestellt.)

### Unabhängige Variablen

Versuchsarrordnungen:

#### BPL-Alt

In einem 2,55 × 3,40 m großen, mit Instrumenten vollgestopften Versuchsraum befindet sich ein metallener Zahnarztstuhl mit Kopfstützen für die Vp. Davor aufgebaut ist ein «Testpult München» mit diversen Einschüben, u. a. einem Token-Expenser, auf dessen gebogenem Trommelrand eine Skala von -3 bis +3 befestigt ist; rechts und links unterhalb dieser Skala befinden sich zwei Allrichtungstasten für die rechte und linke Hand der Vp. Oberhalb des Testpults, schräg in Richtung der Vp geneigt und von einer Klavierlampe beleuchtet, befindet sich eine Lafayette-Gedächtnistrommel. An der rechten Seite des Pultes ist eine Anschlußtafel für mehr als 100 elektrische Steckverbindungen befestigt. Von hier verläuft zu den anderen Geräten der Versuchsarrordnung ebenso wie zwischen diesen eine Vielzahl von Kabelverbindungen, so daß alle Instrumente ausreichend miteinander verkabelt erscheinen. Rechts neben dem Testpult ist die Rückseite eines Wiener Determinationsgerätes, das auf dem Gehäuse des externen Speicherkerns eines Hewlett-Packard 9100B-Calculators steht, sichtbar. Noch weiter rechts, aber noch im Blickfeld der Vp befindet sich der Monitor der HP 9100B (Bildschirmgröße 25 × 33 cm), der (nicht erkennbar) vom Nebenraum aus ansteuerbar ist, darüber ein Event-Recorder mit der Aufschrift «Elektromyograph». Dessen Aufzeichnungen auf drei Kanälen sind deutlich erkennbar, nicht aber die Tatsache, daß es sich um eine Papierschleife handelt. Auf dem Ergebnisschreiber befindet sich eine weitere Anschlußtafel der geschilderten Art. Von

hier laufen Kabelleitungen u. a. zu einigen noch weiter rechts, d. h. neben bzw. hinter der Vp aufgebauten Geräten: einem Schuhfried-Digitaldrucker und einem elektronischen Bettendorf-Chronometer, beide in imponierenden Metallgehäusen, einem elektronischen Steuergerät für ein Projektionstachistoskop und einem kleinen Pult mit Signalknöpfen. Ein Kassettenrecorder zum Abspielen der Instruktion vervollständigt die Apparatesammlung.

#### BPL-Neu

In einem Raum gleicher Größe wie oben befindet sich auf einem Tisch ein etwa büroschreibmaschinengroßes Gerät (ein Siemens PC 100-Mikrocomputer). Das Gerät besitzt eine Tastatur und darüber eine Leuchtschriftenanzeige. An den beiden Seiten des Gerätes ist je ein beweglicher Metallhebel angebracht, von dem zwei Kabel über einen Stecker zum Gerät führen. Ebenfalls durch Kabel mit dem Gerät verbunden sind zwei Armmanschetten, die innen zwei Metallplättchen aufweisen. Ein Drucker mit einem abdeckbaren Papiervorschub ist in das Gerät eingebaut.

#### Paper-Pencil

In einem Raum gleicher Größe befinden sich lediglich wenige Büromöbel.

### Versuchsablauf

#### Versuchspersonen

Untersucht wurden 108 Versuchspersonen, überwiegend Lehramts- und Pädagogik-Studenten, und zwar überwiegend Studienanfänger (60 männlich, 48 weiblich), die nach Zufall auf die verschiedenen Bedingungen des Versuchsplanes aufgeteilt wurden.

#### BPL-Bedingungen

Die Versuchsperson füllt in Anwesenheit eines Helfers des Versuchsleiters einen Kurzfragebogen mit vier Items aus; sie beurteilt das Zutreffen der folgenden Statements auf einer siebenstufigen Skala:

- Ich bin fußballbegeistert;
- Ich höre gern Rockmusik;
- Ich sehe gern Krimis zur Entspannung;
- Ich bin für ein Tempolimit auf Autobahnen.

Anschließend nimmt die Versuchsperson vor dem BPL-Gerät Platz, und es wird ihr per Tonband die Instruktion gegeben:

*BPL-Alt:* «Im folgenden Teil des Versuches geht es um eine objektive physiologische Gefühlsmessung. Vor Ihnen befindet sich ein sogenannter adaptierter Elektromyograph, ein an der Yale-Universität in den USA entwickeltes Gerät zur genauen Registrierung unbewußter kleinster Muskelspannungen. Dazu muß Ihnen der Versuchsleiter an den beiden Unterarmen zwei Metallplättchen anbringen. Sie brauchen keine Angst zu haben, hier fließt keinerlei Strom, sondern es werden nur die in den Muskelspindeln stets vorhandenen Mikroschwingungen abgeleitet und aufgezeichnet.» (Der VI reinigt zwei Hautstellen an den Innenseiten der Unterarme der Vp mit Wattebausch und Alkohol und legt zwei an Uhrenarmbändern befestigte Elektroden an.)

«Die folgende Registrierung geht so vor sich: Legen Sie bitte den Kopf an die Kopfstützen, umfassen sie mit beiden Fäusten fest die beiden Metallhebel. Dabei sollen die Daumen am Schaft anliegen, und es soll möglichst viel Handfläche an den Hebeln anliegen. Betrachten Sie bitte dort die Skala mit den Stärkegraden von  $-3$  bis  $-1$  auf der linken Seite und von  $+3$  bis  $+1$  auf der rechten Seite. Null bedeutet die Mitte zwischen beiden Seiten. Bei dieser Skala handelt es sich um eine Meinungs- bzw. Gefühlsskala. Wenn gleich oben in dem Ausschnitt ein Satz oder ein Wort, d. h. ein Meinungsgegenstand erscheint, dann sollen Sie diesen Meinungsgegenstand daraufhin beurteilen, wie stark sie ihn ablehnen (linke Seite) bzw. wie stark sie ihm zustimmen (auf der rechten Seite). Der adaptierte Elektromyograph mißt während des Beurteilungsvorganges die ersten unbewußten und unmerklichen Muskelbewegungen im Unterarm in die eine oder andere Richtung. Diese Muskelbewegungen sind unabhängig von den groben Muskelbewegungen. Vermeiden Sie aber trotzdem grobe Bewegungen und halten Sie die Metallhebel ruhig. Sie brauchen sich also nur, nachdem Sie das Testwort bzw. den Testsatz gelesen haben, auf denjenigen Punkt der Skala zu konzentrieren, der Ihrer tatsächlichen Meinung bzw. Ihrem ersten unmittelbaren Gefühl am besten entspricht. Um das Gerät jetzt einregulieren zu können, weil jede Person einen unterschiedlichen Grundwert aufweist, sollen Sie zunächst mit den Meinungsgegenständen beginnen, die Sie vorhin schon einmal beurteilt haben. Die Meinungsgegenstände werden in einem festgelegten Abstand in dem Fensterchen dargeboten. Sobald Sie die Wörter wahrgenommen haben, konzentrieren Sie sich auf Ihre Antwortskala. Sie brauchen nur den richtigen Antwortskalpunkt anzupeilen und die Hände fest an den Griffen liegenzulassen. Das EMG, also das Elektromyogramm, pegelt sich dann ein und wird rechts auf der Mattscheibe, also über dem EMG-output, wiedergegeben. Ich wiederhole noch einmal: Sie brauchen die Griffe nicht zu bewegen, da nur die unwillkürlichen Muskelbewegungen registriert werden. Konzentrieren Sie sich auf den Ihrer tatsächlichen Meinung nach zutreffenden Antwortskalpunkt. Nach dem Ausschalten des Tonbandes erfolgt die erste Darbietung.»

(Der VI setzt die gesamte Apparatur in Gang, woraufhin in der Gedächtnistrommel nacheinander im Abstand von 16 Sekunden die vier oben bereits aufgeführten Items erscheinen. Beim Erscheinen eines jeden Items gibt der VI seinem Gehilfen, der sich die Antworten der Vp auf diese Items zuvor gemerkt hat, ein von der Vp nicht bemerktes Zeichen, woraufhin der Gehilfe vom Nebenraum aus die Antwort der Vp mit geringer Abweichung (z. B. statt  $+3$  jetzt 2.870) auf den Bildschirm überträgt. Der VI liest die Angaben auf dem Monitor hörbar ab und notiert den Wert in einer Liste. Nach dem letzten Wert schaltet er die Anlage wieder ab. Gemeinsam mit der Vp vergleicht er die «Meßwerte» mit den von ihr zuvor angekreuzten Werten. Er zeigt sich mit der Übereinstimmung zufrieden und stellt fest, daß das Gerät keiner weiteren Einregulierung bedürfte.)

«Im nächsten Teil des Versuches möchten wir untersuchen, wie genau jemand seine eigenen gefühlsmäßigen Einschätzungen kennt bzw. wie genau er sie vorhersagen kann. Ihnen werden jetzt wieder einige Gegenstände zur Beurteilung dargeboten. Dabei wird Ihnen aber der Blick auf den EMG-output verbaut, damit Sie die Computerdaten nicht sehen können. Stattdessen werden Sie jedes Mal um Ihre Einschätzung bzw. Vorhersage des Computerergebnisses gebeten. Sie brauchen also, wenn das betreffende Wort erscheint, nur Ihr nach Ihrer Meinung zutreffendes persönliches Urteil anzugeben, und hinterher werden diese Werte mit den output-Werten verglichen, und es wird ein Übereinstimmungsmaß berechnet. Zur Vereinfachung sollen Sie stets denselben Urteilsgegenstand beurteilen. Nehmen wir z. B. einmal an, dieser Urteilsgegenstand sei «Neger». Nacheinander werden Ihnen von dem Versuchsleiter Eigenschaftswörter dargeboten, die Sie daraufhin beurteilen sollen, wie sehr sie auf Neger zutreffen oder nicht zutreffen.  $-3$  bedeutet also, daß das exponierte Eigenschaftswort überhaupt nicht zutrifft,  $+3$  bedeutet, daß das jeweilige Eigenschaftswort vollständig zutrifft. Bemühen Sie sich bitte, die gesamte Urteils-skala auszunutzen und scheuen Sie sich nicht vor eindeutigen und ganz ursprünglichen Urteilen. Alle Daten werden anonym und lediglich zu Forschungszwecken verwendet. Nach dem Ausschalten des Tonbandes gibt Ihnen der Versuchsleiter an, welches Urteilsobjekt Sie mittels der dargebotenen Eigenschaftswörter beurteilen sollen. Dann werden Ihnen in festgelegter Reihenfolge die Eigenschaftswörter dargeboten, und Sie geben bitte jeweils anschließend Ihre Schätzung des Computerwertes an.»

(Der VI benennt das erste Urteilsobjekt (TÜRKEN, DEUTSCHE, HOLLÄNDER, Reihenfolge variiert) und stellt zusätzlich ein Pappschild mit der Bezeichnung dieses Urteilsobjektes auf das Testpult. Er bittet die Vp, die Metallhebel wieder anzufassen und setzt die Apparatur in Betrieb. Jeweils im Abstand von etwa 6 sec., orientiert am Schleifgeräusch der Gedächtnistrommel, legt der VI nacheinander 12 Karten mit den Eigenschaftswörtern (BESCHIEDEN, AUFDRINGLICH ... UNSAUBER) auf eine am Testpult angebrachte Halterung. Die Vp gibt laut ihren Schätzwert an, der vom VI notiert wird. Nachdem alle Eigenschaften zu einem Urteilsobjekt bewertet sind, stoppt der VI die Geräte und erklärt der Vp, welches Urteilsobjekt als nächstes mit den gleichen Adjektiven zu beurteilen ist und stellt das entsprechende Schild auf das Testpult. Während des «Meßvorganges» produziert das

Testpult in Verbindung mit der Gedächtnisstrommel ein Schleifgeräusch sowie ein rhythmisches Klicken. Der Drucker druckt hörbar (sinnlose) Werte aus und der «Elektromyograph» produziert den o.g. Output.

Nach dem letzten Einstellungsobjekt schaltet der VI die Geräte ab, sagt der Vp, damit wäre dieser Teil des Versuchs beendet und entfernt die «Elektroden».)

Die Vp wird nun wieder in einen Büroraum geführt, und es werden ihr nochmals die 12 Adjektive in einer Liste vorgelegt. Sie wird gebeten, die Adjektive auf einer Sieben-Punkte-Skala dahingehend zu beurteilen, für wie wünschenswert sie sie als Persönlichkeitsmerkmale hält. Schließlich wird sie noch gebeten, einige Fragen schriftlich zu beantworten (postexperimentelle Befragung).

*BPL-Neu:* Die Vp erhält zunächst die identische Tonbandinstruktion wie in der BPL-Alt-Bedingung. Ebenfalls werden ihr an der entsprechenden Passage die Armmanschetten mit den «Elektroden» angelegt. Nachdem der Vp erklärt wird, daß sie «zur Einregulierung des Gerätes» nochmals die schon einmal in dem Kurzfragebogen beurteilten vier Statements beantworten soll, weicht die Instruktion, bedingt durch die unterschiedliche Apparatur, geringfügig von der BPL-Alt-Bedingung ab:

«... Die Meinungsgegenstände werden in einem festgelegten Takt vom Versuchsleiter über der Skala placiert. Sobald Sie die Wörter wahrgenommen haben, konzentrieren Sie sich auf Ihre Antwortskala. Sie brauchen nur den richtigen Antwortskalenpunkt anzupeilen und die Hände fest an den Griffen liegenzulassen. Das EMG, also das Elektromyogramm, pegelt sich dann ein und wird auf der Computeranzeige wiedergegeben sowie auf dem Papierstreifen ausgedruckt. Ich wiederhole noch einmal: Sie brauchen die Griffe nicht zu bewegen, da nur die unwillkürlichen Muskelbewegungen registriert werden. Konzentrieren Sie sich auf den Ihrer tatsächlichen Meinung nach zutreffenden Antwortskalenpunkt. Nach dem Ausschalten des Tonbandes erfolgt die erste Darbietung.»

(Während des Abspielens der Instruktion schaut der VI unbemerkt auf den zu Beginn von der Vp ausgefüllten Kurzfragebogen und prägt sich die von der Vp angekreuzten Werte ein. Nach dem Abspielen der Instruktion nimmt der VI, während die Vp vor dem Gerät sitzen bleibt, verschiedene Eingaben vor. Er sagt der Vp, zunächst müsse er noch die Programm-Nr. eingeben. Er gibt eine 4stellige Zahl ein, nämlich die von der Vp vorher angekreuzten Werte, umkodiert in Ziffern von 1 (= -3) bis 7 (= +3) und anschließend die Vp.-Nr. und das Datum. Nachdem auf der Leuchtschriftanzeige das Signal «Startbereit» auftaucht, bittet der VI die Vp, die Metallhebel anzufassen und sich auf die auf dem Gerät angebrachte Antwortskala zu konzentrieren.

Zunächst werden vom VI die vier Items des Kurzfragebogens im festgelegten Takt in die dafür vorgesehene Halterung gelegt. Nachdem die Vp das jeweilige Item gelesen hat, erscheint auf dem Display «Achtung: Messung» und anschließend der jeweilige «Meßwert», der zusätzlich ausgedruckt wird (die vorher eingegebenen Werte erscheinen mit der Zufallsabweichung von bis zu +/-0,4). Wie in der Bedingung 1 werden die ausgedruckten Meßwerte mit der Vp gemeinsam verglichen und anschließend der 2. Teil der Instruktionen abgespielt.)

«Im nächsten Teil des Versuchs möchten wir untersu-

chen, wie genau jemand seine eigenen gefühlsmäßigen Einschätzungen kennt bzw. wie genau er sie vorhersagen kann. Ihnen werden jetzt wieder einige Gegenstände zur Beurteilung dargeboten, wobei Ihnen jedoch der Blick auf den Papierstreifen verbaut wird, damit Sie die Computerdaten nicht sehen können.»

(Der VI deckt den Drucker ab.) «Zusätzlich ist das Gerät so programmiert, daß auch die Computeranzeige dunkel bleibt. Stattdessen werden Sie jedesmal um Ihre Schätzung bzw. Vorhersage des Computerergebnisses gebeten. Sie brauchen also, wenn das betreffende Wort erscheint, nur Ihr nach Ihrer Meinung zutreffendes persönliches Urteil abzugeben, und hinterher werden diese Werte mit den Output-Werten verglichen, und es wird ein Übereinstimmungsmaß berechnet. Zur Vereinfachung sollen Sie stets denselben Urteilsgegenstand beurteilen. Nehmen Sie zum Beispiel an, dieser Gegenstand sei «Neger». Nacheinander werden Ihnen vom Versuchsleiter Eigenschaftswörter dargeboten, die Sie daraufhin beurteilen sollen, wie sehr sie auf «Neger» zutreffen oder nicht zutreffen. -3 bedeutet also, daß das exponierte Eigenschaftswort überhaupt nicht zutrifft, +3 bedeutet, daß das jeweilige Eigenschaftswort vollständig zutrifft. Bemühen Sie sich bitte, die gesamte Urteilsskala auszunutzen und scheuen Sie sich nicht vor eindeutigen und ganz ursprünglichen Urteilen. Alle Daten werden anonym und lediglich für Forschungszwecke verwendet. Nach dem Ausschalten des Tonbandes gibt Ihnen der Versuchsleiter an, welches Urteilsobjekt Sie mittels der dargebotenen Eigenschaftswörter beurteilen sollen. Dann werden Ihnen in festgelegter Reihenfolge die Eigenschaftswörter dargeboten und Sie geben bitte anschließend Ihre Schätzung des Computerwertes an.»

(Die Präsentation der Urteilsobjekte sowie der Eigenschaftswörter entspricht der BPL-Alt-Bedingung; der (abgedeckte) Drucker produziert alle 6 Sekunden ein Geräusch - scheinbar den Ausdruck des jeweiligen Meßwertes - durch das der Takt für die Vorlage der Eigenschaftswörter vorgegeben wird.)

Die anschließende postexperimentelle Befragung entspricht der BPL-Alt-Bedingung.

*PP-Bedingung:* Die Vpn werden vom Helfer in den Versuchsraum geführt, wo sie am Schreibtisch schräg gegenüber vom VI Platz nehmen.

Der VI erklärt der Vp in Anlehnung an die Instruktion der BPL-Bedingungen zunächst die Skala. Er bittet die Vp, möglichst spontan zu antworten und sich nicht vor eindeutigen Urteilen zu scheuen und weist auf die Anonymität der Vp sowie den wissenschaftlichen Zweck der Untersuchung hin. Dann benennt er das zuerst zu beurteilende Objekt (TÜRKEN, DEUTSCHE, HOLLÄNDER) und stellt das entsprechende Schild vor der Vp auf. Die Eigenschaftswörter werden der Vp ebenfalls auf Karten vorgelegt und die Antworten notiert.

Die postexperimentale Befragung erfolgt wie in den BPL-Bedingungen.

## Ergebnisse

Zur Vermeidung einer Vielzahl von Varianzanalysen (für die verschiedenen Eigenschafts-



urteile) und angesichts der Tatsache, daß die elf Eigenschaften faktoriell eine Skala bilden, wurde ein *Gesamtscore* pro Versuchsperson und ethnische Gruppe berechnet. Dieser Score wurde nach Umpolung zweier Items, die negativ auf dem Faktor laden, durch Summenbildung aller Ratings bestimmt und wird im folgenden als *einfacher Score* bezeichnet. Er gibt somit zusammengefaßt das Urteil einer Versuchsperson über eine ethnische Gruppe wieder.

Es wurde eine  $3 \times 3$ -Varianzanalyse mit einem abhängigen Faktor, und zwar für die Faktoren «Methode» (unabhängig) und «Ethnische Gruppe» (abhängig) für diesen einfachen Score gerechnet. Das Ergebnis der Varianzanalyse ist in Tabelle 2 wiedergegeben.

Demnach wird für den *Gesamtscore* der Eigenschaftsbeurteilungen die Wechselwirkung «Methode  $\times$  Ethnische Gruppe» signifikant, neben dem signifikanten Haupteffekt für den Faktor «Ethnische Gruppe». Das Resultat wird deskriptiv in Abbildung 3 verdeutlicht; ein ho-

Tab. 2: Ergebnisse der Varianzanalyse «Methode  $\times$  Ethnische Gruppe» für den einfachen Gesamtscore der 11 Eigenschaftsbeurteilungen.

Quelle	df	MS	F
Methode (A)	2	184.18	1.57
Ethnische Gruppe (B)	2	2952.53	58.18***
A $\times$ B	4	197.83	3.90**

\*\*  $p < .01$      \*\*\*  $p < .001$

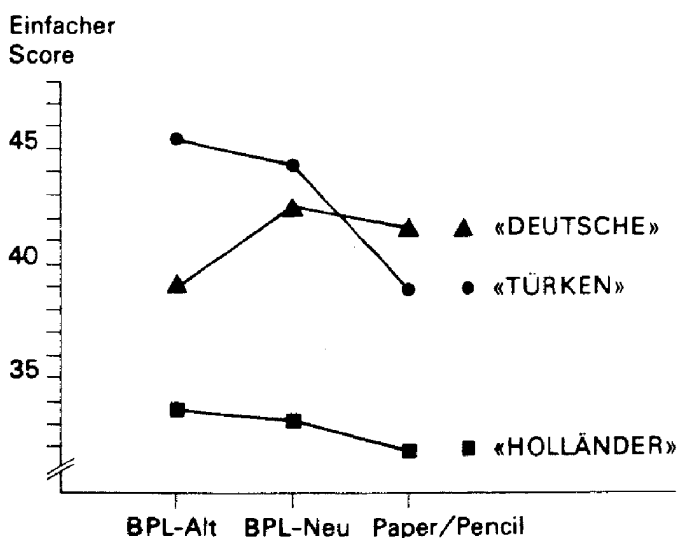


Abb. 3: Darstellung der Mittelwerte des einfachen Gesamtscores (minimale/maximale Ordinatenwerte 11 bzw. 77).

her Gesamtscore entspricht dabei hohen Beurteilungswerten hinsichtlich «AUFDRINGLICH», «RECHTHABERISCH», «TRIEBHAF» usw.

Gemäß dem NEWMAN-KEULS-Test werden sämtliche ethnischen Gruppen im Durchschnitt von den Versuchspersonen auf dem Fünf-Prozent-Niveau unterschiedlich beurteilt. Betrachtet man die entsprechenden Posttests für die signifikante Interaktion «Methode  $\times$  Ethnische Gruppe», so werden für jede der drei Methoden-Bedingungen einzeln Unterschiede zwischen den ethnischen Gruppen auf dem 0.001-Niveau signifikant. Stellt man die entsprechenden Vergleiche getrennt für die drei ethnischen Gruppen an, so ergibt sich lediglich bei dem Urteilsobjekt «TÜRKEN» eine auf dem 0.01-Niveau gesicherte Differenz zwischen den drei Methodenarten. Für das Objekt «TÜRKEN» gilt, daß der Unterschied zwischen BPL-Alt und BPL-Neu nicht signifikant ist ( $t = 0,59$ ), daß dagegen die Differenz zwischen BPL-Neu und der Paper-Pencil-Bedingung signifikant ist ( $t = 2,52$ ).

Auf der Grundlage des einfachen Gesamtscores der Eigenschaftsbeurteilungen wird damit die den *Hypothesen 1*, teilweise *2b* sowie *3* entsprechende Null-Hypothese zurückgewiesen: Es zeigen sich Unterschiede in der Beurteilung der ethnischen Gruppen, und die Wechselwirkung zwischen den Faktoren «Urteilsobjekt» und «Meßmethode» wird signifikant. Unterschiede zwischen den beiden BPL-Bedingungen einerseits und der Paper-Pencil-Bedingung andererseits ergeben sich lediglich für das Urteilsobjekt «TÜRKEN».

Ein weiterer *Gesamtscore* wurde für die Urteile jeder Versuchsperson über jede ethnische Gruppe unter Berücksichtigung der persönlichen sozialen Erwünschtheit jeder Eigenschaft ermittelt; er wird im folgenden als *Bewertungsscore* bezeichnet. Dazu wurde jedes Rating mit dem Rating der Erwünschtheit (aufgrund der abschließenden Bewertung aller Eigenschaften durch die Versuchsperson) jeder Eigenschaft multipliziert und die Summe aller Produkte pro Person gebildet. Wiederum wurde eine  $3 \times 3$ -Varianzanalyse mit einem abhängigen Faktor gerechnet. Das Resultat gibt Tabelle 3 wieder.

Auch für den *Bewertungsscore* wird die ent-



Tab. 3: Ergebnisse der Varianzanalyse «Methode  $\times$  Ethnische Gruppe» für den Bewertungsscore der 11 Eigenschaftsbeurteilungen.

Quelle	df	MS	F
Methode (A)	2	883.80	1.33
Ethnische Gruppe (B)	2	12350.87	41.98***
A $\times$ B	4	1122.34	3.81**

\*\* p < .01    \*\*\* p < .001

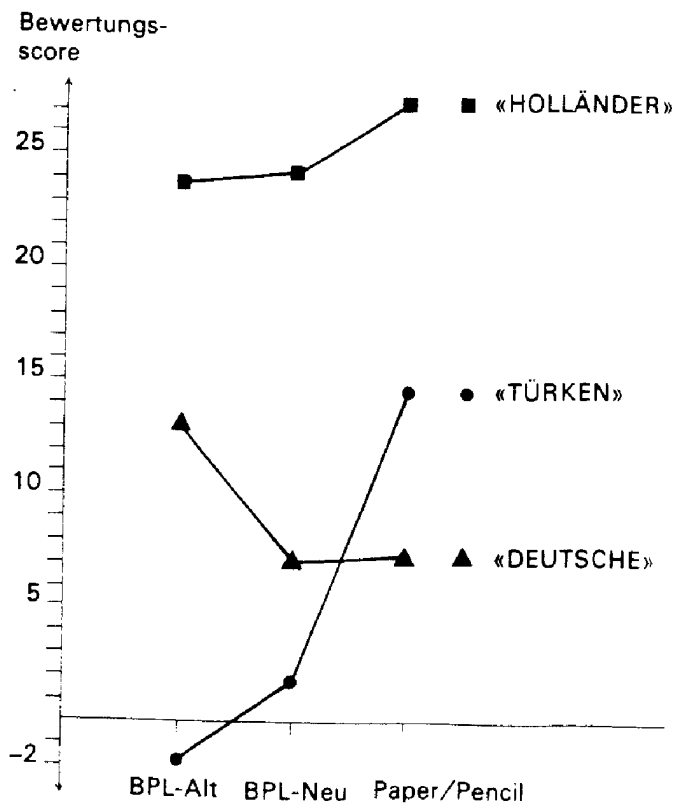


Abb. 4: Darstellung der Mittelwerte des Bewertungsscores (minimale/maximale Ordinatenwerte  $-99$  bzw.  $+99$ ). Hohe Scores bedeuten positive Bewertung des Urteilsobjektes.

scheidende Interaktion «Methode  $\times$  Ethnische Gruppe» neben dem Faktor «Ethnische Gruppe» signifikant. Abbildung 4 gibt die Ergebnisse graphisch wieder. Hohe Scores bedeuten positive Bewertung des Objektes.

Aufgrund des NEWMAN-KEULS-Tests unterscheiden sich «HOLLÄNDER» sowohl von «TÜRKEN» als auch von «DEUTSCHEN» im Mittel auf dem 0.05-Niveau signifikant; zwischen «TÜRKEN» und «DEUTSCHEN» konnte kein signifikanter Unterschied nachgewiesen werden. Die Posttests für die Analyse der Interaktion «Methode  $\times$  Ethnische Gruppe» führen zu denselben Resultaten wie oben

für den einfachen Score berichtet: Hochsignifikant sind Unterschiede zwischen ethnischen Gruppen pro Methoden-Bedingung, dagegen unterschieden sich die Methodenarten nur beim Einstellungsobjekt «TÜRKEN». Für «TÜRKEN» ergibt sich auch hier, daß kein signifikanter Unterschied zwischen BPL-Alt und BPL-Neu vorliegt ( $t = 0,88$ ); dagegen ist der Unterschied zwischen BPL-Neu und Paper-Pencil signifikant ( $t = 2,55$ ).

Damit wird auch für den Bewertungsscore, der den Grad der positiven Beurteilung einer ethnischen Gruppe im Sinne der persönlichen sozialen Erwünschtheit direkt wiedergibt, die den Hypothesen 1, teilweise 2b sowie 3 entsprechende Nullhypothese zurückgewiesen.

Zusätzlich wurden für jede der elf zu beurteilenden Eigenschaften getrennt  $3 \times 3$ -Varianzanalysen mit dem Wiederholungsfaktor «Meßmethode» berechnet; die Ergebnisse sind in Tabelle 4 zusammengefaßt. Demnach kommt es in drei von elf Fällen zu signifikanten Interaktionen zwischen «Ethnische Gruppe» und «Meßmethode»; in zwei weiteren Fällen wird für die Wechselwirkungen das 10%-Niveau unterschritten. Signifikante Haupteffekte für den Faktor «Meßmethode» zeigen sich in keinem einzigen Fall, dagegen sind mit einer Ausnahme sämtliche Unterschiede zwischen den drei ethnischen Gruppen signifikant. Damit zeigt sich auf der Ebene einzelner Items, daß der Bogus-Pipeline-Effekt für einzelne Items (Adjektive) in durchaus unterschiedlichem Maße zutrifft. Eindeutig am stärksten ist er für die Eigenschaft AUFDRINGLICH sowie für VERTRAUENSWÜRDIG; so gut wie kein Effekt zeigt sich dagegen für SCHWATZHAFT.

Schließlich wurde eine *multivariate Varianzanalyse* (MANOVA) mit den Faktoren «Ethnische Gruppe» und «Meßmethode» für jeweils unabhängige Stichproben gerechnet. Bei dieser Auswertungsmethode wird für jede Versuchsperson nur die Beurteilung der jeweils ersten zu beurteilenden ethnischen Gruppe verwendet. Als Ergebnisse der  $3 \times 3$ -MANOVA werden die auf der Basis von WILKS Lambda geschätzten F-Werte für den Faktor «Ethnische Gruppe» auf dem 0.001-Niveau und für die Interaktion «Meßmethode  $\times$  Ethnische Gruppe» auf dem 0.002-Niveau signifikant.

Tab. 4: Mittelwerte der Beurteilungen unterschiedlicher ethnischer Gruppen mittels 11 Eigenschaften (pro Versuchsbedingung «Methode» jeweils 36 Vpn).

Eigenschaft	TÜRKEN			DEUTSCHE			HOLLÄNDER			F-Werte		
	BPL alt	BPL neu	PP	BPL alt	BPL neu	PP	BPL alt	BPL neu	PP	(A)	(B)	A × B
AUFDRINGLICH	4.83	4.58	3.50	3.81	3.86	4.50	2.89	2.08	3.50	0.00	20.47***	7.00***
ZUVERLÄSSIG	4.25	4.39	4.36	5.53	4.89	5.14	4.92	4.78	5.31	1.15	16.30***	1.75
RECHTHABERISCH	4.17	4.00	3.50	5.31	5.53	5.61	3.72	3.06	3.28	0.75	87.35***	2.39*
VERTRAUENSWÜRDIG	3.78	4.06	4.83	5.19	4.47	4.78	5.44	5.42	5.33	1.58	25.18***	4.26**
TRIEBHAF	4.19	4.58	3.89	3.31	3.67	3.83	3.03	3.14	3.03	0.62	25.81***	1.69
UNGEBILDET	3.47	3.81	3.69	3.17	2.72	2.53	2.58	2.81	2.08	1.28	21.59***	1.65
HINTERHÄLTIG	3.44	3.31	2.64	3.06	3.25	3.14	2.39	2.39	2.28	0.96	16.48***	1.55
GROSSMÄULIG	4.42	3.81	3.42	4.81	4.97	5.11	3.64	3.28	3.64	0.64	26.26***	1.99 <sup>T</sup>
UNBEHERRSCHT	4.69	4.78	4.19	3.86	4.50	4.06	3.25	2.83	3.03	0.64	42.83***	2.22 <sup>T</sup>
SCHWATZHAFT	4.72	4.36	3.86	4.36	4.31	4.22	4.22	4.39	3.92	2.08	0.28	0.73
UNSAUBER	3.50	3.58	3.39	2.08	2.86	2.50	2.44	2.50	1.89	1.49	28.36***	1.67

(A) = Faktor «Methode» (BPL-Alt/BPL-Neu/Paper-Pencil) (unabhängiger Faktor)

(B) = Faktor «Ethnische Gruppe» (Türken/Deutsche/Holländer) (abhängiger Faktor)

\* p < .05    \*\* p < .01    \*\*\* p < .001    <sup>T</sup> p < .10

In entsprechender Weise wurden zum Vergleich je zweier Einstellungsobjekte bzw. ethnischer Gruppen 2 × 3-MANOVAs gerechnet. Beim Vergleich von «TÜRKEN»/«HOLLÄNDER» ergeben sich wiederum signifikante Effekte des Faktors «Ethnische Gruppe» und der Wechselwirkung zwischen «Meßmethode» und «Ethnische Gruppe». Das gleiche Resultat zeigt sich für den Vergleich «TÜRKEN»/«DEUTSCHE». Dagegen ergibt sich beim Vergleich «DEUTSCHE»/«HOLLÄNDER» je ein signifikanter Haupteffekt der Faktoren «Meßmethode» und «Ethnische Gruppe», nicht dagegen ein Effekt der statistischen Wechselwirkung.

Bei Einbeziehung aller elf Eigenschaftsurteile in eine multivariate Varianzanalyse mit den Faktoren «Ethnische Gruppe» und «Meßmethode» wird die der Hypothese 1 entsprechende Nullhypothese durchweg zurückgewiesen. Die der Hypothese 2 entsprechende Nullhypothese wird nur beim Vergleich von «DEUTSCHE» und «HOLLÄNDER» zurückgewiesen. Die der Hypothese 3 entsprechende Nullhypothese wird in allen Fällen mit Ausnahme des Vergleichs «DEUTSCHE»/«HOLLÄNDER» zurückgewiesen.

### Postexperimentelle Befragung

Die postexperimentelle Befragung der Versuchspersonen hatte zum Ziel, Aufschluß über die mögliche Transparenz des Experimentes, die Glaubwürdigkeit der Bogus-Pipeline-«Messung», Motiviertheit und Aufmerksamkeit der Probanden zu erhalten. Es zeigte sich, daß die «Zuverlässigkeit» und die «Anschaffungskosten» beider BPL-Versuchsanordnungen von den Vpn als relativ hoch eingeschätzt wurden. Als vermeintliche Zielsetzung der experimentellen Untersuchungen sehen die Versuchspersonen unter der Paper-Pencil-Bedingung häufiger die Untersuchung von «Einstellungen gegenüber ethnischen Gruppen», dagegen geben die Versuchspersonen unter den BPL-Bedingungen häufiger Antworten in Richtung auf «Selbsterkenntnis». BPL-Vpn nehmen eher als Paper-Pencil-Vpn an, daß die Untersuchung dem «wissenschaftlichen Fortschritt» diene. Unter den BPL-Vpn schätzen solche, die den wissenschaftlichen Nutzen der experimentellen Untersuchung als hoch einschätzen, die Kosten der BPL-Apparatur als vergleichsweise hoch ein. Sämtliche anderen erfragten Merkmale (z. B. Teilnehmermotivation, Kritik an der Untersuchung und experimentelle Vorerfahrung) verteilen sich gleich-

mäßig über die experimentellen Bedingungen. Als Bemerkungen zu der Untersuchung (geäußert von 26 der 108 Vpn) wurde zumeist Kritik an der Art der experimentellen Situation geäußert (Unpersönlichkeit, unfreundliche Räume usw.).

Nachträglich vorgenommene Kovarianzanalysen über den einfachen Gesamtscore sowie den Bewertungsscore, bei denen als Kovariaten jeweils die Einschätzung des «wissenschaftlichen Nutzens» der Untersuchung, die «Zuverlässigkeit» der BPL-Apparatur, die «Kosten» dieser Apparatur und die «Vorerfahrung als Versuchsperson» eingesetzt wurden, erbrachten keine Veränderungen der oben berichteten Ergebnisse.

## Diskussion

Die Ergebnisse des vorliegenden Experiments haben vollständige Unterstützung für die Hypothese einer Wechselwirkung zwischen der angewendeten Meßmethode (Bogus-Pipeline-Messung oder Papier-Bleistift-Messung) und den beurteilten Einstellungsobjekten (verschiedene ethnische Gruppen) gebracht. Demnach werden ethnische Gruppen, wie «TÜRKEN», unterschiedlich beurteilt je nachdem, ob die beurteilende Person ein einfaches Rating-Urteil abgibt oder ob sie annehmen muß, die physiologische Begleiterscheinung ihrer sozialen Einstellung werde durch ein entsprechendes Gerät exakt registriert. Berücksichtigt man den Grad der sozialen Erwünschtheit der zur Beurteilung verwendeten Adjektive, so zeigt sich der Bogus-Pipeline-Effekt besonders deutlich für das «Vorurteilsobjekt» «TÜRKEN»: Während «Türken» unter Papier-Bleistift-Bedingungen nicht besser und nicht schlechter beurteilt werden als «Deutsche», werden sie unter den beiden BPL-Bedingungen wesentlich ungünstiger eingeschätzt.

Die vorliegenden Ergebnisse scheinen zu belegen, daß sich gewisse methodische Verbesserungen gegenüber den früheren Untersuchungen von SIGALL & PAGE (1971) bzw. MUMMENDEY et al. (1979) positiv ausgewirkt haben: Die Verwendung einer Reihe von aufgrund von Voruntersuchungen ausgewählten und item-analysierten Eigenschaftswörtern zur Beurtei-

lung der ethnischen Gruppen ermöglichte die Berechnung globaler Einstellungsmaße anstelle einer Betrachtung von Wechselwirkungen in bezug auf jede einzelne zu beurteilende Eigenschaft; während der BPL-Effekt sich für die globalen Einstellungsmaße durchweg einstellte und auch bei gleichzeitiger Berücksichtigung sämtlicher Eigenschaften in multivariaten Varianzanalysen nachzuweisen war, ergaben sich auf der Ebene von Einzelitems durchaus unterschiedliche Ergebnisse, was die wesentlich ungünstigeren Resultate der früheren Untersuchungen zum Teil erklären mag. Aufgrund postexperimenteller Befragung in der vorliegenden Untersuchung scheint sich auch erwiesen zu haben, daß mögliche Störvariablen für den BPL-Effekt, wie zum Beispiel der vermutete wissenschaftliche Nutzen der Untersuchung, die Beurteilung der Zuverlässigkeit der Versuchsgeräte, die geschätzten Kosten der Versuchsapparaturen und bisherigen Vorerfahrungen als Versuchsperson seitens der Probanden keinen bedeutsamen Einfluß auf die Ergebnisse, und das heißt auf die Gültigkeit des BPL-Paradigmas ausüben.

Allerdings wird auch durch das vorliegende Validierungsexperiment die schon von JONES & SIGALL (1971) aufgeworfene Frage nach dem «Warum» des Auftretens der Urteilsdifferenzen unter Bogus-Pipeline- und Paper-Pencil-Bedingungen nicht abschließend beantwortet. Alternative Erklärungen des BPL-Effektes zu der Annahme einer verringerten Antworttendenz der sozialen Erwünschtheit in der BPL-Situation lassen sich nach wie vor formulieren. Zum Beispiel wiesen JONES & SIGALL (1971, p. 362) auf die Möglichkeit hin, daß die BPL-Apparatur mit ihrem explizit genannten Anspruch, physiologische Begleiterscheinungen sozialer Einstellungen zu messen, die affektive Komponente einer Einstellung im Vergleich zu den mehr kognitiven Komponenten salient mache:

«If you ask a person, «What do you think?» he may respond differently than when you ask him «How do you feel?» Similarly, to ask the subject to predict what his autonomic nervous system is telling the machine is to focus him on the more primitive, affective components of the impression or attitude structure» (Hervorhebung durch uns).

Mit anderen Worten würde unter den unterschiedlichen Versuchsbedingungen (BPL ver-

sus Paper-Pencil) zumindest teilweise nicht das gleiche gemessen, sondern es würden in Termini des Komponentenmodells der Einstellung jeweils schwerpunktmäßig unterschiedliche Komponenten einer Einstellung erfaßt. Dies scheint bereits dadurch plausibel zu sein, daß im Experiment zwangsläufig mit unterschiedlichen Instruktionen gearbeitet werden muß. So wird zum Beispiel in der Paper-Pencil-Bedingung kein besonderer Hinweis darauf gegeben, daß die Versuchsperson sich auf ihre «wahren Gefühle» konzentrieren solle; unter den BPL-Bedingungen ist dies jedoch zur Begründung des Versuchsablaufes erforderlich. Gewisse Hinweise auf eine unterschiedliche Salienz der verschiedenen Einstellungskomponenten lassen sich aus Ergebnissen der postexperimentellen Befragung ableiten: Unter Papier-Bleistift-Bedingungen wird als Ziel der experimentellen Untersuchung häufiger die Erforschung von Einstellungen gegenüber ethnischen Gruppen angegeben, während unter BPL-Bedingungen häufiger die übrigen angebotenen Antwortkategorien angekreuzt wurden (Erprobung einer Befragungsmethode/Kennntnis eigener gefühlsmäßiger Einschätzungen/physiologische Reaktionen auf Einstellungsobjekte). Die Frage, ob der Bogus-Pipeline-Effekt primär dadurch zustande kommt, daß Antworttendenzen der sozialen Erwünschtheit vermindert und unter BPL-Bedingungen unverzerrt geantwortet wird oder ob in den unterschiedlichen Situationen unterschiedliche Aspekte von «Einstellung» gemessen werden, kann auch durch das vorlie-

gende Experiment nicht schlüssig beantwortet werden und bedarf weiterer experimenteller Bearbeitung.

### Literatur

- BRACKWEDE, D. 1980. Das Bogus-Pipeline-Paradigma: Eine Übersicht über bisherige experimentelle Ergebnisse. *Zeitschrift für Sozialpsychologie*, 11, 50–59.
- JONES, E. E. & SIGALL, H. 1971. The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76, 349–364.
- KARLINS, M., COFFMAN, T. L. & WALTERS, G. 1969. On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, 13, 1–16.
- KATZ, D. & BRALY, K. W. 1933. Racial stereotypes of one-hundred college students. *Journal of Abnormal and Social Psychology*, 28, 282–290.
- MUMMENDEY, H. D. & BOLTEN, H.-G. 1981. Die Veränderung von Social-Desirability-Antworten bei erwarteter Wahrheitskontrolle (Bogus-Pipeline-Paradigma). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 2, 151–156.
- MUMMENDEY, H. D., SCHIEBEL, B., TROSKE, U., HESENER, B. & BOLTEN, H.-G. 1979. Experimentelle Replikation des Bogus-Pipeline-Effekts für ethnische Stereotype. *Bielefelder Arbeiten zur Sozialpsychologie*, Nr. 55.
- OSTROM, T. M. 1973. The bogus pipeline: A new ignis fatuus? *Psychological Bulletin*, 79, 252–259.
- QUIGLEY-FERNANDEZ, B. & TEDESCHI, J. T. 1978. The bogus pipeline as a lie detector: Two validity studies. *Journal of Personality and Social Psychology*, 36, 247–256.
- SIGALL, H. & PAGE, R. 1971. Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology*, 18, 247–255.

