

# Neuere Erhebungsmethoden

Hans Werner Bierhoff

Für die psychologische Forschung steht eine Vielzahl von Erhebungsmethoden zur Verfügung (vgl. Breakwell, Hammond & Fife-Schaw, 1995). Traditionell wird zwischen Experiment, Beobachtung, Befragung und Inhaltsanalyse als den vier Hauptmethoden der Datengewinnung unterschieden (Roth, 1993; vgl. auch die Beiträge von Bredenkamp und Roskam, in diesem Band).

Eine erweiterte Perspektive ergibt sich, wenn den direkten die indirekten Meßverfahren gegenübergestellt werden (Kidder & Campbell, 1970; Petermann & Noack, 1993). Unter die direkten Methoden, bei denen Befragte und Untersucher in ihrem Verständnis der Zielsetzung der Studie übereinstimmen, lassen sich das Interview und die schriftliche Befragung subsumieren. Unter die indirekten Methoden, bei denen die Dateninterpretation von dem Verständnis der Untersuchung auf Seiten der Versuchsteilnehmer abweicht, fallen die nicht-reaktiven Verfahren (teilnehmende Beobachtung, Analyse von Spuren, Inhaltsanalyse von Texten). Der Begriff „nicht-reaktiv“ bezieht sich darauf, daß die Antworten der Versuchspersonen durch den Vorgang der Datenerhebung nicht oder nur geringfügig verfälscht werden. Dies ist ein bedeutender Vorteil gegenüber den direkten Verfahren. So könnte z.B. die direkte Frage danach, ob jemand an der letzten Bundestagswahl teilgenommen hat, zu übertriebenen Angaben über die Wahlbeteiligung führen, weil bestimmte Befragte nicht zugeben wollen, daß sie am Wahltag zu Hause geblieben sind.

Neben der Verfälschungsproblematik gibt es die Datenschutzproblematik, die Verfahren der Datenerhebung belastet. In vielen Fällen ist eine Person mit ihren Antworten unmittelbar oder potentiell identifizierbar. Daher entsteht das berechtigte Interesse, verbale Verfahren der Datenerhebung zu entwickeln, die die Anonymität der Befragten schützen bzw. grundsätzlich ausschließen, daß eine einzelne Person mit einer bestimmten Antwort (z.B. im Hinblick auf Drogenkonsum) identifiziert werden kann.

Viele Standardverfahren der Forschung (z.B. das Semantische Differential, vgl. Schäfer, 1983, oder das Fishbein-Ajzen-Modell der Einstellungsmessung, vgl. Frey, Stahlberg & Gollwitzer, 1993) werfen Probleme auf, wenn sie auf „sensible“ Themenbereiche wie Einschätzungen stigmatisierter Gruppen zur Erfassung von ambivalenten Gefühlen oder Selbsteinschätzungen abweichenden Verhaltens angewandt werden. Die drei Erhebungsmethoden, die in diesem Beitrag behandelt werden, dienen der Reduktion von Verfälschungstendenzen, der Sicherung des Datenschutzes oder beidem. Daß weniger Verfälschung und mehr Datenschutz positiv miteinander zusammenhängen können, wird deutlich, wenn man bedenkt, daß eine Person, für deren Antworten der Datenschutz gewährleistet ist, vielfach auch weniger motiviert sein wird, ihre Antworten in Richtung auf soziale Erwünschtheit zu verfälschen.

Die drei Erhebungsmethoden sind die randomisierte Antworttechnik, die Reaktionszeitmessung in der Einstellungsforschung und die Erfassung ambivalenter Gefühle. Alle drei Verfahren stellen verbale Verfahren dar. Die randomisierte Antworttechnik ist zu den direkten Meßverfahren zu rechnen, während die beiden anderen Verfahren unter die indirekten Meßverfahren fallen. Während die randomisierte Antworttechnik in vielen Bereichen der Meinungs- und Einstellungsforschung eingesetzt werden kann, ist die Reaktionszeitmessung auf die Erfassung von Assoziationsstrukturen anwendbar, die Einstellungen und Stereotypen zugrunde liegen. Schließlich stellt das Verfahren zur Erfassung von ambivalenten Einstellungen eine spezifische Methode dar, mit deren Hilfe erfaßt werden kann, ob die Beurteiler ambivalente Gefühle gegenüber den Zielgruppen empfinden.

## 1 Randomisierte Antworttechnik

Untersuchungen, in denen die randomisierte Antworttechnik (*randomized response technique*) eingesetzt wurde, beziehen sich z.B. auf die Frage an junge Frauen, ob sie vergewaltigt wurden (Soeken & Damrosch, 1986) oder auf die Frage an eine Zufallsstichprobe amerikanischer Männer, ob sie jemals ein Kind sexuell mißbraucht haben (Finkelhor & Lewis, 1988). Andere Beispiele für die Verwendung der randomisierten Antworttechnik liegen im Gesundheitsbereich, z.B. im Zusammenhang mit AIDS (Williams, Suen & Baffi, 1993).

Der Gedanke, der der randomisierten Antworttechnik zugrunde liegt, geht auf Warner (1965) zurück, der ein Zufallselement in die Befragung einführte, um die Identifizierbarkeit einer Antwort mit einer befragten Person auszuschließen. Nach diesem Verfahren wird den Befragten z.B. eine bestimmte Frage gestellt („Haben Sie Sympathien mit den Reps?“), wenn sie eine grüne Kugel aus einem Behälter per Zufall ziehen, und eine Frage, die das Gegenteil beinhaltet („Haben Sie keinerlei Sympathien mit den Reps?“), wenn sie eine rote Kugel ziehen. Dabei kann es z.B. so sein, daß 8 von 10 Kugeln grün sind und die restlichen zwei rot. Wenn die Wahrscheinlichkeit der Stellung der beiden Fragen ungleich  $1/2$  ist, läßt sich aus den Antworten mit einiger Verlässlichkeit ableiten, wie groß die Wahrscheinlichkeit ist, daß eine Person die in Frage stehende Einstellung (Sympathie mit den Reps) vertritt.

Auf der Basis des Ansatzes von Warner entwickelten Himmelfarb und Edgell (1980) ein quantitatives Verfahren der Meinungs- und Einstellungsforschung. Den Befragten wird Anonymität gewährleistet, indem ihnen prinzipiell die Möglichkeit geboten wird, die Frage korrekt oder verfälscht zu beantworten, ohne daß die Fragesteller wissen, ob die befragte Person die korrekte oder die verfälschte Antwort abgegeben hat. Die Version von Himmelfarb und Edgell beinhaltet die Methode der additiven Konstanten. Wenn die interessierende Frage z.B. lautet

„Wie hoch ist das Einkommen, das Ihrer Familie monatlich zur Verfügung steht?“,

könnte nach diesem Verfahren das folgende Vorgehen gewählt werden: Die Befragten erhalten jeweils 100 Karten, auf denen die zu beantwortende Frage steht. Sie ziehen eine der Karten aus dem Stapel, die sie beantworten. Diese Frage hat in 80% der Fälle die oben genannte Formulierung, während der Text in 10% der Fälle lautet:

„Nehmen Sie die Summe des Einkommens, das Ihrer Familie monatlich zur Verfügung steht, und addieren Sie die Zahl 500. Wie hoch ist die resultierende Zahl?“

In weiteren 10% der Fälle wird dieselbe Frage gestellt, allerdings mit dem Unterschied, daß 500 von dem Wert des Einkommens abgezogen wird. Wenn die Befragten diese Technik der Datenerhebung richtig verstehen und wenn sie bei der Auswahl und Beantwortung der Frage unbeobachtet bleiben, dann können sie sicher sein, daß die Fragesteller die wirkliche Höhe des Einkommens im Einzelfall nicht erschließen können.

Wenn nun die zu addierende Konstante mit  $k_i$  ( $i = 1, \dots, c$ ) bezeichnet wird (im obigen Beispiel weist die Konstante die Werte +500, 0 und -500 auf), wobei jedes  $k_i$  die Wahrscheinlichkeit  $p_i$  des Auftretens hat (im Beispiel .10, .80, .10), dann läßt sich die Antwort  $Z$  der Befragten wie folgt zerlegen:

$$Z = X + k_i, \quad (1)$$

wobei  $X$  die gesuchte Antwort ist und  $k_i$  mit der Wahrscheinlichkeit  $p_i$  auftritt. Die erwartungstreue Schätzung des Populationsmittelwertes der Variablen  $X$  ergibt sich dann als

$$\hat{\mu}_x = M_z - \sum_{i=1}^c p_i k_i, \quad (2)$$

wobei  $M_z$  den Stichprobenmittelwert von  $Z$  bezeichnet. Für die Varianz von  $\hat{\mu}_x$  ergibt sich unter Berücksichtigung von  $N$  (Anzahl der Befragten):

$$Var(\hat{\mu}_x) = Var(M_z) = Var(Z)/N. \quad (3)$$

Ausgehend von der Definition der Varianz einer Zufallsvariablen  $Z$  ergibt sich

$$Var(Z) = E(Z^2) - [E(Z)]^2. \quad (4)$$

Die Beziehung zwischen  $Var(Z)$  und  $\sigma_x^2$ , der Populationsvarianz von  $X$ , läßt sich aus (4) ableiten (vgl. Himmelfarb & Edgell, 1980):

$$Var(Z) = \sigma_x^2 + \sum_{i=1}^c p_i k_i^2 - \left( \sum_{i=1}^c p_i k_i \right)^2. \quad (5)$$

Betrachten wir nun das Beispiel  $\sigma_x = 700$ ,  $p_1 = .8$  für  $k_1 = 0$  und  $p_2 = .2$  für  $k_2 = 500$ . Wenn bisher die Vorteile der randomisierten Antworttechnik dargestellt wurden, so ist nun auch der Preis zu erwähnen, den man für die Anwendung dieser Technik zahlen muß. Die Varianz der Stichprobenverteilung der Mittelwerte fällt größer aus, als wenn man eine direkte Frage stellt. Unser Beispiel kann das verdeutlichen: Während  $Var(Z)$  den Wert von 530000 annimmt, was bei einem angenommenen  $N$  von 1000 zu einem Wert von  $Var(\hat{\mu}_x) = 530$  führt, beträgt  $\sigma_x^2 = 490000$  und somit  $\sigma_{M_x}^2 = 490$  bei  $N = 1000$ . Diese Vergrößerung der Varianz von 490 auf 530 läßt sich als relativ gering bezeichnen. Fügt man noch eine weitere Konstante  $k_3 = -500$  hinzu und setzt  $p_2$  und  $p_3$  auf .1 fest, nimmt die Varianz zu, da sich nun ein Wert von

$Var(Z) = 540000$  ergibt bzw. von  $Var(\hat{\mu}_x) = 540$ . Die Varianz nimmt auch zu, je kleiner die relative Anzahl der Fälle wird, in denen die Befragten ihre wirkliche Antwort geben sollen. Je größer die wahrgenommene Anonymität und je geringer die Wahrscheinlichkeit der Aufdeckung also wird, desto höher fällt die Varianz der Mittelwerte aus.

Wenn die randomisierte Antworttechnik angewandt wird, besteht das Interesse darin, den Mittelwert der Variablen  $X$  über die Erfassung der Variablen  $Z$  zu bestimmen (vgl. Gleichung (1)). In dem Spezialfall, in dem die Konstante mit gleicher Wahrscheinlichkeit addiert oder subtrahiert werden soll (wenn die Frage nicht direkt beantwortet werden soll), ergibt sich aus Gleichung (2)

$$\hat{\mu}_x = M_z. \quad (6)$$

In dem Fall, daß eine Konstante  $k_2$  mit der Wahrscheinlichkeit  $p_2$  addiert werden soll, ergibt sich

$$\hat{\mu}_x = M_z - p_2 k_2. \quad (7)$$

In diesem zweiten Spezialfall muß die mit der Wahrscheinlichkeit des Auftretens multiplizierte Konstante von dem Mittelwert von  $Z$  subtrahiert werden, um die erwartungstreue Schätzung des Populationsmittelwertes der Variablen  $X$  zu erhalten.

Die randomisierte Antworttechnik liegt in vielen Varianten vor, die an dieser Stelle nicht ausführlicher behandelt werden können (vgl. aber Boruch & Cecil, 1979; Fox & Tracy, 1986). Eine schon frühzeitig vorgeschlagene Alternative besteht darin, die interessierende Frage und eine nicht interessierende Frage zu stellen, deren Beantwortung harmlos ist (Greenberg, Abu-Ela, Simmons & Horvitz, 1969; vgl. auch Folsom, Greenberg, Horvitz & Abernathy, 1973; Scheers & Dayton, 1987). In diesem Fall werden zwei Stichproben benötigt, um die Proportionen für die sensitive und die nicht-sensitive Frage zu schätzen, es sei denn, die Wahrscheinlichkeit für das Auftreten einer Zustimmung bei der harmlosen Frage ist im vorhinein bekannt (z.B. „Wurden Sie im Monat Juli geboren?“).

Generell läßt sich vermuten, daß die randomisierte Antworttechnik dazu beiträgt, daß sozial wenig erwünschte Verhaltensmuster eher zugegeben werden (Himmelfarb & Lickteig, 1982) bzw. daß validere Antworten in kritischen Fragebereichen gegeben werden (Scheers & Dayton, 1987). Zum Beispiel zeigte sich, daß bei der Feststellung „Ich habe bei einer Prüfung gemogelt, indem ich die Antworten meines Nachbarn abgeschrieben habe“ von den schwachen Studenten, die die niedrigsten Werte im Grade Point Average (GPA) aufwiesen, 86% ein Mogeln nach der randomisierten Antworttechnik einräumten, während nur 47% dieser Studenten bei einer anonymen Direktbefragung ein Mogeln eingestanden. (Bei den Studenten mit den besten GPA-Werten waren es 21% bzw. 13%).

Ein weiterer potentieller Validitätshinweis besteht darin, daß die Ergebnisse mit der randomisierten Antworttechnik mit Resultaten aus amtlichen Statistiken übereinstimmen. Ein gutes Beispiel für den zuletzt genannten Punkt könnte die Erhebung der Teilnahme an einer Wahl sein, da die objektive Wahlbeteiligung amtlich ermittelt wird.

Die randomisierte Antworttechnik wirft eine Reihe von Problemen auf, die bisher nur teilweise gelöst sind. Dazu zählt die Frage von statistischen Tests, die z.B.

durchgeführt werden können, um zu prüfen, ob die Technik ihr Ziel der Verringerung der Verfälschungstendenz erreicht hat (Boruch & Cecil, 1979). Angenommen, eine Stichprobe von Befragten wird nach Zufall in zwei Gruppen unterteilt: Eine Gruppe wird nach einem sensiblen Thema mit der randomisierten Antworttechnik befragt, eine zweite Gruppe wird in konventioneller Weise direkt befragt. Werden in der ersten Gruppe mehr sozial unerwünschte Verhaltenstendenzen eingestanden als in der zweiten? Zur Beantwortung dieser Frage lassen sich Tests zum Vergleich von Proportionen einsetzen (vgl. Levy, 1977).

Alternative Methoden der Gewinnung von wenig verfälschten Daten, die auch dem Anspruch des Datenschutzes dienen, stehen zur Verfügung (Boruch & Cecil, 1979). Dazu zählt insbesondere die individuelle Aggregation von Antworten, die hier nur durch ein Beispiel illustriert werden kann. Angenommen, es soll erfaßt werden, ob eine Person harte Drogen nimmt. Eine Frage könnte lauten: „Wie oft haben Sie im letzten Monat harte Drogen genommen?“. Außerdem wird die befragte Person gebeten, eine „harmlose“ Frage zu beantworten, z.B. „Wie oft waren Sie im letzten Monat im Kino?“. Die Befragten geben die Antworten aber nicht einzeln ab, sondern werden gebeten, nur die Summe der beiden Antworten anzugeben. Wenn nun in einer vergleichbaren Stichprobe dieselben Fragen vorgegeben werden und die Instruktion gegeben wird, die Differenz zu bilden und anzugeben, besteht die Möglichkeit, die Häufigkeit des Drogenkonsums zu schätzen (vgl. Boruch & Cecil, 1979).

## 2 Reaktionszeitmessungen

Reaktionszeitmessungen bei der Abgabe von Stellungnahmen zu Einstellungsfeststellungen können wertvolle Zusatzinformationen liefern, die sich auf die Stärke und die assoziative Struktur der Einstellungen beziehen. Dies wird im folgenden an zwei Beispielen verdeutlicht, von denen sich das erste allgemein mit der Erfassung der Einstellungsstärke befaßt, während das zweite von der Annahme ausgeht, daß Stereotype sich in Assoziationen zwischen negativen/positiven Eigenschaften und ethnischen Gruppen zeigen, deren Stärke über die Reaktionszeiten erfaßt wird.

### 2.1 *Einstellungsstärke*

Es ist schon seit längerem bekannt, daß die Abfrage von Einstellungen – insbesondere bezogen auf solche Einstellungen, die außerhalb des Erfahrungshorizonts der Befragten liegen – von den Befragten mehr oder weniger willkürlich beantwortet werden, ohne daß die Angabe eine Vorhersage auf das entsprechende Verhalten gestattet. Wie kann die Stärke einer Einstellung bestimmt werden und zwischen solchen Einstellungen, die momentan elaboriert werden, und solchen, die das Verhalten der Beurteiler beeinflussen, unterschieden werden? Es ist offensichtlich von großem angewandten Nutzen, wenn die Möglichkeit besteht, neben der Richtung und Ausprägung der Einstellung auch etwas über ihre Stärke aussagen zu können.

Die Theorie der Zugänglichkeit von Einstellungen, die Fazio (1989) entwickelt hat, befaßt sich mit der Erfassung der Stärke von Einstellungen, die als Assoziationen zwischen einem gegebenen Einstellungsobjekt und seiner Bewertung aufgefaßt werden. Objekte können politische Themen, Situationen, Personen oder Ge-

genstände sein. Bewertungen können bedingte emotionale Reaktionen, positiv oder negativ bewertete Meinungen über die Nutzen eines Objekts oder positive/negative Erfahrungen aus früheren Begegnungen mit dem Einstellungsobjekt sein. Schließlich gilt, daß die Assoziationen zwischen Einstellungsobjekt und Bewertung stärker oder schwächer sein könnten, was als Indiz für die Einstellungsstärke aufgefaßt wird.

Damit eine Einstellung bei der Bewertung eines Einstellungsobjekts benutzt wird, muß sie aus dem Gedächtnis abgerufen werden. Das ist umso eher der Fall, je zugänglicher die Einstellung ist. Nach der Auffassung der Einstellung als Assoziation zwischen Objekt und Bewertung ergibt sich die Konsequenz, daß die Einstellungsstärke Ergebnis assoziativen Lernens ist. Je häufiger die Bewertung mit dem Einstellungsobjekt verbunden wurde, desto zugänglicher sollte die Einstellung sein, und desto schneller sollte sie abrufbar sein.

Untersuchungstechnisch erfaßt Fazio (1989) die Zugänglichkeit einer Einstellung über die Reaktionszeit. Schnelle Bewertungen eines Einstellungsobjekts verweisen auf hohe Zugänglichkeit, langsame Bewertungen auf niedrige Zugänglichkeit. Seitdem über die Erfassung von Reaktionen am Personalcomputer eine problemlose Messung der Reaktionszeit möglich ist, bietet dieses Verfahren eine vielfach anwendbare Messung der Einstellungsstärke, und zwar unabhängig davon, welche inhaltliche Einstellung erfaßt wird.

In einem Experiment von Powell und Fazio (1984) wurde gezeigt, daß auf ein Einstellungsthema umso schneller reagiert wurde, je häufiger es zuvor dargeboten worden war. Bei sechs Wiederholungen war die Reaktionszeit und damit die Zugänglichkeit der Einstellung größer als bei einer vorherigen Darbietung oder bei keiner vorherigen Darbietung.

Die angewandte Bedeutung der Messung der Einstellungsstärke über die Reaktionszeit wurde in einer Untersuchung von Fazio und Williams (1986) anhand der 1984er US-Präsidentenwahlen, bei der Reagan und Mondale zur Wahl standen, aufgezeigt. Bei Wahlprognosen besteht das bekannte Problem, das tatsächliche Wahlverhalten vorherzusagen. Der Ansatz von Fazio und Williams ermöglicht nun die Unterscheidung zwischen Personen, deren Wahlverhalten gut prognostizierbar ist, und solchen Personen, deren Wahlverhalten weniger gut prognostizierbar ist.

245 Wähler nahmen an einer Befragung teil, bei der 25 Einstellungsfeststellungen auf einer Fünf-Punkte-Skala so schnell wie möglich zu beurteilen waren. Die Fragen wurden über Tonband dargeboten und die Reaktion in einen Computer von den Befragten durch Knopfdruck eingegeben, der die Reaktionszeit erfaßte. Die für die Studie kritischen beiden Feststellungen lauteten: „Ein guter Präsident für die nächsten vier Jahre wäre Ronald Reagan (Walter Mondale)“.

Die Wahl fand drei Monate nach der Einstellungsmessung statt. Das Wahlverhalten korrelierte hoch mit der Einstellung zu den Kandidaten, insbesondere mit der gegenüber Reagan. Um den Einfluß der Einstellungszugänglichkeit zu untersuchen, wurde aufgrund der Reaktionszeit am Median eine Trennung nach hoher bzw. niedriger Verfügbarkeit der Einstellung durchgeführt. Bei niedriger Verfügbarkeit sollte die Einstellung mit der Wahl niedriger korrelieren als bei hoher Verfügbarkeit. Die Ergebnisse zeigen, daß die Korrelation bei hoher Zugänglichkeit generell höher ausfiel. Bei hoher Zugänglichkeit betrug die Korrelation  $r = .89$ , bei niedriger Zugänglichkeit  $r = .66$ ; dieser Unterschied war statistisch signifikant.

## 2.2 *Stereotype*

Während sich das Verfahren von Fazio (1989) allgemein auf die Erfassung der Einstellungsstärke über die Reaktionszeiten richtet, entwickelten Gaertner und McLaughlin (1983) ein Verfahren zur Erfassung ethnischer Stereotype, das auf den Vergleich von Reaktionszeiten auf unterschiedliche Zielgruppen hinausläuft. Sie gingen von der Annahme aus, daß Stereotype Assoziationen zwischen (Bezeichnungen von) ethnischen Gruppen (z.B. Schwarze) und negativen oder positiven Eigenschaften darstellen. Aus lexikalischen Entscheidungsaufgaben entlehnten sie die Erkenntnis, daß hohe Assoziationen schneller zu einer Antwort auf bestimmte Fragen (wie „Sind beide Begriffe sinnvoll?“) führen als geringe Assoziationen.

Aufgrund dieser Überlegungen lassen sich ethnische Stereotype durch Reaktionszeitmessungen erfassen. Die Beurteiler müssen angeben, ob zwei Buchstabenfolgen, die gleichzeitig dargeboten werden, sinnvolle Worte sind. Für fehlerlose schnelle Reaktionen wird ein finanzieller Anreiz geboten. Den Beurteilern werden nun sowohl Paare dargeboten, in denen eine erfundene Buchstabenfolge auftritt, als auch solche Paare, in denen ein Begriff eine ethnische Gruppe („Schwarze“ oder „Weiße“) und ein Begriff eine positive (z.B. „smart“) oder negative (z.B. „faul“) Eigenschaft bezeichnet.

Unter der Annahme, daß Stereotype sich in Assoziationen zwischen negativen bzw. positiven Eigenschaften und ethnischen Gruppen zeigen, läßt sich aus den Reaktionszeiten das negative/positive Stereotyp der Beurteiler ableiten. In zwei Experimenten (Gaertner & McLaughlin, 1983) wurde mit weißen Studenten als Versuchspersonen festgestellt, daß bei schwarzen und weißen Zielgruppen keine Unterschiede in den Reaktionszeiten für negative Begriffe auftraten, während für positive Begriffe beobachtet wurde, daß die Reaktionszeiten bei der weißen Zielgruppe kürzer waren als bei der schwarzen Zielgruppe. Das Ergebnismuster, das für ein explizit positives Stereotyp von Weißen gegenüber Weißen spricht, konnte auch in einer direkten Einschätzung in einem weiteren Experiment bestätigt werden.

In diesem Zusammenhang ist darauf hinzuweisen, daß durch neue Erhebungsmethoden ältere Verfahren nicht ersetzt werden sollen. Vielmehr sind diese neuen Verfahren als ergänzende Methoden zu verstehen, die – von einem Multiplen-Indikator-Ansatz mit heterogenen Verfahren ausgehend (multipler Operationalismus, vgl. Bierhoff & Rudinger, in diesem Band) – einen großen Wert für die „Triangulation“ oder „Anpeilung und Vermessung“ eines theoretischen Konstrukts besitzen (Crano, 1981). Alternative Verfahren zur Erfassung der Intergruppen-Diskriminierung (z.B. über Unterschiede im hilfreichen Verhalten gegenüber Mitgliedern verschiedener ethnischer Gruppen) werden von Crosby, Bromley und Saxe (1980) und Petermann und Noack (1993) beschrieben.

## 3 Das Ambivalenz-Indifferenz-Problem

Bei Verwendung von bipolaren Ratingskalen stellt sich die Frage, wie neutrale Antworten zu interpretieren sind. Das semantische Differential (Schäfer, 1983) beinhaltet die Einschätzung bestimmter Begriffe (wie ethnische Gruppen) auf bipolaren Skalen nach folgendem Schema:

Negativer Pol (-3) (-2) (-1) (0) (1) (2) (3) Positiver Pol

Auf einer solchen Urteilsskala läßt sich eine 0-Antwort entweder als neutrale Antwort der Indifferenz oder aber als Ambivalenz interpretieren. Im letzteren Fall sind die Beurteiler sowohl der Ansicht, daß das Konzept positive Merkmale hat, als auch der Ansicht, daß es negative Merkmale hat. Wenn sich die positiven und negativen Merkmale genau aufwiegen, käme die Antwort „0“ zustande.

Unter ambivalenter Einstellung läßt sich verstehen, daß eine Person sowohl positive als auch negative Gefühle gegenüber dem Konzept hat und es z.B. sowohl für „gut“ als auch für „schlecht“ hält. Die Erfassung von ambivalenten Gefühlen ist für die Motivations- und Einstellungsforschung von großer Bedeutung, weil die Berücksichtigung der Ambivalenz eine differenzierte Erfassung des Einstellungssystems in Übereinstimmung mit bedeutsamen theoretischen Ansätzen (wie Ambivalenz-Vergrößerungs-Theorie und Autoritäre Persönlichkeit) ermöglicht.

Außerdem ist es für bipolare Ratingskalen von grundsätzlicher Bedeutung, zwischen Ambivalenz und Indifferenz zu unterscheiden. Kaplan (1972) stellte eine Variante der Technik des Semantischen Differentials dar, die eine solche Unterscheidung ermöglicht. Die Grundidee besteht darin, daß die Endanker eines Semantischen Differentials einzeln auf Ratingskalen eingeschätzt werden. Betrachten wir z.B. das Gegensatzpaar „intelligent – dumm“, das in einem Semantischen Differential vorgegeben werden könnte. Eine Aufspaltung der positiven und negativen Gefühlskomponente in zwei Einschätzungen führt zu folgenden unipolaren Skalen:

(0)	(1)	(2)	(3)	Positiver Pol (Intelligent)
				Negativer Pol (Dumm)
				(-3) (-2) (-1) (0)

In der Praxis können den Beurteilern vierstufige Ratingskalen mit den Endpunkten 0 (= stimmt nicht) und 3 (= stimmt zweifellos) vorgelegt werden. Die Beurteilung des negativen Endpols wird dann entsprechend umkodiert (z.B. 3 in -3). Die getrennte Einschätzung der gegensätzlichen Pole ermöglicht es den Beurteilern, dem positiven und dem negativen Endpol der Eigenschaftsdimension gleichermaßen zuzustimmen, da sie sowohl ein Urteil über die positive Einstellungskomponente ( $A_p$ ) als auch über die negative Einstellungskomponente ( $A_n$ ) abgeben. Wenn beide Komponenten gleich stark ausgeprägt sind, nimmt die Ambivalenz zu, umso größer die Ausprägungen sind (z.B. ist die Ambivalenz bei einem 3, -3 Antwortmuster mit dem Wert 6 größer als bei einem 1, -1 Antwortmuster mit dem Wert 2). Die positive und die negative Einstellungskomponente gegenüber stigmatisierten Gruppen wie Behinderten oder Schwarzen korrelieren nur schwach negativ (Katz, Wackenhut & Glass, 1986).

Wenn man davon ausgeht, daß die positive und die negative Komponente voneinander unabhängig sind, dann lassen sie sich als orthogonale Dimensionen im zweidimensionalen Raum geometrisch darstellen. Von dieser geometrischen Repräsentation ausgehend leitete Kaplan (1972) mehrere Indizes ab, auf deren geometrische Interpretation an dieser Stelle nicht weiter eingegangen wird. Der Gesamtaffekt  $TA$  wird



wie folgt definiert:

$$TA = A_p + |A_n| . \quad (8)$$

Nur wenn der Gesamteffekt 0 beträgt (wenn also  $A_p$  und  $A_n$  gleicherweise den Wert 0 aufweisen), kann man von Indifferenz sprechen. Der Gesamteffekt setzt sich aus der Ambivalenz ( $AMB$ ) – definiert als das Ausmaß der genauen Ausbalancierung von positiver und negativer Komponente – und der Polarisierung ( $POL$ ) – der absolute Wert des gerichteten Residuals – zusammen, die wie folgt berechnet werden:

$$POL = |A| , \quad (9)$$

wobei

$$A = A_p + A_n \quad (10)$$

und

$$AMB = TA - POL. \quad (11)$$

Der Ambivalenzindex entspricht bei balancierten Antworten (wie 1, -1) dem Gesamteffekt, da dann die Polarisierung 0 ist. Bei abweichender Stärke der Antworten auf beiden Komponenten (wie 1, -2) entspricht der Ambivalenzwert aber nur einem Teil des Gesamteffektes, während ein weiterer Beitrag durch die Polarisierung zustandekommt. Wenn entweder  $A_p$  oder  $A_n$  den Wert 0 aufweist, ist die  $AMB = 0$  und  $TA$  entspricht der Polarisierung.

In einer unveröffentlichten Studie (Bierhoff, Seegel, Mehrke & Dietz) verwandten wir diesen Ambivalenzwert, um ambivalente Gefühle gegenüber Behinderten bei dem Pflegepersonal eines Rehabilitationszentrums für geistig und körperlich Schwerstbehinderte zu erfassen. Zum Vergleich wählten wir zwei Stichproben aus. Die erste stammte aus der allgemeinen Bevölkerung und wurde nach Alter und Geschlecht dem Pflegepersonal angeglichen. Die zweite bestand aus Studenten, die zusammen mit behinderten Personen in einem Studentenheim wohnten. Allen Beurteilern wurden die Endanker von 65 Semantischen Differentialen – also 130 Eigenschaftsratings – vorgelegt. Die Eigenschaften lassen sich den Bereichen Gefühle gegenüber Behinderten (z.B. angsterzeugend–beruhigend), körperliche Merkmale (z.B. unbeholfengeschickt), geistig–intellektuelle Merkmale (z.B. undifferenziert–differenziert) und Persönlichkeit (z.B. albern–ernst) zuordnen.

Die Ergebnisse zeigten, daß das Pflegepersonal des Rehabilitationszentrums in allen Bereichen die höchsten Ambivalenzwerte erreichte. Dieser Unterschied war besonders stark für die geistig–intellektuellen Merkmale ausgeprägt. Hier unterschied sich die Rehabilitationsgruppe auf dem Index der Ambivalenz ( $M_{AMB} = 5.51$ ) bedeutsam ( $p < .05$ ) von den beiden anderen Gruppen, während sich die Stichprobe der allgemeinen Bevölkerung ( $M_{AMB} = 2.37$ ) und die Stichprobe der Studenten, die in einem Wohnheim mit Körperbehinderten zusammenwohnten ( $M_{AMB} = 1.30$ ), nicht bedeutsam voneinander unterschieden.

Die Ergebnisse sind nur mit großer Vorsicht zu interpretieren, da die Untersuchung keine Zufallsaufteilung auf die drei Versuchsgruppen beinhaltet. Sie sollte auch hier an erster Stelle das Verfahren der Bestimmung von Ambivalenzwerten demonstrieren. Durch die Parallelisierung der Stichprobe des Rehabilitationspersonals und der Stichprobe aus der allgemeinen Bevölkerung wurde der Versuch unternommen, Störvariablen, die in der Stichprobenzusammensetzung liegen, zumindest

ansatzweise zu kontrollieren. Durch die zweite Vergleichsgruppe der Studentenheimbewohner sollte überprüft werden, ob der Kontakt mit Behinderten ausreicht, um höhere Ambivalenzwerte zu erzielen.

Generell wurden nur wenige Untersuchungen durchgeführt, in denen der Ambivalenz-Index von Kaplan (1972) eingesetzt wurde. Eine Ausnahme bildet die Studie von Katz, Glass, Lucido und Farber (1977), in der ebenfalls die Ambivalenz gegenüber Behinderten erfaßt wurde. Die Autoren konnten zeigen, daß der Ambivalenz-Index mit der Abwertung behinderter Opfer zusammenhängt: Je höher die Ambivalenz, desto größer war die Abwertung ausgeprägt. Dieses Ergebnis wurde im Sinne der Ambivalenz-Vergrößerungs-Theorie interpretiert, die besagt, daß Ambivalenz gefühlsmäßige Reaktionen gegenüber stigmatisierten Gruppen verstärkt.

#### 4 Abschließender Kommentar

Die dargestellten neueren Erhebungsmethoden lassen sich alle zur Erfassung von Einschätzungen negativer Begriffe (wie sozial unerwünschter Handlungen und stigmatisierter Gruppen) einsetzen. Dieser Schwerpunkt ist nicht zufällig, sondern bringt zum Ausdruck, daß die Erfassung der Beurteilung negativer Sachverhalte Probleme der sozialen Erwünschtheit aufwirft, deren Lösung dem Ziel der Erhöhung der Validität der Beurteilung dient. Bei den besprochenen Verfahren liegt der Schwerpunkt auf verbalen Meßverfahren, die entwickelt wurden, um Verfälschungstendenzen entgegen zu wirken bzw. Fehlerquellen zu reduzieren. Der besondere Vorteil der randomisierten Antworttechnik besteht in diesem Zusammenhang darin, daß sie ein direktes Meßverfahren darstellt, das von der Kooperation der Befragten ausgeht. Andererseits stellen Reaktionszeitmessungen und die Erfassung ambivalenter Gefühle indirekte Meßverfahren dar, die sich im Bereich des ethisch Vertretbaren bewegen und die eine ergänzende Auswertung verbaler Antworten ermöglichen.

#### 5 Weiterführende Literatur

Die meisten Veröffentlichungen zu neueren Erhebungsmethoden konzentrieren sich auf einen bestimmten methodischen Aspekt oder auf ein konkretes angewandtes Forschungsthema. Sie sind durchaus lesenswert, weil sie die Variationen der Verfahren und die kreative Forschungsweise in diesem Bereich verdeutlichen. Fox und Tracys (1986) Band stellt eine gute Übersicht über Ansätze dar, die die Methode der randomisierten Antworttechnik verwenden. Neuere Untersuchungen zur Messung der Einstellungsstärke über die Reaktionszeit werden von Fazio (1989) zusammengefaßt. Die Erfassung gefühlsmäßiger Ambivalenz besprechen Katz, Wackenhut und Glass (1986) sowie Breckler (1994).

#### Literaturverzeichnis

- Bierhoff, H. W., Seegel, F., Mehrke, S. & Dietz, F. (unveröffentlicht). *Ambivalente Einstellungen gegenüber Behinderten*. Marburg: Fachbereich Psychologie der Universität.
- Boruch, R. F. & Cecil, J. S. (1979). *Assuring the confidentiality of social research data*. Philadelphia: University of Pennsylvania Press.

- Breakwell, G. M., Hammond, S. & Fife-Schaw, C. (Eds.) (1995). *Research methods in psychology*. London: Sage.
- Breckler, S. J. (1994). A comparison of numerical indexes for measuring attitude ambivalence. *Educational and Psychological Measurement*, 54, 350–365.
- Crano, W. D. (1981). Triangulation and cross-cultural research. In M. B. Brewer & B. E. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 317–344). San Francisco: Jossey Bass.
- Crosby, F., Bromley, S. & Saxe, L. (1980). Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological Bulletin*, 87, 546–563.
- Fazio, R. H. (1989). On the power and functionality of attitudes: The role of attitude accessibility. In A. R. Pratkanis, S. J. Breckler & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 153–179). Hillsdale: Erlbaum.
- Fazio, R. H. & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: An investigation of the 1984 president elections. *Journal of Personality and Social Psychology*, 51, 505–514.
- Finkelhor, D. & Lewis, I. A. (1988). An epidemiologic approach to the study of child molestation. *Annals of the New York Academy of Sciences*, 528, 64–78.
- Folsom, R. E., Greenberg, B. G., Horvitz, D. G. & Abernathy, J. R. (1973). The two alternate questions randomized response model for human surveys. *Journal of the American Statistical Association*, 68, 525–530.
- Fox, J. A. & Tracy, P. E. (1986). *Randomized response. A method for sensitive surveys*. Beverly Hills, CA: Sage.
- Frey, D., Stahlberg, D. & Gollwitzer, P. M. (1993). Einstellung und Verhalten: Die Theorie des überlegten Handelns und die Theorie des geplanten Verhaltens. In D. Frey & M. Irle (Hrsg.), *Theorien der Sozialpsychologie, Band 1: Kognitive Theorien* (S. 361–398). Bern: Huber.
- Gaertner, S. L. & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascription of positive and negative characteristics. *Social Psychology Quarterly*, 46, 23–30.
- Greenberg, B. G., Abu-Ela, A., Simmons, W. R. & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, 64, 520–539.
- Himmelfarb, S. & Edgell, S. E. (1980). Additive constants model: A randomized response technique for eliminating evasiveness to quantitative response questions. *Psychological Bulletin*, 87, 525–530.
- Himmelfarb, S. & Lickteig, C. (1982). Social desirability and the randomized response technique. *Journal of Personality and Social Psychology*, 43, 710–717.
- Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin*, 77, 361–372.
- Katz, I., Glass, D. C., Lucido, D. J. & Farber, J. (1977). Ambivalence, guilt and the denigration of a physically handicapped victim. *Journal of Personality*, 45, 419–429.
- Katz, I., Wackenhut, J. & Glass, D. C. (1986). An ambivalence-amplification theory of behavior toward the stigmatized. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 103–117). Chicago: Nelson-Hall.
- Kidder, L. H. & Campbell, D. T. (1970). The indirect testing of social attitudes. In G. F. Summers (Ed.), *Attitude measurement* (pp. 333–385). Chicago: Rand McNally.
- Levy, K. J. (1977). The randomized response technique and large sample comparisons among the parameters of  $k$  independent binomial populations. *Psychological Bulletin*, 84, 244–246.

- Petermann, F. & Noack, H. (1993). Nicht-reaktive Meßverfahren. In E. Roth (Hrsg.), *Sozialwissenschaftliche Methoden* (S. 440–460). München: Oldenbourg.
- Powell, M. C. & Fazio, R. H. (1984). Attitude accessibility as a function of repeated attitudinal expression. *Personality and Social Psychology Bulletin*, 10, 139–148.
- Roth, E. (Hrsg.) (1993), *Sozialwissenschaftliche Methoden* (3. Aufl.). München: Oldenbourg.
- Schäfer, B. (1983). Semantische Differential Technik. In H. Feger & J. Bredekamp (Hrsg.), *Datenerhebung* (= Enzyklopädie der Psychologie, Themenbereich B, Serie I, Band 2, S. 154–221). Göttingen: Hogrefe.
- Scheers, N. J. & Dayton, C. M. (1987). Improved estimation of academic cheating behavior using the randomized response technique. *Research in Higher Education*, 26, 61–69.
- Soeken, K. L. & Damrosch, S. P. (1986). Randomized response technique: Application to research on rape. *Psychology of Women Quarterly*, 10, 119–125.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63–69.
- Williams, B. L., Suen, H. K. & Baffi, C. R. (1993). A controlled randomized response technique. *Evaluation and the Health Professions*, 16, 225–238.