

Aided Recall und Recognition in Belson's Leserschaftsuntersuchung von 1962

Im Jahre 1962 wurde Belson's Buch "Studies in Readership" veröffentlicht, in dem er über das von ihm durchgeführte grundlegende Forschungsprojekt berichtete. Seine Ergebnisse, Schlussfolgerungen und Empfehlungen hatten einen ziemlich grossen Einfluss auf die Leserschaftsforschung in Europa und möglicherweise darüber hinaus. Es ist bedauerlich, dass das Buch vergriffen ist, denn es verdient, auch noch heute und nicht nur von Mediaforschern gelesen und wiedergelesen zu werden: es behandelt methodologische Themen, die generell für Forscher relevant sind.

Der gesunde Menschenverstand hat gesagt, und die Erkenntnistheoretiker haben es bestätigt: Die Gewinnung von Wissen ist ein zirkulärer Prozess, – hoffentlich von der Art einer Wendeltreppe. Rupert Riedl, Biologe und Mitbegründer der Evolutionären Erkenntnistheorie, hat diesen Prozess sehr überzeugend dargestellt (Riedl 1980). Dies als eine Erklärung dafür, warum ich zu Belson's Forschung zurückgekehrt bin: Indem man neues Wissen auf alte Daten anwendet, hat man die Möglichkeit, zusätzliche Einsichten zu gewinnen.

Das Belson'sche Vorgehen

Da der Zugang zu Belson's Buch begrenzt ist, gebe ich eine kurze Beschreibung seiner Methoden und Ergebnisse. Sein Konzept war, zunächst Interviews mit dem üblichen National Readership Survey -Verfahren des I.P.A. (Institute of Practitioners in Advertising, Verband der Werbeagenturen) durchzuführen, um in einer zweiten Befragungswelle bei den gleichen Befragten eine ziemlich intensive Ermittlung des Lesens vorzunehmen.

Belson verzichtete darauf, das zweite Interview als einen Validierungsversuch zu bezeichnen, und schrieb stattdessen:

„Das zweite Interview war so gestaltet, dass es im besten Fall eine Überprüfung der Genauigkeit von gewissen Schätzungen der Befragten im ersten Interview war; im schlimmsten Fall kann das zweite Interview als eine Art von Zuverlässigkeitsprüfung betrachtet werden, bezogen auf diese Publikationen.“

In seinen Tabellen nannte Belson die Differenzen zwischen den Resultaten der normalen und der intensiven Interviews „Fehler“, erklärte aber mit angemessener Vorsicht: „Ich gebe nicht vor, dass der Begriff „Fehler“ ein völlig befriedigender Ersatz für „Diskrepanz zwischen I.P.A. und intensiven Schätzungen“ ist;

aber es gibt sehr gute Gründe anzunehmen, dass die Technik der intensiven Befragung die Begriffe in etwa austauschbar macht.“

Belson's Zurückhaltung, von „Validierung“ zu sprechen, ist relevant für die gegenwärtige Diskussion darüber, ob Validierung überhaupt möglich ist (z. B. Opfer und Müller-Veeh 1983; Wendt 1984). Sie entspricht der grundsätzlichen Feststellung von Popper, dass eine Verifizierung, damit Validierung nicht möglich ist; inzwischen hat man erkannt, dass auch die Falsifizierung nicht möglich ist.

Belson verglich die Resultate dieser zwei Messungen unter den Überschriften „I.P.A.“ und „Intensive“. Mir scheint, dass es guten Grund gibt, sie in „Aided Recall“ und „Recognition“ umzubenennen. Die jeweiligen Methoden gehören zu diesen Forschungskategorien; und indem man sie in den weiteren Zusammenhang bringt, kann man vielleicht eines Tages – hier wie anderswo – besser verstehen, was getan wurde und was die Resultate bedeuten.

Aided Recall

In dem ersten Interview wurde die Methode der damaligen I.P.A.- Leserschaftsumfragen angewandt, mit Interviewern, die üblicherweise in diesen Umfragen eingesetzt wurden. Dabei wurde für die Identifikation der 91 behandelten Publikationen eine Aided Recall-Methode verwendet, die aus einer Art Büchlein bestand, die Fotografien der Titelseile oder „Titel-Signets“ von jeder Zeitung und jeder Zeitschrift enthielt. Die Befragten sollten durch dieses Büchlein durchgehen und für jede Publikation sagen, ob sie diese in den letzten drei Monaten angesehen haben oder nicht.

Es ist zu beachten, dass der Begriff „looked at“ („angesehen“) benutzt wurde. Belson's Forschung entdeckte die Tatsache, dass dieser Begriff zu Missverständnissen führte, was gemeint sei: Nur auf die Titelseite geblickt oder ins Heft selbst hineingeschaut? Die I.P.A. - Ergebnisse haben mindestens teilweise darunter gelitten. Die Zeitidentifikation beruhte auf der Frage: Wann war das letzte Mal, als Sie sich diese angesehen haben? Es wurde keine optische Hilfe gegeben, aber die Interviewer hatten nachzufragen, bis es „möglich war, die Antwort als innerhalb einer bestimmten Periode befindlich oder außerhalb derselben zu klassifizieren“.

Recognition

Die Verfahrensweise der zweiten, sogenannten „Intensiv-Befragung“ hatte drei wesentliche Bestandteile:

- 1.) Fragen nach dem „üblichen“ Leseverhalten;
- 2.) Identifikation der Publikationen;
- 3.) Feststellung des letzten Lesens der Publikationen.

Ad 1.) Übliches Leseverhalten

Wie die praktische Erfahrung in der Markt-, Media-, Werbe- und Sozialforschung zeigt, haben die Befragten eine Tendenz, im Sinne ihres „gewöhnlichen“ oder „normalen Verhaltens“ zu antworten, wenn sie nach ihrem speziellen Verhalten gefragt werden. Dies ist leicht verständlich, denn ...

- a) die Leute wollen nicht aufgrund ihres ungewöhnlichen oder abnormalen Verhaltens „beurteilt“ werden;
- b) sie erwarten, dass die Forscher daran interessiert sind, ein „normales“ Bild der Handlungen und Gedanken der Leute zu bekommen, und nicht ein „irreguläres“.

Diese Reaktionsweise ist von den Befragten gut gemeint, verdirbt aber die Ergebnis. Das muss der Forscher wissen und darauf irgendwie reagieren.

Belson hat deshalb zuerst nach dem üblichen Verhalten gefragt, um es sozusagen aus dem Wege zu räumen, bevor nach dem letzten Lesevorgang gefragt wurde.

Ad 2.) Objekt - Identifikation

Die korrekte Identifikation von Publikationen ist eine der entscheidenden Aufgaben der Leserschaftsforschung. Wie sowohl die akademische als auch die praxisbezogene Forschung zum Thema Recognition immer wieder gezeigt hat, sind die Antworten in dem Masse korrekter, je mehr der benutzte Stimulus im Interview dem ursprünglichen Stimulus entspricht.

Wie es Politz in seiner prägnanten Weise formuliert hat: „ ... das grundlegende Gesetz psychologischer Messung: die für Recognition benutzte Einheit muss der wahrgenommenen Einheit nahekommen“ (Politz 1953).

Belson ließ mehrere der jüngsten Ausgaben jeder der behandelten Publikationen den Befragten vorlegen; diese wurden gebeten, sie durchzusehen, um festzustellen, ob sie etwas in ihnen wiedererkannten. Falls eine Befragte sagte, dass sie sich eine bestimmte Publikation niemals ansehen würde, wurde die Anweisung verändert: „Schauen Sie, ob es irgendetwas gibt, an dem sie besonders interessiert sind“.

Belson betonte: „Es ist für das Verständnis entscheidend, dass dieser Vorgang nicht dazu gedacht war, den Nachweis des Lesens zu gewinnen, sondern ein-

fach der befragten Person ganz deutlich zu machen, zu welcher Publikation gefragt wurde.“

Es sollte vermerkt werden, dass Politz mit der Frage nach dem „Interesse“ an den Inhalten begann, was ein noch vorsichtigerer Ansatz zu sein scheint.

Ad 3.) Zeitidentifikation

Der zweite wesentliche Teil der üblichen Verfahrensweise, um Leserschaftsdaten bezogen auf spezifische Ausgaben ermitteln können, – ???

und, indem man mehrere einbezog, hätte man Durchschnitte für sie gewinnen können.

Aber die I.P.A. - Methode basiert, wie bekannt, auf der Annahme, dass die Anzahl der Leser von irgendwelchen Ausgaben innerhalb eines Erscheinungsintervalls gleich der durchschnittlichen Leserschaft von einer (durchschnittlichen) Aussage während deren Lebensdauer ist. (Dieses Konzept wird seit Beginn im Jahr 1954 von der Arbeitsgemeinschaft Leser-/ Mediaanalyse und in den meisten anderen Leseranalysen verwendet.

(Man muss immer daran denken, dass dies nur unter der Bedingung des „*ceteris paribus*“ gilt; und es hat sich immer wieder gezeigt, dass *ceteris* nicht notwendigerweise *paribus* ist.)

Deshalb war es eines der Ziele der Belson'schen Intensivuntersuchung, soviel wie möglich darüber herauszufinden, wann der letzte Lesevorgang stattfand, so dass diese Information mit der Befragtenantwort zur I.P.A.-Abfrage verglichen werden konnte.

Aufgrund des Studiums einer Reihe von Büchern über Kognition und Gedächtnis habe ich den Eindruck gewonnen, dass ein Gedächtnis für die Zeit nicht existiert, sicherlich nicht in der exakt funktionierenden Art, die man benötigt, um unsere Fragen zu beantworten. Einen vagen Eindruck von Distanzen, je; auch in qualitativer Weise im Sinne von „kürzlich“, „eine Weile zurück“, „lange her“ (s. a. Gerloff 1986). Statt dessen muss der Mensch ein Ereignis zu anderen Ereignissen in Beziehung setzen, um auf diese Weise die Vergangenheit zu rekonstruieren (z. B. Brown und Mitautoren 1976; mit einem Hinweis auf Bartlett 1932; Anmerkung anno 2004: Man hätte sich auf William Stern und das Jahr 1902 beziehen können und sollen).

Folgerichtig wurden die Befragten in den intensiven Interviews darum gebeten, Details des letzten Lesevorganges anzugeben: Ort, Tageszeit, was sonst vorgeing, was sie sich in der betreffenden Publikation angesehen und wie es dazu kam, dass sie sich diese angesehen hatten.

Belson betonte in seinem Bericht: „Diese Details wurden dann bei dem Versuch benutzt, den Zeitpunkt des Vorganges festzulegen, wobei das Prinzip war, eher die Umstände eines Lesevorganges zeitlich festzulegen, als das Lesen selbst.“

Alles zusammen genommen, dürfte diese Prozedur sehr weit gehen, um den Befragten eine Hilfestellung für das (positive oder negative) Wiedererkennen einer speziellen Publikation sowie für die Rekonstruktion des letzten Lesevorganges zu liefern. Wenn man es jedoch streng betrachtet, dann reicht es nicht, um es als ein „Validierungsverfahren“ zu bezeichnen. Denn man verlässt sich immer noch auf das menschliche Gedächtnis. Es ist nicht bekannt, was tatsächlich geschah, da die Untersuchung kein kontrolliertes Experiment war.

Dies darf nicht als eine Kritik des Belson'schen Verfahrens betrachtet werden, und es mindert seine Bedeutung in keiner Weise. Es ist vielmehr die triviale Feststellung einer Wahrheit, die allzu leicht vergessen oder übersehen wird.

Belson's Resultate

In Tabelle 1 werden die grundlegenden Resultate gezeigt. Absolut betrachtet, sind die Abweichungen der I.P.A.-Leserschaftszahlen von denen der intensiven Interviews nicht sehr groß: Im Schnitt ungefähr 4 %.

Natürlich gibt es einige größere Differenzen und zwei besonders auffallende Diskrepanzen: für die Publikationen Nr. 25 und 26.

Wenn man diese Abweichungen jedoch vom Standpunkt der einzelnen Publikationen betrachtet, dann sind die relativen Größenordnungen beunruhigend: ziemlich viele sind im Bereich von 20 % und 30 %, und drei um die 70 %.

Durch Überkreuzauszählungen der Antworten zu den beiden Ermittlungsmethoden zeigte Belson die prozentuale Übereinstimmung der beiden (s. Tabelle 1, letzte Kolonne). Diese Übereinstimmung betrug um die 85 % im Schnitt, was wiederum nicht so schlimm erscheint; aber wenn bei einigen Zeitschriften die Diskrepanzen nur 4 % bis 5 % betragen, bei anderen aber 20 % bis 26 %, dann sind solche Ungleichheiten beunruhigend.

Tabelle 1: Belson´s hauptsächliche Ergebnisse				
	"Intensive" (Recognition)	Differenzen		Durch Überkreuz- auswertung festgestellte Übereinstimmung zwischen Intensiven + I.P.A.-Interviews
		abs.	rel.*	
	%	%	%	%
Tageszeitungen				
1) Daily Mirror	48	0	0	82
2) Evening News	40	-3	-8	80
3) Daily Express	30	-1	-3	84
4) Daily Sketch	19	-1	-5	92
5) Daily Telegraph	13	1	8	95
6) Daily Herald	11	1	9	96
7) The Times	5	0	0	96
Sonntagszeitungen				
8) News of the World	51	-2	-3	82
9) Sunday Pictorial	47	-2	-3	84
10) The People	41	-5	-11	85
11) Sunday Express	33	-7	-20	84
12) Sunday Dispatch	19	-6	-29	92
13) Sunday Times	17	-7	-39	87
14) Observer	14	-5	-35	93
Wochenzeitschriften				
15) Radio Times	59	-4	-6	89
16) TV Times	46	-3	-6	84
17) Woman	42	-4	-10	80
18) Woman´s Own	41	-7	-18	78
19) Reveille	24	-6	-24	86

*) die relativen Unterschiede beruhen auf den absoluten Zahlen und nicht auf gerundeten Prozenten
– außer bei Tageszeitungen, für die keine absoluten Zahlen veröffentlicht wurden.

	"Intensive" (Recognition)	Differenzen		Übereinstimmung
		abs.	rel.*	
Monatszeitschriften				
20) Reader's Digest	26	-3	-15	74
21) Do it Yourself	19	-7	-37	86
22) Practical Householder	13	-3	-23	86
23) Vogue	12	-2	-16	86
24) Good Housekeeping	8	1	13	90
25) Ideal Home	22	-15	-68	80
26) Woman and Home	21	-15	-71	74
27) True Story	5	-1	-25	93
28) True Romances	7	-5	-71	91

Verallgemeinerungen

In seinem Buch stellt Belson fest, es existierten frühere Hinweise darauf, dass die monatlichen Zeitschriften für Irrtümer am anfälligsten sind; und die vorhergehende Forschung hätte diese Annahme unterstützt. In seiner Studie sind die Diskrepanzen in der Tat in dieser Kategorie grösser als bei den Sonntagszeitungen und wöchentlichen Zeitschriften.

Die Tageszeitungen zeigten ein durchaus anderes Bild: die Resultate von Recall und Recognition liegen nahe beieinander.

Bei der Betrachtung der beiden wöchentlichen Publikationsgruppen bemerkte Belson für die Sonntagszeitungen, dass „wir eine Fehlerzunahme finden, wenn wir in der Größenordnung der I.P.A.- Leserschaftszahlen abwärts gehen“. Und er fuhr fort, bezogen auf die wöchentlichen Zeitschriften: „Wenn sich dieser Trend über den Bereich der fünf Zeitungen hinaus fortschreibt, dann sind die Folgerungen ernsthafter Natur, weil die Zeitungen, die nicht untersucht wurden, kleinere Leserschaftszahlen haben, als diese, – viele von ihnen sehr viel geringere.“

Einen derartigen Trend sah er in der Kategorie der Monatszeitschriften nicht, sondern: „Was wir in der Tat haben, ist die Erscheinung von großer Unterschiedlichkeit (ohne erkennbares Muster) im Fehlerniveau ...“. Dies könnte vom Stichprobenfehler herkommen, fuhr er fort, und „es könnte daher kommen, dass das Erinnern für die Monatspublikationen schwieriger ist“.

Die Mediaforscher, ich selbst eingeschlossen, haben „gelernt“, in Kategorien wie den vier von Belson verwendeten zu denken (s. Tabelle 1). Solche Kategorisierungen werden üblicherweise in zwei Dimensionen vorgenommen:

- 1.) Grundsätzlich eine Unterscheidung nach dem Erscheinungsintervall, d. h. nach täglich, wöchentlich, 14-täglich, monatlich, zweimonatlich erscheinenden Publikationen.
- 2.) „Qualitative Unterscheidungen wie: Zeitungen, Wochenzeitschriften, Sonntagszeitungen, allgemeine Zeitschriften, Frauenzeitschriften, Special - Interest - Zeitschriften etc..“

Solche Einteilungen erscheinen plausibel: Die erste steht in Beziehung zum Zeitintervall zwischen Lesen und Befragung; die zweite basiert auf der Vorstellung, dass derartige Charakterisierungen nicht nur eine Bedeutung für Verleger und Journalisten haben, sondern auch für Leser und Werbungtreibende.

Und wenn man die Durchschnitte für die vier Belson'schen Gruppen rechnet, wie in Tabelle 2 gezeigt, dann scheinen die Resultate die Logik dieser Unterscheidungen zu unterstützen: Wenn nicht in vier Kategorien, dann mögen sie wenigstens, wie Gallien, „divisa est in partes tres“ (Caesar):

- 1.) Die Tageszeitungen zeigen den geringsten Fehler;
- 2.) dann kommen die zwei Gruppen von Publikationen mit wöchentlichem Erscheinungsintervall;
- 3.) und die monatlichen Zeitschriften sind am schlimmsten dran.

Tabelle 2: Durchschnittliche Abweichungen für die vier Zeitschriften Kategorien

	Recognition	Aided Recall	<u>Differenzen</u> abs.	rel.
	%	%	%	%
Tageszeitungen	23.7	23.2	0.5	2
Sonntagszeitungen	31.7	26.9	4.8	15
Wochenzeitschriften	42.4	37.6	4.8	11
Monatszeitschriften	9.2	14.8	5.6	38

Dies passt zu der allgemeinen Vorstellung, dass das Gedächtnis mit dem Ablauf der Zeit progressiv schlechter wird.

Selbst dann, wenn man die beiden extremen Abweichungen aus der Kategorie der monatlichen Zeitschriften herausnimmt (Publikationen Nr. 25 und 26), ist der Durchschnittsfehler dort der höchste: 22 %.

Folgerichtig wurden Belson's Resultate von anderen Forschern zitiert, um die Vorstellung zu unterstützen, dass die monatlichen Publikationen das Problemkind der Leserschaftsforschung seien (Smith 1988).

Warum eine Reanalyse?

Die Belson'sche Analyse basiert auf dem Standpunkt der psychologischen Forschung und des Testens: Man befasst sich mit Individuen, testet Individuen, versucht Diskrepanzen zwischen verschiedenen Antworten/Reaktionen der Individuen herauszufinden. Und dann addiert man, rechnet Durchschnitte und versucht, die Ergebnisse zu verallgemeinern.

Beim psychologischen Testen von Individuen ist das so, wie es sein sollte. Wenn mein Intelligenz-Quotient oder meine Persönlichkeit getestet wird, möchte ich korrekte Resultate für meine Person; das gleiche gilt, wenn ich einem Lügendetektortest unterworfen werde (zumindest dann, wenn ich die Wahrheit sage). Es wäre nicht akzeptabel für mich, wenn solch ein Test für eine Gruppe statistisch korrekt ist, aber für jedes Individuum (wie mich) nur mit einem Fehlerbereich von 20-30 % „stimmt“. Mein Eindruck ist, dass die Psychologen ein Problem haben mögen, eine derartige Forderung zu erfüllen, aber das überlasse ich ihnen.

Was unsere Art von Forschung anbelangt, so sind wir nicht an Individuen interessiert, sondern an Gruppen. Und es ist nicht unsere Aufgabe, Personen zu testen, sondern Produkte, Packungen, Parfums, – und hier: Publikationen.

Das heißt, dass wir ein „korrektes“ Ergebnis erzielen möchten, z. B. bezüglich der Präferenzen für ein spezielles Produkt, für eine Packung oder ein Parfum. Und wir wollen ein „korrektes“ Resultat bezüglich der Reichweiten für eine bestimmte Anzahl von Publikationen.

Wenn wir diesem Ziel relativ nahe kommen, so sollte es uns erst an zweiter oder dritter Stelle interessieren, dass manche Befragte Antworten geben, die sie als „Leser“ ausweisen, obwohl sie es in der Tat nicht sind: und dass ein anderer Teil der Stichprobe sich in der entgegengesetzten Richtung irrt.

Politz benutzte das Beispiel einer Autofahrer-Umfrage, um darauf hinzuweisen, dass die individuellen Antworten nicht das sind, was wir brauchen: „Wie jeder Interviewer weiß, kann der Unterschied zwischen einem 'Ja' und einem 'Nein' winzig sein.“ Dann wies er auf den Vorteil einer Umfrage hin, die groß genug

ist, um eine „Randomness“ von Schwankungen in den Antworten zu produzieren. Obwohl der einzelne Autofahrer ein beinahe bedeutungsloses „Ja“ oder „Nein“ lieferte, konnte man zu eindeutigen Schlussfolgerungen kommen (Politz 1957).

Wendt hatte vor einiger Zeit die Hypothese formuliert, dass die Antworten zu jeglicher Frage einen „probabilistischen“ Charakter haben könnten (Wendt 1977). Das von Wenzel und Speetzen (1985) vorgetragene Konzept von „random errors“ in Media-Untersuchungen geht in dieselbe Richtung. Brown schrieb, dass Recognition wohl kaum eine „Ja-Nein“ - Angelegenheit sei, sondern vielmehr ein Kontinuum (Brown 1974). Dazu passt die Darstellung von Foppa von „probabilistischen Lernmodellen“; er zitiert als ersten, der mit Lernwahrscheinlichkeiten operiert, L. L. Thurstone (1930).

Nach dem, was ich in einem Buch über Neurologie gelesen habe, könnten neurale Zustände sich auf einem Kontinuum bewegen und die neurale Kommunikation probabilistisch sein (Oeser und Seitelberger 1988).

Ein anderer Ausgangspunkt für diese Reanalyse war der philosophisch unterstützte Gedanke (Popper 1980), dass eine wirkliche Validierung von Methoden und Resultaten nicht möglich ist. Das ist so nicht wegen unseres noch immer stark begrenzten Wissens, sondern grundsätzlich. Hier stimme ich wieder mit Wendt (1984) überein.

Ich glaube, dass Belson's „intensive“ Interview-Methode (wie auch der mit Alfred Politz assoziierte Ansatz) ein hohes Maß an Konstrukt-Validität hat. Deshalb würde ich die Resultate, die von einer solchen Verfahrensweise produziert werden, als Messlatte für Resultate benutzen, die auf eine mehr oberflächliche Art und Weise erhoben wurden. Aber man wäre nicht in der Lage nachzuweisen, dass sie ein absolut gesehen wahres Bild produzieren. Eine noch bessere Methode wäre nötig, um das zu überprüfen ... , – und wir enden mit einem infiniten Regress.

Was wir tun können, und das ist der dritte Ausgangspunkt, ist: Die Resultate von verschiedenen Methoden oder Tests systematisch zu vergleichen. Indem wir das tun, können wir Strukturen oder Muster finden. Diese können uns helfen, besser zu verstehen, was verschiedene Methoden messen; und daraufhin dürfen wir hoffen, die Leserschaftsforschungsmethodologie verbessern zu können. Die Formulierung, „dass wir nur ein Gummiband am anderen messen,“ (Scheler 1980), wenn wir vergleichen, ist sehr anschaulich, phänomenologisch korrekt, vermittelt m. E. jedoch den Eindruck einer „Beliebigkeit“, über die man hinausgehen sollte und kann.

Messung von „Reflexionen“

Diese Reanalyse beruht auf der Voraussetzung, dass zwei verschiedene Methoden zwei verschiedene Reflexionen einer unbekanntes Realität auf unterschiedliche Ebenen produzieren, – möglicherweise auf gekrümmte (s. a. Eckland 1985). Aber indem wir die Resultate zweier Methoden in Beziehung setzen, können wir etwas über beide lernen, und auch über einen Ausschnitt der Realität. (Wie in der Navigation wäre eine dritte oder vierte „Peilung“ offensichtlich sehr hilfreich.)

Vor ein paar Jahren fanden wir heraus, dass zwei verschiedene Messwerte der Erinnerung an Marken sehr hoch miteinander korrelieren (Schaefer 1980):

- 1.) Die spontane Markenbekanntheit („Aktualität“) und die gestützte Bekanntheit korrelieren in einer nicht linearen Art und Weise (Exponentialfunktion);
- 2.) je weiter die gestützte Bekanntheit abfällt, umso geringer ist der Anteil derjenigen, der die Marke spontan nennt.

Es erschien wahrscheinlich, dass eine solche Beziehung auch zwischen den Recall- und Recognition-Messwerten der Belson'schen Studie existiert.

Die Regressionsanalyse halten wir für die angemessene Methode, um etwas über Art und Enge des Zusammenhanges herauszufinden; und sie weist auf Abweichungen von den Erwartungswerten hin (Eine sehr nützliche Diskussion der Methode, ihrer Vorzüge und Grenzen findet man z. B. bei Ehrenberg, 1986).

Wir haben¹ ein EDV-Programm mit acht verschiedenen Regressionsformeln angewendet, um die geeignetste für die Beschreibung der Beziehung zwischen den zwei Datensätzen zu finden. Die Resultate wurden dann geeicht.

Wie in jeder solchen Analyse muss man entscheiden, welche Werte als „x“ und welche als „y“ anzusehen sind. Vom Standpunkt der Konstrukt-Validität liefern die Recognition-Werte eher die volle Leserschaft als die Aided Recall-Daten; deswegen wurden letztere als „x“- Werte eingesetzt.

Alle Publikationen

Um ganz generell herauszufinden, ob die beiden Methoden grundsätzlich denselben Sachverhalt messen oder nicht, wurden die Daten für alle 28 intensiv untersuchten Publikationen für die erste Regressionsanalyse verwendet.

¹ bei der Reanalyse 1989

Die beste Anpassung war eine Kurve wie in Graphik 1 reproduziert. Die Korrelation ist mit $r = .974$ sehr hoch. Das bedeutet, dass in ganz allgemeiner Weise die beiden Methoden die gleiche (uns unbekannt) Realität reflektieren. Hohe Korrelationen sind schön, aber Abweichungen sind von größerem Interesse für den Forscher, weil sie eine größere intellektuelle Herausforderung bedeuten, zu vermehrter Einsicht führen und evtl. zu neuen Forschungsprojekten.

Und es gibt Abweichungen. Ich zähle nur jene auf, die 4 % oder mehr von den Erwartungswerten abweichen (Tabelle 3).

Tabelle 3: Recall gegenüber Erwartungswerten			
		<u>Aided Recall</u>	
<u>Nr.</u>	<u>Name</u>	<u>Ergebnisse</u> %	<u>Erwartungswerte</u> %
<u>Höher</u>			
3	Daily Express	29	24
4	Daily Sketch	18	14
5	Daily Telegraph	14	9
6	Daily Herald	12	8
<u>Niedriger</u>			
15	Radio Times	55	60
25	Ideal Home	7	17
26	Woman and Home	6	16

Die erste Gruppe besteht nur aus Tageszeitungen. Mit der Ausnahme der „Times“ sind bei allen die Recall-Werte höher, als aufgrund dieser Regressionsanalyse zu erwarten war. Wie die reproduzierten Zahlen aus Belson's Buch zeigen, lagen die I.P.A.- und die intensiven Werte ziemlich nahe beieinander. Deshalb müssen wir daraus schließen, dass die Tageszeitungen sich etwas anders verhalten als der Rest, und dass sie vielleicht separat zu betrachten sind.

Die zweite Gruppe enthält einen unerwarteten Fall: Die „Radio-Times“; dort gab es einen Unterschied zwischen I.P.A. und intensiv. Aber es gab noch eine andere Publikation in dieser Kategorie mit der gleichen Größenordnung eines

„Fehlers“, sowie zwei mit sogar größeren Abweichungen, – und sie alle zeigen sich hier nicht. Auch die Wochenzeitschriften erscheinen nicht als Abweichungen. Offenbar gibt es eine Ergebnisstruktur, die sich in der einfachen Gegenüberstellung nicht offenbart, wohl aber in dieser Regressionsanalyse.

Die beiden schlimmsten Fälle, beide aus der Kategorie der Monatszeitschriften, tauchen natürlich auch auf. Wir werden uns mit ihnen noch beschäftigen.

Tageszeitungen

Wie schon angedeutet, könnten die Tageszeitungen eine Kategorie für sich sein. Dies entspricht selbstverständlich den Erwartungen: Das tägliche Erscheinen hinterlässt einen solchen Eindruck im Gedächtnis, dass man sie nicht leicht und schnell vergisst; das Zeitintervall zwischen Lesen und Interview ist kurz.

Mehr aus Spaß haben wir auch mit diesen Regressionsanalysen durchgeführt. Vier verschiedene Formeln lieferten erklärte Varianzen von mehr als 99 %, drei von ihnen waren praktisch gerade Linien, die vierte eine sehr leichte Kurve (Graphik 2).

Demnach sind Tageszeitungen tatsächlich eine Kategorie für sich.

Sonntagszeitungen und Wochenzeitschriften

Wenn wir der Hypothese folgen, dass die Qualität unserer Erinnerung an Lesevorgänge unter anderem eine Funktion des Zeitintervalls zwischen Lesen und Interview ist, dann sollten die zwei Kategorien der wöchentlichen Publikationen dieselbe Art von Beziehungen aufweisen. Deshalb habe ich sie in einer Regressionsanalyse zusammengefasst; und sie sind alle nahe an der resultierenden Kurve, mit einer erklärten Varianz von 99,4 % (Graphik 3). Sie gehören also zusammen.

Die Kurve unterstützt Belson's Feststellung: Mit abfallenden Recognitionwerten vermindern sich progressiv die Chancen für Publikationen, bei der I.P.A.- (= Aided Recall) -Verfahrensweise genannt zu werden (Tabelle 4):

**Tabelle 4: Alle wöchentlichen Publikationen
(Sonntagszeitungen plus Wochenzeitschriften)**

<u>Recognition</u>	<u>Aided Recall</u> <u>Erwartungswerte</u>
%	%
60	59.5
50	46.6
40	34.6
30	23.6
20	13.7
10	5.4
5	2.2
4	1.6
3	1.1
2	0.6
1	0.25

Monatszeitschriften

Wie wir in der ursprünglichen Analyse und in der ersten Regressionsanalyse gesehen haben, gibt es zwei Publikationen in dieser Gruppe, bei denen die I.P.A.-Messung von der intensiven weit entfernt war. Es ist deshalb kein Wunder, dass eine Regressionsanalyse mit dieser Zeitschriften-Kategorie auch nicht entfernt eine so hohe Korrelation aufweist wie für die anderen Kategorien. Die beste Anpassung liefert eine Parabel-Kurve (Graphik 4), aber mit nur 46,8 % erklärter Varianz.

Offensichtlich passen die Publikationen 25 und 26 nicht in das allgemeine Muster. Leider sagt die Belson'sche Studie nicht, warum oder in welcher Art und Weise diese beiden so atypisch waren. Insbesondere für eine „qualitative“ Untersuchung ist bemerkenswert, dass nicht der Versuch unternommen wurde,

die beiden Ausreißer zu beschreiben und einen Erklärungsversuch zu unternehmen.

Die Tatsache, dass diese zwei Monatszeitschriften von allen Normen abweichen, erscheint als ausreichender Grund, um Regressionsanalysen ohne sie durchzuführen. Ihren Ausschluss betrachte ich nicht als eine Anwendung von Maier's Gesetz: „Wenn die Fakten nicht mit der Theorie übereinstimmen, müssen sie beiseite geschafft werden“ (Maier 1960).

Das Resultat ist dramatisch: Die erklärte Varianz springt auf 88,9 % (Graphik 5). Daraus dürfen wir schließen, dass die Monatszeitschriften sich im Prinzip wie andere verhalten, – mit zwei Ausnahmen, die nicht erklärt werden können.

Die generelle Beziehung

Wenn man alle Publikationen außer Nr. 25 und 26 in die Analyse einbezieht, dann steigt die erklärte Varianz von 94,9 % auf 97,8 %, wie zu erwarten war. Aber alle Tageszeitungen befinden sich oberhalb der Erwartungswerte. Dies bestätigt endgültig die Hypothese, dass sie eine Kategorie für sich sind und für sich betrachtet werden sollten.

Deshalb haben wir in der letzten Regressionsanalyse die beiden Kategorien der wöchentlich erscheinenden Zeitschriften und die Monatszeitschriften (ohne 25 + 26) untersucht, aber die Tageszeitungen ausgeschlossen (Graphik 6).

Wieder befindet sich die erklärte Varianz in der extremen Höhe von 99 %; und die beste Anpassung ist eine Kurve. Die Erwartungswerte sind in Tabelle 5 dargestellt. Sie zeigen das schon früher beobachtete Muster: Die Chancen, von der I.P.A.- (Aided Recall-)Methode richtig repräsentiert zu werden, nehmen progressiv mit dem Umfang der Leserschaft ab.

Zusammenfassung und Schlussfolgerungen

Die Resultate, die in dieser Analyse vorgetragen werden, enthalten sowohl eine gute als auch eine schlechte Nachricht.

Die gute Nachricht ist, dass selbst das relativ oberflächliche Verfahren der damaligen I.P.A. - Umfragen, Leser von sehr vielen Publikationen einzusammeln, Ergebnisse produzierte, die nicht so sehr weit von denen weg sind, die in einer intensiven Befragung mit wenigen Publikationen erzielt wurden. Dies liefert eine allgemeine Erklärung für die Tatsache, dass so viele Leserschaftsexperimente

mit mehr oder weniger unterschiedlichen Methoden ähnlich aussehende Resultate zu liefern scheinen.

Es unterstützt auch den Standpunkt eines der „Elder Statesmen“ der Leserschaftsforschung. Dr. Ernst Braunschweig, der einmal gesagt hat: „Gott muss auf der Seite der Marktforscher sein, – anders können manche Dinge nicht erklärt werden.“

Die schlechte Nachricht wird angekündigt durch die nicht-lineare Beziehung: Je kleiner eine Publikation ist, desto geringer ist ihre Chance bei einer oberflächlichen Methode.

Die Tageszeitungen werden von Unterschieden der Methode am wenigsten beeinflusst. Dies mag auf das kurze Zeitintervall zwischen Lesen und Interview zurückgeführt werden: Das Gedächtnis hat noch nicht gelitten. Unterstützende Faktoren könnten die größere „Präsenz“ dieser Publikationen und das gewohnheitsmäßige häufige Lesen sein.

Was die anderen Kategorien anbelangt, so unterstützen die Ergebnisse dieser Analyse zwei allgemeine Annahmen nicht:

- 1.) Dass die übliche Kategorisierung der Publikationen nach dem Erscheinungsintervall wichtig und richtig sei;
- 2.) dass die Monatszeitschriften am meisten unter Fehlerinnerungen leiden, und zwar wegen des größeren Zeitintervalls zwischen Lesen und Interview.

Stattdessen finden wir, dass die Unterschiede eine direkte, nicht lineare Funktion des Umfangs der Leserschaft ist; und die Monatszeitschriften, die in der Belson'schen Studie behandelt wurden, hatten kleine Leserschaften.

Die hier vorgestellte Art der Analyse liefert Auskünfte über „Erwartungswerte“ (graphisch: Gerade oder Kurve), mit deren Hilfe Publikationen identifiziert werden können, die sich nicht normal „verhalten“. Diese sind es dann, die eine nähere Untersuchung verdienen. Leider hilft die Belson'sche Studie hier nicht weiter.

Als eine generelle Erkenntnis kann man formulieren, dass offenbar kleinere Publikationen keine faire Chance haben, bei oberflächlicher Behandlung vieler Publikationen korrekt dargestellt zu werden (Tabelle 5).

Tabelle 5:
Alle wöchentlichen und monatlichen Publikationen
 (außer 25 + 26)

Recognition %	Aided Recall
	Erwartungswerte %
60	59
50	46
40	35
30	24
20	15
10	7
5	3.2
4	2.5
3	1.8
2	1.1
1	0.4

Oder in der einer noch allgemeineren Art und Weise: Bei Recall mit beschränkter Hilfestellung im Vergleich zu Recognition hat ein Objekt eine relativ zu kleine Chance, genannt zu werden, wenn es von geringer „Prominenz/Prägnanz/Aktualität“ ist. Die Recall - Aussichten steigen mit wachsender „Prominenz/ Prägnanz/ Aktualität“.

Danksagungen

Der Autor dankt sehr den Herren Prof. Andrew Ehrenberg, Drs. Jan Noordhoff, Friedrich Wendt und Felix Schaefer für die kritische Durchsicht der englischen Version dieses Aufsatzes und für nützliche Korrekturen und Zusätze. Dies ist die etwas erweiterte und revidierte deutsche Fassung eines Aufsatzes, der in der Februar-Ausgabe 1989 von „Marketing und Research Today“ (ehemals „European Research“) veröffentlicht wurde. Redaktion und ESOMAR wird für die freundliche Genehmigung zur Wiedergabe gedankt.

Der Fairness halber ist darauf hinzuweisen, dass gewisse Meinungsverschiedenheiten zwischen Ehrenberg und dem Autor ungelöst blieben. Eine Reihe von Kolleginnen und Kollegen haben freundlicherweise das deutsche Manuskript durchgesehen und kommentiert; insbesondere den Herren Dr. Klaus Peter Landgrebe und Thomas Kiss wird für ihre zahlreichen und gut begründeten Verbesserungsvorschläge gedankt. Die Verantwortung für das Endprodukt und ggf. dessen Mängel verbleibt beim Autor.

Quellenhinweise:

- Bartlett**, F.C.: „Remembering: A Study in Experimental and Social Psychology“, Cambridge University Press; Cambridge, England, 1932.
- Belson**, William A.: „Studies in Readership“, published on behalf of the Institute of Practitioners in Advertising by Business Publications Ltd., London, 1962.
- Brown**, John: „Recognition Assessed by Ratings and Ranking“, British Journal of Psychology, 65, 1, pp 13-22, 1974.
- Brown**, John (ed.): „Recall and Recognition“, Wiley, New York / London, 1976.
- Caesar**, G. J.: „De bello Gallico“, Reclam, Stuttgart, 1980.
- Ehrenberg**, A. S. C.: „Statistik oder der Umgang mit Zahlen“, VCH, Weinheim, 1986.
- Ekeland**, Ivar: „Das Vorhersehbare und das Unvorhersehbare“, Harnack-Verlag, München, 1985.
- Foppa**, Klaus: „Probabilistische Lernmodelle“, in R. Bergius (herg.): „Handbuch der Psychologie“, Bd. 1, 2. Halbband, Hogrefe, Göttingen, 1965.
- Gerloff**, Otfried: „Materialsammlung zur Zeitidentifikation“ (unveröff. Manuskript), Arbeitsgemeinschaft Mediaanalyse e.V., Frankfurt a. M., 1986.
- Maier**, N. R. F.: „Maier's Law“; The American Psychologist, Vol. 15, Nr. 3, 1960.
- Noordhoff**, J., and M. **Brown**: „The AG.MA. Media Research Model: Philosophy and Consequences“, Media-Micro-Census GmbH, Frankfurt a. M., 1983.
- Oeser**, Erhard, und Franz **Seitelberger**: „Gehirn, Bewusstsein und Erkenntnis“, Wiss. Buchgesellschaft, Darmstadt, 1988.
- Opfer**, G. und D. **Müller-Veeh**: „Validierungen im Mediabereich: Zeitschriften“, Arbeitsgemeinschaft Media-Analyse e. V., Band 9, Frankfurt, 1983.
- Politz**, Media Studies: „A Study of Four Media“, LIFE, NY, 1953.
- Politz**, Alfred: „Science and Truth in Marketing Research“, Harvard Business Review, Jan.-Feb. 1957.
- Popper**, Karl R.: „Objektive Erkenntnis“, Hoffmann und Campe, Hamburg, 1974.
- Schaefer**-Marktforschung: „Werbetrend“, Hamburg, 1980-87.
- Riedl**, Rupert: „Biologie der Erkenntnis“, Parey, Hamburg, 1980.
- Scheler**, H.-E.: „Technische, organisatorische und methodische Probleme der MA“, Referat auf dem BDW-Symposium „Media-Analysen zwischen Konvention und Wirklichkeit“, 13. November 1980.
- Smith**, Alan: „Gelesen oder nicht gelesen - das ist die Frage“, Viertel-Jahreshefte für Media und Werbewirkung, 1/1988.
- Stern**, William: „
- Stern**, William: „
- Wendt**, Friedrich, in: „Konsumdynamik - Eine Stufe zu den Marktmechanismen im Kauf- und Informationsverhalten“, Dt. Supplement-Verlag, Nürnberg, 1977.
- Wendt**, I. und F.: „Vom Leser pro Nummer zur Nutzungswahrscheinlichkeit“, Media-Micro-Census GmbH, Frankfurt a. M., 1979.
- Wendt**, Friedrich, in Harry Henry (ed): „Readership Research“, Elsevier Science Publ., Amsterdam/ New York, 1984.
- Wenzel**, W. und R. **Speetzen**: „Debugging Random Errors from Media Analysis Data - A New Type of Validation“, in H. Henry (ed.): „Readership Research“, North Holland, Amsterdam, 1985.