

Brockhoff, Klaus

Working Paper — Digitized Version

Standards für den Einsatz von Rechnern als Experimentierinstrumente

Manuskripte aus den Instituten für Betriebswirtschaftslehre der Universität Kiel, No. 90

Provided in Cooperation with:

Christian-Albrechts-University of Kiel, Institute of Business Administration

Suggested Citation: Brockhoff, Klaus (1981) : Standards für den Einsatz von Rechnern als Experimentierinstrumente, Manuskripte aus den Instituten für Betriebswirtschaftslehre der Universität Kiel, No. 90, Universität Kiel, Institut für Betriebswirtschaftslehre, Kiel

This Version is available at:

<https://hdl.handle.net/10419/190923>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Nr.90

Standards für den Einsatz
von Rechnern als
Experimentierinstrumente

Klaus Brockhoff

Standards für den Einsatz von Rechnern als
Experimentierinstrumente *

Klaus Brockhoff **

* In diesen Beitrag sind Erfahrungen mehrjähriger Projektarbeit eingeflossen, die auch meine Kollegen, Mitarbeiter und Versuchspersonen vermittelt haben. Ihnen allen sei deshalb gedankt.

** Dr. Klaus Brockhoff, o. Professor für Betriebswirtschaftslehre der Universität Kiel, Olshausenstr. 40-60, 2300 Kiel

1. Rechner als Experimentiereinrichtungen

Dialogfähige Rechenanlagen und der Zugang zu ihnen erfahren zur Zeit eine schnell wachsende Verbreitung. Dafür können verschiedene wirtschaftliche Gründe sprechen: erstens die Realisierung der Potentiale einer den spezifischen relativen Leistungsvorteilen entsprechenden Arbeitsteilung zwischen Mensch und Rechner¹⁾; zweitens die absolut wachsenden Leistungsanforderungen.

Die Anlagen bieten u.a. die Möglichkeit, verschiedene Merkmale der über den Rechner geleiteten Benutzeraktivitäten automatisch und vom Benutzer unbemerkt zu erfassen²⁾, d.h. im Extremfall im Sinne einer voll-biotischen Situation³⁾.

Diese Chance unaufdringlicher ("unobstrusive") Beobachtungen kann im Rahmen des routinemäßigen Einsatzes solcher Systeme genutzt werden, um eine dem Benutzerverhalten angepaßte, wirtschaftliche Systemgestaltung zu entwickeln oder um ein Benutzertraining zur besseren Akzeptanz und Ausnutzung von Systempotentialen zu konzipieren. Hierbei können experimentelle oder - realitätsnäher - auch quasi-experimentelle⁴⁾ Veränderungen der Einsatzbedingungen die Erkenntnisgewinnung fördern.

-
- 1) Z.B. Sackman, H., Computers, System Science and Evolving Society. The Challenge of Man-Machine Digital Systems, New York 1967, S. 42f.; Mertens, P., Kress, H., Mensch-Maschinen-Kommunikation als Hilfe bei Entscheidungsvorbereitung und Planung, Zeitschrift für betriebswirtschaftliche Forschung, 22. Jg., 1970, S. 1-21, hier S. 2; Kieser, H.P., Mensch-Maschinen-Kommunikation in Organisationen, Diss. Mannheim 1973; S. 88ff.
 - 2) Vgl. die Vorgehensweise in: Brockhoff, K., Delphi-Prognosen im Computer-Dialog, Tübingen 1979, S. 27ff.
 - 3) Spiegel, B., Werbepsychologische Untersuchungsmethoden, Berlin 1958, S. 45: "Die Versuchsperson weiß nicht um den Zweck des Versuchs, nicht um ihre Aufgabe und auch nicht um ihre Situation als Versuchsperson".
 - 4) Im Sinne von: Campbell, D.T., Stanley, S.C., Experimental and Quasi-Experimental Designs for Research, 10.A., Chicago/Ill. 1973.

Durch experimentelle Beobachtungen, die hier auf das Verhalten von Versuchspersonen gerichtet sind, sollen hypothetische "wenn-dann"-Sätze überprüft werden. Hauptkennzeichen eines Experiments ist, daß der Versuchsleiter dazu kontrolliert Veränderungen vermuteter Einflußvariablen vornimmt und sich gleichzeitig bemüht, die Wirkung potentieller Störvariablen auszuschließen. Die Reaktion der Versuchspersonen auf die Veränderung von Einflußvariablen ist zu beobachten und in Bezug auf die Hypothese über dieses Verhalten auszuwerten⁵⁾.

Zwei Möglichkeiten des Rechnereinsatzes im Experiment sind zu unterscheiden. Im ersten Fall soll der Rechner eine Hilfe zur Lösung einer Sachaufgabe bieten. Er soll dabei nicht stören. Sein Störpotential muß also konstant gehalten und kontrolliert werden, während Aufgabenstellung und Versuchspersonen variieren können. Im zweiten Fall soll das Störpotential des Rechners bei der Lösung bestimmter Sachaufgaben untersucht werden. Dabei wird man die Aufgabenstellung konstant halten und kontrollieren, die rechnerspezifischen Äußerungen gegenüber Versuchspersonen und die Auswahl dieser Versuchspersonen aber variieren.

Im folgenden gehen wir vom ersten Fall aus; nur an einigen markanten Stellen verweisen wir auf die veränderten Fragen und Beobachtungen, auf die in Experimenten der zweiten Art zu achten ist.

- 5) Definitionen des Begriffs "Experiment" umfassen regelmäßig die hier genannten positiven Elemente, erfordern aber die vorgenommenen Einschränkungen nicht. So kann z. B. auch die Lösungsmethode für Planungsprobleme experimentell, d.h. durch simulative Anwendung, überprüft werden. Auf der anderen Seite kann das erwartete oder vermutete Verhalten von Personen aufgrund früherer Feststellungen durch Rechner abgebildet und in seinen Folgen, auch den Interaktionsfolgen, simuliert werden: Coleman, J.S., Analysis of Social Structures and Simulation of Social Processes with Electronic Computers, Ed. psychol. Measurement, Vol. 21, 1961, S. 203-213; Emshoff, J.R., A Computer Simulation Model of the Prisoner's Dilemma, Behavioral Science, Vol. 15, 1970, S. 304-317; Groves, P.H., Computer Simulation of Interaction in Decision Making, Behavioral Science, Vol. 15, 1970, S. 277-285; Guetzkow, H., Kotler, P., Schultz, R., Eds, Simulation in Social and Administrative Science, Englewood Cliffs/N.J. 1972; Hart, E.W., Sung, Y.H., Computer and Experimental Simulation of Triad Decision Making, Behavioral Science, Vol. 21, 1976, S. 532-547. Zusammenfassend zum Experiment: Tietz, R., Experimente in den Wirtschaftswissenschaften, Handwörterbuch der Betriebswirtschaft, 4. A., Bd. 1, Stuttgart 1974, Sp. 1351-1363.

Die häufig mangelhafte Kontrolle potentieller Störvariablen und Beschränkungen der Veränderbarkeit der Einflußvariablen durch Erfordernisse des laufenden Geschäftsbetriebes legen es nahe, neben Feldexperimenten auch Laborexperimente durchzuführen. Dieser Weg ist bereits mehrfach besprochen worden, um Erkenntnisse über den Mensch-Maschine-Dialog zu gewinnen, auch wenn keine routinisierten Entscheidungen zu treffen sind⁶⁾. Im Laborexperiment kann der Rechnerinsatz über eine Simulation seiner Einsatzbedingungen im Feld hinausgehen. Die Versuchssituation wird dabei regelmäßig nicht mehr voll-biotisch sein. Sie kann "quasi-biotisch" sein, wenn die Versuchspersonen gegenüber der voll-biotischen Situation Wissen um die Versuchssituation haben

6) Beispielhaft wird dies aufgrund einer Analyse von 331 US-amerikanischen Dissertationen und Master-Arbeiten zwischen 1973 und 1979 deutlich: Ives, B., Hamilton, S., Davis, G.B., A Framework for Research in Computer-Based Management Information Systems, Management Science, Vol. 26, 1980, S. 910-934, bes. Tab. 14, S. 928.

oder sie kann "nicht durchschaubar" sein, wenn sie zusätzlich auch die "eigentliche" Aufgabe kennen ⁷⁾.

Damit sind drei Gestaltungsmöglichkeiten für Experimente hervorzuheben:

- (1) die experimentelle oder quasi-experimentelle Variation ohnehin ablaufender Mensch-Maschine-Dialoge "on the job", d.h. als Feldexperiment; oder der Vergleich der Reaktionen der Versuchspersonen innerhalb solcher Dialogbeziehungen mit den Reaktionen in Kontrollgruppen, denen kein rechnergestützter Dialog zur Verfügung steht;
- (2) die simulative Abbildung und Durchführung realisierter oder für den Feldeinsatz geplanter Mensch-Maschine-Dialoge im Rahmen von Laborexperimenten;
- (3) die Benutzung des Rechners zur modellhaften Abbildung konkreter Phänomene, die im Feldeinsatz nicht durch den Rechner dargestellt werden, zusätzlich zu den unter (2) genannten Möglichkeiten.

Die erste Möglichkeit ist unmittelbar einleuchtend. Man sollte bei ihr aber nicht allein an betriebliche Anwendungen denken. Die Verbreitung von Zwei-Wege-Kommunikationsmöglichkeiten durch Kombinationen von Kabelfernsehen mit Rückkanaltechnik, Bildschirmtext oder Tasten-Telefon ermöglichen eine Ausdehnung von Marktforschungsuntersuchungen über das Kaufverhalten von Haushalten. Wenn Fallstudien im Rechnerdialog ⁸⁾ im Rahmen des Unterrichts eingesetzt werden, bieten sie ebenfalls Veranlassung und Möglichkeiten zur Untersuchung des Benutzerverhaltens. Dieser Fall ist schon als Übergangssituation zu der folgenden Gestaltungsalternative anzusehen.

7) Spiegel, B., a.a.O., S.44f.

8) Hervorzuheben sind die von F. Böcker begonnenen Entwicklungen solcher Fallstudien. Einen Hinweis auf diese Möglichkeit und weitere "Rationalisierungseffekte für die betriebswirtschaftliche Forschung... etwa bei erforderlichem Vergleich alternativer Modellformulierungen" gibt: Schmidt, R., Zur Verbindung von Modellen, Methoden und Daten bei Unternehmensplanung mit EDV, in: H. Müller-Merbach, Hrsg., Quantitative Ansätze in der Betriebswirtschaftslehre, München 1978, S.95-109, hier S. 108.

Diese zweite Gestaltungsmöglichkeit ist realisiert, wenn ein ausschließlich im Rechnerdialog arbeitendes Programmsystem für Delphi-Prognosen unter Benutzung einer Datenbank unter Laborbedingungen eingesetzt wird, um z.B. den Zusammenhang von Informationsnachfrage und Prognoseleistung zu untersuchen⁹⁾. Das Programm kann ggf. ohne Änderungen auch praktisch eingesetzt werden, wobei die Datensätze den möglicherweise veränderten Anwendungsbedingungen anzupassen sind. Darin kann der simulative Charakter der Abbildung erblickt werden.

Die dritte Möglichkeit erlaubt es, bestimmte kontrollier- te Stimuli in die Experimente einzuführen. Zur Ausschaltung schwer kontrollierbarer Unterschiede im Kenntnisstand realer Phänomene als Grundlage einer Schätzung subjektiver Wahr- scheinlichkeiten wählt z.B. Schütt den Weg, für eine auto- matisch exakt kontrollierbare kurze Zeit Zeichenketten auf einem Bildschirmterminal abzubilden; die unterschiedliche Länge der Zeichenketten wird den Versuchspersonen als Indiz für die unterschiedliche Wahrscheinlichkeit des Auftretens der mit ihnen assoziierten Umweltzustände vor- gestellt¹⁰⁾. Damit wird ein in der Darbietung gleichmäßig kontrolliertes Informationsangebot erzeugt.

Weitere Anwendungen und Variationen dieser Möglichkeit sind vorstellbar. So kann z.B. durch eine unterschiedlich lange Präsentation von Nachrichten, etwa den eben genannten Zeichen- ketten, Zeitdruck als Streßphänomen bei Entscheidungen

9) Vgl. Brockhoff, K., Delphi-Prognosen im Computer- Dialog, a.a.O., pass. Aus dem Gebiet der Planung ist bei- spielhaft zu verweisen auf Schmidt, R., Modellgestützte Generierung von Planbilanzen mit PLASMA, in: P.Stahlknecht, Hrsg. Online-Systeme im Finanz-und Rechnungswesen, Berlin, Heidelberg, New York 1980, S. 201-219.

10) Vgl. Schütt, K.P., Wahrscheinlichkeitsschätzungen im Computer-Dialog, Stuttgart 1981.

modelliert und in seinen Konsequenzen untersucht werden; Zugangsmöglichkeiten zu bestimmten Daten, Methoden oder Modellen können zeit-, ergebnis- oder benutzerabhängig eröffnet oder geschlossen werden, um die resultierenden Effekte zu untersuchen; der Rechner kann in Kombination mit anderen Arbeits- und Entscheidungshilfen eingesetzt werden, um die Aufteilung der Aktivitäten zwischen den verschiedenen Hilfsmitteln zu prüfen: es fällt auf, daß manche Personen Abschriften von Bildschirmangaben anfertigen, obwohl dieselben Daten nach Belieben reproduziert werden können.

Die automatische Registrierung von Benutzeraktivitäten ist ein wesentlicher Grund für die Wirtschaftlichkeit des Rechners als Experimentierinstrument im Vergleich mit einer Aufnahme vergleichbarer Daten durch Versuchsleiter. Damit können Grenzen der Administrierung von Experimenten hinausgeschoben werden, die z.B. Scheuch bei der Zuweisung eines Versuchsleiters zu maximal drei Versuchspersonen gegeben sieht¹¹⁾. Durch Zeitersparnis kann auch dem Ausfall von Versuchspersonen vorgebeugt werden, und zeitabhängige Umwelteinflüsse können reduziert werden.

Letztlich ist es denkbar, daß der Umfang experimenteller Studien in Abhängigkeit von den Ergebnissen nur so weit ausgedehnt wird, wie dies zur Prüfung einer Hypothese bei vorgegebenem Sicherheitsniveau erforderlich ist. Dieser Gedanke folgt analog den bei sequentiellen Stichproben oder der dynamischen Qualitätskontrolle angewendeten Prinzipien.

Gestalterische Möglichkeiten und forschungsökonomische Gründe sprechen also für den Einsatz dialogfähiger Rechner als Experimentierinstrumenten in bestimmten Aufgabenbereichen.

Potentielle Einschränkungen der Brauchbarkeit dieser Instrumente liegen - sieht man von lokalen oder temporären Zugangsproblemen einmal ab - in zwei Punkten:

11) Scheuch, F., Heuristische Entscheidungsprozesse in der Produktionspolitik, Berlin 1977, S. 104

Mit dem Recher kann nur schriftlich (in Zukunft auch mündlich) kommuniziert werden. Das gilt auch für den Nachrichtenaustausch mit anderen Personen über den Rechner. Das darf gegenüber der "normalen", im Feld beobachteten Kommunikationsweise bei einer gegebenen Aufgabenstellung nur kontrollierbare Einschränkungen bewirken (z.B. hinsichtlich der Reichhaltigkeit der Ausdrucksmittel¹²⁾) oder Freiheiten gewähren (z.B. bei der Abgabe potentiell unpopulärer Vorstellungen).

Die Entwicklung der Programme verursacht erhebliche Aufwendungen. Bisher können diese nur beschränkt durch Rückgriff auf standardisierte Programmteile gesenkt werden. In die Entwicklungsaufwendungen gehen aber auch Beträge ein, die aus mangelnder Erfahrung mit dem Einsatz des Instruments resultieren. Im folgenden wird über einige dieser Erfahrungen berichtet.

2. Rahmenbedingungen

Experimente laufen in einer Umwelt ab, die ihnen bestimmte Rahmenbedingungen auferlegt. Sie gehen auf gesellschaftliche Normen zurück. Wer nicht die bei einer Verletzung der Normen in Aussicht gestellten Sanktionen akzeptiert, muß die Normen einhalten. Für experimentelle Forschungen sind verschiedene Normen gesetzt, die z.T. sogar unter dem besonderen Schutz des Grundgesetzes stehen. Hier soll nur auf einen Standard von spezifischer Bedeutung verwiesen werden¹³⁾ :

12) Hiltz, St.R., A Social Scientist Looks at Computer Conferencing, ICC, 1976; dies., Computer Conferencing, Assessing the Social Impact of a New Communications Medium, Technological Forecasting and Social Change, Vol. 10, 1977, S. 225-238, hier S. 227.

13) Simitis, Sp., et al., Kommentar zum Bundesdatenschutzgesetz, Baden-Baden 1978.

Standard 1: Die Anforderungen des Datenschutzes sind zu beachten.

Besonders einschränkend kann dieser Standard bei persönlichen Daten wirken. Eine persönliche Identifikation von Versuchspersonen ist allerdings nur erforderlich, wenn Nachfragen erfolgen sollen oder Fortsetzungen und Replikationen von Experimenten mit identischem Personenkreis notwendig sind. Solche Experimente könnten erstens durch die Einholung von Einverständniserklärungen ermöglicht werden. Zweitens ist die Bezeichnung von Versuchspersonen mit Codeziffern oder fiktiven Namen möglich, die nur die Versuchsperson selbst kennt. In einem solchen Falle ist durch eine besondere Prüfung zu sichern, daß diesselbe fiktive Personenkennzeichnung nicht von mehreren Versuchspersonen gewählt wird. Das ist bei Codeziffern mit hohem symbolischen oder semantischen Gehalt (4711,007) ebenso zu befürchten, wie bei Namen, mit denen sich Mitglieder derjenigen Subkulturen, aus denen die Versuchspersonen stammen, gerne identifizieren.

3. Standards in der Phase des Versuchsaufbaus

Feld- oder Laborexperimente werden in einer Aufbauphase vorbereitet und anschließend durchgeführt, gegebenenfalls nach Modifikationen aus Erkenntnissen, die in der Aufbauphase gewonnen wurden. Die Auswertung der Ergebnisse kann sich mit der Durchführungsphase überlappen.

In der Aufbauphase ist die Erfüllung dreier Standards notwendige Voraussetzung für ein Gelingen der späteren experimentellen Arbeit.

Standard 2: Auch der Einsatz eines Rechners als Experimentierinstrument erfordert die sorgfältige Planung der Versuchsanordnung ("experimental design").

Die Planung der Versuchsanordnung gestattet es, wesentliche Störungsquellen der Reliabilität und der Validität der aus der Durchführungsphase erwarteten Aussagen vor auszusehen¹⁴⁾. Die Beurteilung der Bedeutung einzelner Störungsquellen kann beim Rechnereinsatz von einer Versuchsdurchführung mit anderen Instrumenten in typischer Weise abweichen.

An die Stelle der Interaktionseffekte zwischen Versuchsleiter und Versuchsperson in den herkömmlichen Experimenten können ähnliche Interaktionseffekte zwischen Versuchsperson und Rechner treten. Das haben z.B. Johnson und Baker¹⁵⁾ gezeigt.

Die motivierenden oder demotivierenden Wirkungen des Rechners selbst müssen von vergleichbaren Aspekten z.B. der Aufgabenstellung meßtechnisch zu trennen sein. Schaut man sich die dritte der eingangs genannten Gestaltungsmöglichkeiten für Experimente mit Rechnern als Experimentierinstrumenten an, so wird deutlich, daß eine solche Trennung der Effekte nur sehr schwer zu realisieren ist. In diesem Falle hilft nämlich auch die Einrichtung einer Kontrollgruppe, die ohne Rechner arbeitet, nicht weiter, weil dann auch die durch den Rechner darzustellenden Versuchsteile nicht mehr zu realisieren sind¹⁶⁾. In diesem Zusammenhang empfiehlt sich auch die Standardisierung der kontrollierbaren Außeneinflüsse auf die Versuchspersonen im Experiment, sei es durch Personen oder durch Hilfsmaterial (Instruktionen, Taschenrechner).

14) Hinweise geben neben dem Standardwerk von Campbell und Stanley, a.a.O., auch: Lienert, A., Testaufbau und Testanalyse, 3.A., Weinheim, Berlin, Basel 1969; Zimmermann, E., Das Experiment in den Sozialwissenschaften, Stuttgart 1972.

15) Johnson, E.S., Baker, R.F., The Computer as Experimenter: New Results, Behavioral Science, Vol.18, 1973, S.377-385.

16) Delbecq, A.L., van de Ven, A.H., Gustafson, D.H., Group Techniques for Program Planning. A Guide to Nominal Group and Delphi Processes. Glenview/Ill. 1975, S.84; Johansen, R., Miller, R.H., Vallée, J., Group Communication through Electronic Media. Fundamental Choices and Social Effects. In: H.A.L. Linstone, M. Turoff, The Delphi Method and Applications, Reading/Mass. 1975, S. 516-534, hier S. 529.

Die praktisch unkontrollierbaren und durch wechselnde Kapazitätsauslastung des Rechners bedingten Veränderungen seines Antwortverhaltens können besonders dann Störungen verursachen, wenn die Versuchspersonen bestimmte Erwartungen an das Antwortverhalten stellen, die nicht erfüllt werden.

In den klassischen experimentellen Anordnungen können "Instrumentierungs-Effekte" störend eingreifen. Sie resultieren aus einem Wechsel der Instrumente, hier z.B. aus dem Wechsel verschiedener Bildschirme. Dieser Typ von Effekten ist vergleichsweise leicht kontrollierbar.

Der Versuchsaufbau sollte es ermöglichen, Verzerrungen auszu-sondern, die durch den Auf- oder Abbau von Akzeptanzbarrieren des Rechnereinsatzes bei den Versuchspersonen entstehen können. Bedienungsprobleme, Gewöhnung an das System, Ermüdung sind Beispiele für Symptome solcher Barrieren. Sie können auch bei Experimenten ohne Einsatz von Rechnern auftreten, nehmen hier aber besondere Formen an.

Aus zwei Gründen kann der Rechnereinsatz die Einhaltung experimenteller Designs in besonderer Weise fördern. Erstens verfügen praktisch alle Rechner über Zufallszahlengeneratoren, die die Durchführung zufallsgesteuerter Experimente erleichtern. Zweitens kann die Variation der experimentellen Bedingungen nach festgelegten Plänen objektiv und automatisch gesteuert werden. Es hat sich bewährt, die verschiedenen Möglichkeiten über einen "Kontextvektor" zu steuern.

Mit einem weiteren Standard wollen wir der Forderung entsprechen, daß die Möglichkeit intersubjektiver Überprüfbarkeit experimentell gewonnener Ergebnisse Priorität hat gegenüber systemspezifischen Vorteilen der Gestaltung eines Experiments.

Standard 3: Wahlmöglichkeiten bei Hardware und Software erfordern die Abwägung zwischen der möglichst vollständigen Ausnutzung spezifischer Vorteile von Anlagen, Betriebssystem sowie Programmiersprachen im Interesse effizienter Experimentdurchführung und der Chance der Übertragbarkeit (Portabilität) des Experiments im Interesse der Replikation.

Die Übertragbarkeit kann schon beim Versuchleiter eingeschränkt werden, wenn zwischen Aufbau- und Durchführungsphase oder während dieser Phase Maschinenwechsel oder Veränderungen am Betriebssystem unvermeidbar sind. Je "systemischer" der Versuchsaufbau, um so schwieriger ist der Übergang zu vollziehen.

Dieselbe Überlegung gilt für die Ermöglichung von Replikationen an anderen Orten. Hierfür sind identische Versuchsaufbauten vorzuziehen, die durch Programmübertragung erleichtert werden¹⁷⁾.

Standard 4: Der Versuchsaufbau ist einem Pretest zu unterziehen.

Was vor dem Einsatz anderer Instrumente zur Datengewinnung erforderlich ist, darauf kann auch beim Rechnereinsatz nicht verzichtet werden. Zusätzlich zu den üblichen Funktionen eines Pretest, spezifische Fehler und Mißverständnisse aufzudecken, die aus der Benutzung des Rechners resultieren. An erster Stelle ist dabei an Programmierungsfehler zu denken. Eine ausführlichere Beschreibung der Fehlermöglichkeiten wird bei der Erörterung des Standards 7 gegeben.

Ein Pretest kann seine Funktion nur erfüllen, wenn er unter Bedingungen realisiert wird, die gegenüber den Versuchsbedingungen im Experiment keine Erleichterungen vorsehen.

17) Die Verlegung der schon zitierten Delphi-Experimente von Kiel nach ^①Frankfurt^② setzte zunächst die Suche nach einem entsprechenden Rechner voraus, sodann das Einverständnis ihn nutzen zu können und schließlich die Aktivierung eines auf diesem Rechner zur Versuchszeit nicht gebräuchlichen Betriebssystems.

Studenten und Mitarbeiter sind dann wenig geeignete Teilnehmer an Pretests, wenn sie motiviert sind oder gelernt haben, den Versuchsaufbau schonend zu behandeln. Sie können in einem Ausmaß kooperieren, wie es von den Versuchspersonen der Durchführungsphase nicht erwartet werden kann¹⁸⁾.

Die weiteren Standards kommen in der Durchführungsphase zur Anwendung, sind aber beim Aufbau der Versuche antizipierend zu berücksichtigen.

4. Standards in der Phase der Versuchsdurchführung

Abgesehen von einer Veränderung der experimentellen Variablen sollen alle übrigen Bedingungen während der Durchführung von Experimenten möglichst gleich bleiben oder nur mit kontrollierbarer Wirkung veränderlich sein. Störungen könnten durch die eingesetzten Geräte und durch die Versuchspersonen ausgelöst werden. Beide Ursachen sind durch die Einhaltung von Standards zu erfassen und partiell zu vermeiden.

Standard 5: Alle für die Fortführung beliebig unterbrochener Experimente oder für die Auswertung ihrer Ergebnisse notwendigen Daten sind parallel zur und gleichzeitig mit der Durchführung der Experimente so zu speichern, daß nach einem Abbruch unmittelbar ein neuer Start an der Abbruchstelle möglich ist.

18) Die Delphi-Experimente zur Zinsprognose sind durch Mitarbeiter und Studenten ohne Beanstandungen getestet worden. Vor ihrem Einsatz mit Versuchspersonen aus der Praxis, denen die benutzten Bildschirmterminals in ihren Funktionen unbekannt waren, wurden Journalisten und Politiker zu einer Wahlprognose eingeladen. Hierbei traten Bedienungsfehler auf, die vor Eintritt in die Durchführungsphase beseitigt werden konnten.

Die Einhaltung dieses Standards beugt Datenverlusten vor, die z.B. bei Maschinenstillstand (Stromausfall ohne Kompensation, Zusammenbruch des Betriebssystems) auftreten können. Werden diese Störungen sehr schnell behoben, kann man ein Experiment an der Abbruchstelle neu starten. Damit wird eine vollständige Wiederholung vermieden und sparsam mit der knappen Zeit der Versuchspersonen umgegangen.

Allerdings ist ein Neustart problematisch, wenn er die experimentellen Bedingungen verzerrt. Den übrigen stochastischen Störgrößen sind die Effekte solcher Neustarts zuzuordnen, bei denen die Auswirkungen unter der Schwelle bleiben, bei deren Überschreitung sie als Bias feststellbar wären.

Standard 6: Versuchspersonen dürfen weder Zugang zu dem Rechnerprogramm haben, das die Versuche steuert, noch zu den Ausgangs- oder zu den Ergebnisdaten.

Ist der Programmzugang nicht blockiert, kann weder ein unkontrollierter Abbruch des Experiments vermieden noch die Datensicherung oder die Stabilität der experimentellen Bedingungen garantiert werden. Der Standard berührt nicht die Unterscheidung experimenteller Situation zwischen den Polen "voll biotisch" und "offen"¹⁹⁾.

Ein (ebenfalls auszuschließender) Extremfall des "Zugangs" liegt vor, wenn die Versuchspersonen auf den Monitorlevel des Rechners zurückgehen können, wodurch sie sich ohne Zustimmung des Versuchsleiters aus dem Experiment ausschalten (Betätigung der Tasten "Control" und "C" des Eingabefeldes). Das Programm muß "Control C - fest" gemacht werden, um diese Möglichkeit auszuschließen. Dieselben Vorkehrungen sind zu treffen, wenn die Versuchspersonen unerlaubte Löschungen (nicht aber Eingabekorrekturen) vornehmen könnten.

19) Vg. Spiegel, B., a.a.O., S.44.

Standard 7: Benutzerreaktionen gegenüber dem Rechner sind auf Fehler zu prüfen und gegebenenfalls zurückzuweisen; die festgestellten Fehler sollen in einer parallel geführten Fehlerstatistik erfaßt werden.

Die Häufigkeit von unerwünschten Fehlern von Versuchspersonen wird sich verringern lassen, wenn die im Zusammenhang mit dem Standard 3 (Pretest) vorgeschlagenen Maßnahmen eingesetzt werden. Versuchspersonenfehler können aber auch insoweit erwünscht sein, als sie Indizien für Schwierigkeiten der Aufgabenlösung darstellen. Liegt der Zweck eines Experiments in einer solchen Feststellung, dann könnten Versuche einer starken Fehlervermeidung dem experimentellen Zweck widersprechen. Wir wollen deshalb auch nur von einer Fehlerprüfung sprechen.

Bei solchen Prüfungen treten verschiedene Arten von Fehlern auf. Es konnte z.B. beobachtet werden, daß bestimmte Versuchspersonen sich nicht in der Lage sahen, entsprechend den erteilten Weisungen Eingaben am Bildschirm nur dann zu machen, wenn sie dazu aufgefordert werden. Vorzeitige Eingaben werden aber in der Regel beim nächsten Lesebefehl als die eigentlich gewünschte Eingabe eingelesen. Damit können nicht nur Fehler auf der syntaktischen Ebene auftreten, sondern auch Verschiebungen in der Zuordnung von Antworten zu Fragen eines Dialogs vorkommen, die bei syntaktischer Richtigkeit von der Versuchsperson oder vom Versuchsleiter kaum bemerkt werden können.

Aus solchen Beobachtungen folgt der Vorschlag, entsprechend den häufig beobachteten Fehlertypen drei Arten von Kontrollmaßnahmen vorzusehen:

- vollständige Löschung des Datenendgerät-Buffers einer Versuchsperson unmittelbar vor einer erwarteten Eingabe oder Antwort der Versuchsperson; das sichert die Zuordnung von Stimuli (Fragen) und Reaktionen (Antworten);
- syntaktische Kontrolle auf Verwendung der für die Reaktionen zugelassenen Zeichen (z.B. keine Zulassung von Bsträgen in Worten, wenn sie in Ziffern darzustellen sind);
- semantische Kontrolle auf Plausibilität der Reaktionen (Kontrolle, ob numerische Angaben innerhalb zulässiger oder vernünftiger Grenzen liegen; automatische Inhaltsanalyse von Texten und Vergleich mit einem Wörterbuch plausibler Antworten ²⁰⁾; gegebenenfalls kann eine Rückbestätigung eingeholt werden).

In den Bereich der semantischen Kontrolle fällt die Prüfung auf zulässig oder unzulässig verwendete Abkürzungen und andere Zeichenketten (strings), die die Beantwortung mehrerer aufeinanderfolgender Stimuli vorwegnehmen können. Routinisierte Versuchspersonen verwenden solche Beschleunigungen der Dialoge gern. Es ist zu prüfen, ob dies im Experiment unerwünscht ist

20) Hierfür kann ein System von Ahrens verwendet werden:
L. Ahrens, Inhaltsanalyse als Instrument zur Beschreibung von Unternehmensstrukturen, Dipl. Arb. (Lehrstuhl für Finanzwirtschaft), Kiel 1978. Vgl. Zimmermann, W.L., Datenverarbeitung, 2. Bd., Wiesbaden 1977, S. 217.

oder nicht²¹⁾. Werden Abkürzungen und Zeichenketten erlaubt so kann es zweckmäßig sein, einzelnen Versuchspersonen gegenüber diese Erlaubnis automatisch einzuschränken oder zu entziehen, wenn ein vorgegebenes Fehlerniveau bei den Eingaben überschritten wird²²⁾.

Die Einhaltung der Standards 6 und 7 sichert die an anderer Stelle als "Robustheit" bezeichnete Eigenschaft des Programms²³⁾.

Standard 8: Der Versuchsleiter muß die Möglichkeit haben, sich über den Stand des Versuchs bei jeder Versuchsperson zu informieren. Er soll über den Rechner mit den Versuchspersonen kommunizieren können.

Die durch diesen Standard eröffneten Möglichkeiten sollten nur sparsam verwendet werden, um nicht als experimentelle Variable zu wirken. Sie sind aber besonders dann erforderlich, wenn Versuchspersonen und Versuchsleiter nicht an demselben Ort sind. Es kann notwendig sein, Versuchspersonen an die Abgabe ("return") einer schon formulierten Reaktion zu erinnern, sie über außergewöhnliche Ereignisse in Kenntnis zu setzen (z.B. Abbruch oder neuer

21) Wird die Benutzung erlaubt, so steigt das Aktivitätsniveau des Benutzers an. Er nimmt folgende Stimuli und Reaktionen vorweg. Das kann sein Antwortverhalten, seine Motivation, sein Lernen usw. im Vergleich zu anderen beeinflussen. Katter hat darauf hingewiesen, daß die Schnelligkeit der EDV-Anlage den Benutzer zu schnellerer als optimaler Reaktion veranlassen kann, d.h. zu Unbedachttheiten und Unachtsamkeiten: Katter, R.V., On the On-Line User of Remote-Access Citation Retrieval Services, SDC, Santa Monica/Cal. 1970, Memorandum L4494/000/00. Vgl. Nickerson, R.S., Elkind, J.I., Carbonell, J.R., Human Factors and the Design of Time Sharing Computer Systems, Human Factors, Vol. 10, 1968, S. 127-134. Zur optimalen Antwortrate vgl.: Martin, J., Design of Man-Computer Dialogues, Englewood Cliffs/N.J. 1973, S. 12.

22) Vgl. Ling, R.F., Roberts, H.V., IDA: An Approach to Interactive Data Analysis in Teaching and Research, Journal of Business, Vol. 48, 1975, S. 411-451. Realisiert ist eine solche Sprachniveau-Veränderung z.B. im Siemens-System IPROS und in den CALL-Routinen von IBM.

23) Kurbel, K., Die Qualität von Softwareprodukten. Ziele und Kosten der Entwicklung von Anwendungssoftware, Die Unternehmung, 34. Jg., 1980, S. 237-262, hier S. 241.

Start) usw.²⁴⁾. Zur besseren Eingrenzung des Verhaltensspektrums des Versuchsleiters kann es zweckmäßig sein, allein feststehende Texte für seine Kommunikation mit Teilnehmern zuzulassen. Weitergehend wäre der Vorschlag, die Verwendung solcher Texte allein durch ein "master program" zu gestatten, dessen Eingriffe aufgrund objektiver Versuchsbedingungen ausgelöst wird statt durch die subjektiven Eindrücke des jeweiligen Versuchsleiters.

Standard 9 (Metastandard): Die für die Dialogprogrammierung üblichen Standards zur Erzielung verständiger Reaktionen der Versuchspersonen sind zu beachten.

Eine detaillierte Erläuterung zu den Konsequenzen dieses Standards erübrigt sich durch Hinweis auf die Literatur²⁵⁾. Mißachtung des Standards kann erhebliche Akzeptanzprobleme bringen oder zusätzliche Fehlerquellen durch Verletzung der "funktionellen Fähigkeiten" der Programme eröffnen²⁶⁾.

Es ist z.B. nicht auszuschließen, daß einander widersprechende Ergebnisse über die Bereitschaft, Graphiken an Stelle von Tabellen in Entscheidungs- und Urteilsprozessen zu verwenden, auf Unterschiede in der Darstellung zurückzuführen sind²⁷⁾. - Gelegentlich bleibt auch unbeachtet, daß die Anzahl der Spalten und Zeilen je "Seite" zwischen den verschiedenen Ein- und Ausgabemedien schwankt (z.B. 20-Zeilen/80-Spalten Bildschirme im Vergleich mit 50-Zeilen/132-Spalten Papier).

24) Vgl. die Möglichkeiten bei Brockhoff, K., Delphi-Prognosen im Computer-Dialog, a.a.O., S. 189.

25) Z.B. Seibt, D., Organisation von Software-Systemen, Wiesbaden 1973, S. 93ff.; Brockhoff, K., Prognoseverfahren für die Unternehmensplanung, Wiesbaden 1977, S. 153ff. Kurbel, K., a.a.O., unterscheidet vier Ebenen von Qualitätszielen für Programme: 1. Ebene: Zuverlässigkeit; 2. Ebene: Verständlichkeit, Adaptabilität, Portabilität; 3. Ebene: Robustheit, Benutzerfreundlichkeit, Universalität; 4. Ebene: Effizienz.

26) Seibt, D., a.a.O.

27) K. Brockhoff, Delphi-Prognosen im Computer-Dialog, a.a.O., S. 105ff., wo Graphik-Abfrage relativ selten vorkommt und bei mangelnder Vertrautheit mit Aufgabe und System eher bevorzugt wird. Anders: Kroeber-Riel, W., Konsumentenverhalten, 2.A., München 1980, S. 109 f. und die dort angeg. Lit.

Die mit Standard 9 angesprochene Forderung nach Benutzerfreundlichkeit wird in der Literatur unterschiedlich weit interpretiert. In Experimenten muß nicht immer die benutzerfreundlichste Version realisiert werden, wenn z.B. der Einfluss unterschiedlich benutzerfreundlicher Realisierungen von Programmen als experimentelle Variable aufzufassen ist.

5. Standards in der Auswertungsphase

Auch die Auswertungsphase wird von einem Metastandard beherrscht, nach dem die angemessenen Auswertungstechniken einzusetzen sind. Die Techniken setzen aber interpretierbares Datenmaterial voraus. Mit der Sicherung dieser Datengewinnung, die teilweise schon durch Standard 5 berührt wird, wollen wir uns hier befassen.

Standard 10: Unabhängig von variierenden Untersuchungshypothesen sollen eine Fehlererfassung und eine Zeiterfassung geführt werden.

Beide Übersichten dienen der Verbesserung der Programme, der verbesserten Vorbereitung von Versuchspersonen und der Interpretation von Ergebnisunterschieden.

Die Fehlererfassung unterscheidet nach Programmfehlern oder Softwarefehlern, Maschinenstörungen oder Hardwarefehlern und Fehlern der Versuchspersonen, die nach den oben erläuterten Kategorien auszuzählen sind. Bei Maschinenstörungen ist gegebenenfalls eine Regel für Neustarts anzuwenden. Programmfehler sollten zwar die Experimente nicht belasten, doch können sie auch bei Beachtung des Standards 4 nicht völlig ausgeschlossen werden: Es kann zwar die Anwesenheit von Fehlern ausgetestet werden, nicht aber ihre Abwesenheit 28) .

28) Vgl. Dijkstra, E.W., Notes on Structured Programming, in: ders., C.A. Hoare, Eds., Structured Programming, London, New York 1972, S. 1-82, hier S. 6. Vgl. die Hinweise auf Korrektheitsbeweise bei Kurbel, K., a.a.O., S. 240.

Die Zeiterfassung unterscheidet nach Reaktionszeiten der Versuchsperson und Arbeitszeiten des Rechners. Die erste Zeitkategorie kann Aufschlüsse über besonders schwierige Aufgaben für einzelne Versuchspersonen zulassen²⁹⁾. Die zweite Zeitkategorie dokumentiert die Folgen unterschiedlicher Kapazitätsauslastung des Rechners. Im time-sharing-Betrieb kann die Bewältigung anderer Benutzeraufgaben den Ablauf der Experimente z.B. durch Verzögerungen beeinflussen. Das kann die Vergleichbarkeit zwischen verschiedenen Experimenten beeinflussen.

Zum Abbau der vom Versuchsleiter unerwünschten Spannungen, die aufgrund der von den Versuchspersonen unerwarteten Wartezeiten auftreten können, lassen sich Nachrichten über den Grund von Verzögerungen gut verwenden. ("Jetzt können Wartezeiten auftreten, da noch die Ergebnisse anderer Teilnehmer abzuwarten sind", wurde in einem Experiment mit Gruppen ausgegeben. In einem Individualdialog kommt die Mitteilung vor:"Jetzt werden die von Ihnen gewünschten Rechnungen durchgeführt. Bitte, gedulden Sie sich bis zur Ausgabe der Ergebnisse".)

Standard 11: Es soll ein vollständiges Versuchsprotokoll erstellt werden.

Das Experiment wird zur Datengewinnung bei der Prüfung bestimmter Hypothesen aufgebaut. Deshalb wird die Auswertung häufig auch allein auf diese ursprünglich für relevant gehaltenen Daten beschränkt. Es können aber später weitere Hypothesen auftreten, die nur bei Speicherung aller Beobachtungen überprüfbar sind. Beispielsweise kann sich eine Hypothese zunächst ausschließlich auf die Erfassung einer Informationsmenge unterschiedlicher Informationsarten

29) Vgl. Boies, S.J., User Behavior on an Interactive Computer System, IBM Systems Journal, Vol. 13, 1974, S. 2-18, hier S. 9.

richten. Erst später werden Aussagen über die Reihenfolge der Abfrage der Informationsarten aufgestellt. Die Überprüfung dieser Aussagen ist unmöglich, wenn im Hinblick auf die erste Hypothese nur eine Häufigkeitsstatistik der Informationsnachfrage nach Informationsarten angelegt wurde. Ein vollständiges Versuchsprotokoll würde aber die Reihenfolge der Abfrage noch erkennen lassen und könnte daraufhin ausgewertet werden.

"Vollständigkeit" ist insoweit ein relativierbares Kriterium, als vereinbarte Abkürzungen, Zusammenfassungen usw. verwendet werden können, wenn damit die Rekonstruktion der experimentellen Ergebnisse nicht verhindert wird.

6. Schluß

Die elf vorgestellten Standards werden als ein System von Grundsätzen ordnungsgemäßer Realisierung von Experimenten mit Rechnern angesehen, die in Kommunikation mit den Versuchspersonen stehen. Die ordnungsgemäße Realisierung ist ordnungsgemäß zu dokumentieren. Die Standards sind weitgehend aus der Beobachtung von Reaktionen einzelner Versuchspersonen gewonnen worden. Es ist deshalb denkbar, daß sie in dem Umfang zu revidieren sind, wie sich das Verhalten von Versuchspersonen gegenüber der Peripherie von Rechnern ändert, was sicher von den wahrgenommenen Eigenschaften der Rechner und ihrer Peripherie abhängt.

Weiterhin ist denkbar, daß die mitgeteilten Erfahrungen ausgebaut werden. So liegen z.B. bisher nur wenige, am Standard 6 (Zugang der Versuchspersonen zu Programm und Daten) anknüpfende Erfahrungen vor, die eher auf die Notwendigkeit einer deutlich abgeschlossenen Programmierung der experimentellen Aufgabe hinweisen als darauf, dem in anderem Zusammenhang geforderten Standard der Universalität³⁰⁾ zu entsprechen. Fraglich ist auch, ob Aussagen über die Ermüdung von Versuchspersonen gewonnen werden können, die zu einem Standard über die zeitliche Ausdehnung von Experimenten führen könnten. Auch deshalb darf als sicher gelten, daß die hier vorgestellten Standards sich im Laufe der Zeit als nicht erschöpfend erweisen werden.

³⁰⁾ Kurbel, K., a.a.O., S.246.

Zusammenfassung

Auf drei Weisen können dialogfähige Rechenanlagen im Rahmen von betriebswirtschaftlichen Experimenten eingesetzt werden. Vor- und Nachteile dieser Vorgehensweisen werden kurz erwähnt. Es folgt die Aufstellung von elf Standards, die beim ordnungsgemäßen Aufbau der Durchführung und der Auswertung der Experimente zu beachten sind.

Summary

There are three ways to use computer dialogues to perform experiments in business administration. Benefits and drawbacks of such experiments are mentioned. This is followed by the derivation of eleven standards that serve as guidelines for the set-up of experiments, their performance and the derivation of results.