



World Library and Information Congress: 69th IFLA General Conference and Council

1-9 August 2003, Berlin

Code Number: 015-G
Meeting: 127. National Libraries & Statistics and Evaluation
Simultaneous Interpretation: -

Leistungsvergleichstests: Übersicht und Zusammenhänge

Rowena Cullen

School of Information Management
Victoria University of Wellington
Wellington
New Zealand

„Management hat mit Wandel zu tun und der Garantie ständigen Wandels“.
Sir John Harvey-Jones (*Making it Happen*)

Zusammenfassung

Der Aufsatz untersucht einige der mit Leistungsvergleichstests verwandten Themen und deren erfolgreiche Anwendung in wissenschaftlichen, öffentlichen und speziellen Bibliotheken mit dem Ziel der Qualitätssteigerung. Diskutiert wird der Bedarf an Leistungsvergleichen und Fachgutachten zur Bewertung und Analyse von Ergebnissen, ebenso die herkömmlichen Eingaben, Verarbeitungen und Ausgabedaten unter Berücksichtigung der Bedeutung sozialer Zusammenhänge. Erläutert werden die Bedeutung von Bewertungsprogrammen für Nationalbibliotheken und einige Möglichkeiten, Tests und Fachgutachten für Qualitätssteigerungen einzusetzen. Der Aufsatz behandelt Verfahren, wie die besonderen Bedürfnisse von Nationalbibliotheken zu berücksichtigen sind, und deren Schwierigkeiten bei der Suche nach Vergleichspartnern. Elemente des „bestmöglichen Verfahrens“ auf diesem Gebiet werden bestimmt.

Einführung

Wir leben in einer Zeit großer Veränderungen, und manchmal ist es schwierig, gleichzeitig mit dem Wandel fertig werden zu müssen und dabei unsere langfristigen, festen Ziele nicht aus den Augen zu verlieren: einer möglichst großen Zahl von Kunden und Interessenvertretern Dienstleistungen auf höchstem Niveau anzubieten. Um das zu erreichen, müssen wir herausfinden, welche Veränderungen extern ablaufen, welche intern zu geschehen haben, und wir müssen den Prozeß des Wandels steuern, um Beides zu verbinden. Externe

Veränderungen, auf die wir zu reagieren haben, sind die sich wandelnde Welt der Informations- und Kommunikationstechnologie, die alle Aspekte unseres Berufs beeinflusst, die wechselnden Erwartungen und Kunden und die sich ändernden Anforderungen hinsichtlich Wirtschaftlichkeit, Dienstleistungsqualität sowie nationaler und weltweiter Zusammenarbeit.

Auf vielen anderen Gebieten des Bibliothekswesens, teilweise auch bei Nationalbibliotheken, sind Leistungsvergleichstests akzeptiert worden als ein Weg, mit äußeren Einflüssen umzugehen und interne Veränderungen zu fördern. In diesem Aufsatz untersuche ich einige der mit Leistungsvergleichstests verwandten Themen, ihre erfolgreiche Anwendung, beschreibe einige Verfahrensweisen und befaße mich dann mit der Lage der Nationalbibliotheken, dem Wert und der Bedeutung von Tests in diesem Bereich. Da Vergleiche oft herangezogen werden, um sogenannte „beste Verfahren“ zu finden, behandle ich schließlich die Frage, wie „beste Verfahren“ auf diesem Gebiet aussehen könnten.

Aktuelle Themen bei Vergleichstests

Obwohl Foot (1998), der für kommunale Dienststellen und Bibliotheken schreibt, eine „benchmark“ als „ein(en) Fixpunkt an dem man sich messen kann, wie ein begrenztes Ziel oder ein Standard“, ist diese Definition doch etwas eng. Es stimmt zwar, daß viele bibliothekarische Testverfahren in der Vergangenheit zu statistischen Normen geworden sind, ob sie nun einer Gruppe vergleichbarer Bibliotheken entstammen oder einer nationalen Datensammlung, mit Bezug auf Eingaben wie Finanzmittel und Personalausstattung, Wirkungsgrad und Ergebnisdaten (z.B. Entleihungen u.ä. Geschäftsvorfälle), und die für Vergleichszwecke genutzt wurden (Pritchard 1995). Dies ist noch immer Bestandteil der Verfahren, die im Vereinigten Königreich bei LISU (Library and Information Statistics Unit at Loughborough University),¹ IPF (Institute of Public Finance)² und SCONUL (Society of College, National and University Librarians).³ Aber im industriellen und privatwirtschaftlichen Bereich, aus dem die Testmethoden stammen und wo sie noch immer viele Anhänger haben, sind die Vergleichstests viel stärker auf Geschäftsprozesse ausgerichtet und befassen sich weniger mit Aktivitäten im Herstellungs- oder Dienstleistungsbereich (Camp, 1998). Die Internetseite des „Benchmarking Network“ verzeichnet viele und umfassende Beispiele an Vergleichstestverfahren; zu den zehn Spitzenreitern für die Mitgliedsgesellschaften zählen „Personalentwicklungstraining“, „Technologie von Informationssystemen“, „Arbeitsreserven“ und „Entwicklung von Leistungsbewertung“ als die vier wichtigsten.⁴ Obwohl man Beispiele für solche Vergleichsverfahren auch im Bibliotheksbereich findet, sind sie dort weit weniger verbreitet als statistische Vergleiche von Eingabe- und Ergebnisdaten sowie die gern benutzten, aber nicht erforschten Leistungsvergleiche. Dies geschieht im Gegensatz zum privatwirtschaftlichen Sektor und seinem Interesse daran, wie Verfahren zur Leistungsbewertung entwickelt und angewandt werden.

¹ <http://www.lboro.ac.uk/departments/dils/lisu/lisuhp.html>

² <http://www.ipf.co.uk/benchmarking>

³ <http://www.sconul.ac.uk>

⁴ <http://www.benchnet.com>

Foot's zweite Definition des Leistungsvergleichstests ist hilfreicher: „Der Vorgang der Bewertung eigener Dienstleistungsprozesse und Ergebnisse im systematischen Vergleich mit den Ergebnissen Anderer, um größtmöglichen Nutzen zu erzielen.“ Diese zweite Definition betont die nützlicheren Konzepte, systematische Vergleiche, Verfahrensabläufe und Ergebnisse und die Suche nach größtmöglichem Nutzen. Diese Definition entspricht viel mehr der Verwendung von Tests im privatwirtschaftlichen Sektor und ist noch sehr neu für Bibliotheken.

Daher müssen wir bei der Beschäftigung mit Vergleichstests in Bibliotheken auf Folgendes konzentrieren:

- 1) Sammeln und Vergleichen von Daten - Eingaben vor Allem, und Ausgabedaten. Dies umfaßt bibliothekarische Statistik im Bereich nationaler, universitärer und öffentlicher Bibliotheken, wobei Etats verglichen werden, Anzahl und Wert der Erwerbungen, Anzahl des Fachpersonals, Auskunftsfälle und neuerdings elektronisch gestützte Dienstleistungen.
- 2) Verfahrenstests in Zusammenhang mit Analyse von Funktionen und Abläufen, in Zusammenarbeit mit einer oder mehreren ähnlichen Institutionen(en) mit dem Ziel, daß alle Teilnehmer davon profitieren und voneinander lernen.

Die Bedeutung der erstgenannten Testarten wird jetzt vor allem als diagnostisches Werkzeug begriffen, obwohl auch Verfahrenstests zur Diagnose eingesetzt werden können. Allerdings ist nur die auf Verfahren gründende Vorgehensweise von Bedeutung bei der Suche nach dem größtmöglichen Nutzen. Während Verfahrenstests bei einer Institution oder Abteilung (auch mehreren) gewöhnlich mit Hilfe von Fachgutachtern durchgeführt werden, könnte man sie doch auch durch unabhängige Gutachter ausführen lassen, die Praxiskenntnisse aus anderen Institutionen mitbringen, die als Vergleich dienen oder aus denen sich Modelle für „bestmögliche Verfahren“ ableiten lassen.

Die Praxis des Sammelns und Vergleichens von Daten bei einer Anzahl unterschiedlicher Bibliotheken bringt viele Fragen zur Datendefinition mit sich, die den Mitgliedern der Statistiksektion der IFLA gut bekannt sind und von ihnen beschrieben wurden. Diese Themen zu bearbeiten war eine der Stärken von LISU und eine der vorrangigen Aufgaben des Instituts für öffentliche Finanzen (Institute of Public Finance), das Vergleichstests bei Gemeinden und öffentlichen Bibliotheken im Vereinigten Königreich betreibt. Weil die Datensammlung sich vom Zusammenstellen der Eingabedaten fortbewegt in Richtung Ablauf- und Ausgabedaten, können diese Probleme noch umfassender werden. Unterschiede bei Berechnung von Aufwand und Fixkosten sowie Messung von Dienstleistungen und anderen Resultaten vergrößern sich, sobald die Verhältnisse zwischen solchen Daten untersucht werden. Außerdem ist es sehr verlockend, bei solchen Daten nach Ursache und Wirkung zu suchen. Jedoch ist der diagnostische Wert beim Vergleich von nackten statistischen Angaben zu Eingaben, Ausgabedaten und Abläufen sowie der Beziehungen zwischen ihnen beschränkt auf Hinweise drauf, wo Probleme vorhanden sind, die nähere Untersuchung erfordern; die Daten selber verraten nichts zur Kausalität. Man darf keine Schlüsse Rückschlüsse ziehen hinsichtlich der Zusammenhänge von Eingaben, Abläufen und Resultaten, ohne die Plausibilität der benutzten Meßverfahren zu untersuchen und andere Vermutungen, auf denen sie beruhen.

Dies wurde erforscht von Pratt und Altman (1997), die mit einer einfachen Untersuchung der statistischen Angaben von 24 der größten Bibliothekssysteme in den Vereinigten Staaten, die im Bericht über die öffentlichen Bibliotheken Amerikas für 1996 publiziert wurden, den Nachweis erbrachten, daß nicht nur keine Beziehung zwischen Eingaben und Resultaten bestand oder zwischen Aufwand und Wirkung unter Anwendung der fünf benutzten Wirkungsmaßstäbe (Pro-Kopf-Zahlen für Ausleihverbuchungen, Ausleihe, Beratungsvorgänge, Bestandumsatz und Bestandsdaten), sondern daß auch die verwendeten Eingabedaten auf verschiedenen Grundlagen erhoben waren, ebenso die Ergebnisse, mit denen sie verglichen wurden. Aber selbst dann, wenn man diese Unterschiede bei der Datenauswahl und -sammlung vermied, konnten die gewonnenen Zahlen noch immer nicht erklären, ob die offensichtlichen Ergebnisdifferenzen auf mehr oder weniger wirksamer Nutzung des Materials beruhten oder ob sie durch strukturelle, kulturelle und strategische Unterschiede zwischen den einzelnen Organisationen bedingt waren, ihre Umgebung, ihren Kontext. Verfahrenstests müssen durchgeführt werden zur gründlicheren Erforschung des Problems, wenn es denn als Problem wahrgenommen wird.

Obwohl die Gefahren solcher Vergleiche seit einiger Zeit bekannt sind, hat man im Vereinigten Königreich ganz ähnliche Ausgabedaten zugrundegelegt bei der Bestimmung des „größtmöglichen Nutzens“, der von den Indikatoren für Höchstmögliche Effizienz (Best Value Performance Indicators, BVPIs) bei öffentlichen Bibliotheken angewandt wird. „Seit April 2000 müssen Gemeindeverwaltungen zur Erfüllung der Anforderungen an bestmöglichen Nutzen klare Angaben liefern zu Kosten und Nutzen, und zwar mit Hilfe der gründlichsten und erfolgreichsten Methoden“ (Favret 2000). Bei Bibliotheken werden, neben einer großen Zahl an Kennziffern beim Jahresbibliotheksplan, folgende spezielle BVPIs ermittelt:

- Kosten pro Besuch in öffentlichen Bibliotheken
- Anzahl der Bibliotheksbesuche pro Kopf der Bevölkerung
- Prozentualer Anteil der Nutzer, die den gewünschten Titel fanden
- Anzahl der pro Kopf der Bevölkerung ausgeliehenen Titel.

Favret fährt fort: „Die Regierung wird bei einigen Indikatoren Spitzenwerte festlegen, so daß, allgemein gesagt, die Kosten ganz unten liegen sollen, während Dienstleistungsergebnisse oben liegen müssen.“

Nicht nur muß man sich auf die Integrität, den Wert und die Vergleichbarkeit der für diese Kennziffern verwendeten Daten verlassen können - es gibt noch zwei inhärente Probleme bei dieser Art von Kennziffern. Zum einen ist, trotz der Berücksichtigung des scheinbar eindeutigen Indikators „Prozentualer Anteil der Nutzer, die den gewünschten Titel fanden“, der doch anscheinend das Hauptziel berücksichtigt, das Bibliotheken anstreben sollen, dieser Indikator sehr schwer akkurat meßbar. Zum anderen liegt ein Problem in der Tatsache, daß das Verfahren die Mehrzahl der Bibliotheken auf eine inakzeptable Position in den unteren drei Vierteln der Skala verweist, was bezeichnet wurde als eine Kultur des „Name am Schandpfahl“ (name and shame) und kaum förderlich ist für Bestleistungen. Ein drittes und vielleicht noch bedeutenderes Problem liegt darin, daß kontextuelle Daten, aus der internen und externen Umgebung, bei diesen Kennziffern vollständig fehlen. Daher können diese Kennziffern nur als diagnostische Signale verwendet werden, sie können notwendigerweise

keine sinnvolle Verbindung zu Erfolg, bestmöglichem Verfahren oder höchstmögliche Effizienz garantieren. Diese Themen sind für Bibliothekare im Vereinigten Königreich nicht neu, die selber eine zeitlang über die Bedeutung und die Brauchbarkeit der wichtigsten BVPI-Kennziffern debattiert haben und über die größere Zahl von Indikatoren im Jahresbibliotheksplan.

Neuere Initiativen in Europa zur Einbeziehung elektronischer bibliothekarischer Dienstleistungen beleuchten die gleichen Probleme. Das EQUINOX-Projekt sollte Maßstäbe für elektronische bibliothekarische Dienstleistungen entwickeln und vermutlich den Wertvergleich der Ergebnisdaten ermöglichen, da man gewaltige Anstrengungen unternommen hat, um Übereinstimmung bei den Maßeinheiten selbst zu erzielen und bei der Definition der Daten und ihrer Erhebungsmethoden.⁵ Die entwickelten Leistungsindikatoren betreffen die bekannten Bereiche: prozentuale Höhe der erfaßten Bevölkerung, Zahl der Anfragen an das System, seine individuellen Hilfsmittel und die von der Bevölkerung untersuchten Dokumente, ebenso Kosten pro Verbindungssitzung und von Nutzern besuchte Trainingsveranstaltungen. Geoffrey Ford hat erfolgreich EQUINOX und verschiedene andere im Angebot befindliche Systeme für die Bewertung elektronischer Bibliotheken verglichen (solche wie das e-lib-Projekt, das MIEL -Projekt von Brophy und Wynne, Bertots und McClures Arbeit für die ARL und die „Nutzen für Geld (Value for Money)“-Initiative im Vereinigten Königreich) und kommt bezüglich des „Nutzen für Geld“-Projekts zu dem Ergebnis:

„Ein interessanter Teil der Abschlußuntersuchung bestand in einer Reihe von Vergleichstests aus einem Querschnitt von 20 Einrichtungen ... Diese finanziellen Testergebnisse sind nutzlos für Bewertungen, es ist reine Ansichtssache, ob diese Zahlen hoch oder niedrig sein sollten ... Sie sind nur Verhältniszahlen, die zufällig angefallen sind.

Beispiele für Leistungsvergleichstests und was wir daraus lernen können

Die Menge an Literatur zur Anwendung von Vergleichstests im Bibliotheks- und Informationsbereich belegt die zahlreichen Arten ihrer Nutzung. Abgesehen vom Gesundheitssektor (Marshall und Buchanan 1995, Robbins und Daniels 2001, Todd-Smith und Markwell 2002) und dem Bereich der Spezialbibliotheken (Henczel 2002, Polling 2002) wurden Vergleichsuntersuchungen vor allem zwischen universitären Bibliotheken im Vereinigten Königreich durchgeführt, insbesondere von SCONUL (Society of College, National and University Librarians), das eine kleine Gruppe von Universitätsbibliotheken untersucht, in Australien betreibt dies die gleichartige Organisation CAUL (Committee of University Librarians). Trotzdem ergab eine Untersuchung zur Verwendung von Vergleichstests in den Hochschul- und Spezialbibliotheken des Vereinigten Königreichs (Brockman 1997), daß seinerzeit nur 7,4% der Auskunftgebenden Vergleichstests zu Bewertungszwecken einsetzten, obwohl fast 25% ihre Einrichtung mit anderen verglichen, wobei sie Leistungsindikatoren oder standardmäßige Statistikzahlen benutzten (wie Ausgaben, Erwerbungen und/oder Ausleihvorgänge, Geschäftsvorfälle pro Vollzeitstudent. Die Untersuchung unternahm es dann, drei Institutionen einem Vergleichstest zu unterziehen.

⁵ <http://equinox.dcu.ie/index.html>

Es wurde eine funktionale Definition des Testverfahrens formuliert: „Das Ziel des Vergleichstests ist die Dokumentation und Bewertung eines Schlüsselprozesses und anschließender Vergleich der gewonnenen Daten mit denen aus ähnlichen Prozessen in anderen Organisationen“ (Brockman 1997, 345). Brockman und seine Mitarbeiter arbeiteten fünf Formen von Vergleichstests heraus:

- Wettbewerber (Vergleich mit führenden Organisationen auf demselben Gebiet)
- generisch (Vergleich ähnlicher Geschäftsabläufe, gleich in welchem Tätigkeitsbereich)
- intern (Vergleich interner Abläufe innerhalb einer Organisation)
- funktional (Vergleich ähnlicher Funktionen innerhalb eines Tätigkeitsbereichs)
- kundenorientiert (Vergleich des Ausmaßes, in dem Kundenerwartungen erfüllt oder übertroffen wurden)

Herausgearbeitet wurden zwölf Schritte, die bei einem amerikanischen Projekt über (wissenschaftliche) Spezialbibliotheken entwickelt wurden (Allen 1993), die korrelieren mit Standardverfahren, die Fachleuten zu Fragen des Qualitätsmanagements entwickelt haben (Bullivant 1994, 7). Diese Schritte dienen der Identifizierung, Definition und dem Verstehen des im Vergleich befindlichen Schlüsselprozesses, der Bestimmung eines geeigneten Partners, der Identifizierung, Definition und Erhebung von Daten in beiden Institutionen, ihrer Analyse, Empfehlungen für Veränderungen, Neueinführungen und Überwachung, was alles an Testprojekten unter Einbeziehung der Mitarbeiter beider Institutionen stattfindet. In der Anfangsphase des Prozesses, viel früher als von Bullivant vorgeschlagen und vermutlich früher als im privatwirtschaftlichen Bereich, wo die Verwaltung viel früher einbezogen werden kann und den Testablauf bestimmen kann, findet der wichtige Schritt der Einbeziehung der Verwaltung statt mit starker Beteiligung der Mitarbeiter (Brockman 1997, 39-5). Eine generische Vorgehensweise wurde beim Leistungsvergleichstest verwendet, weil diese „den organisatorischen Bedürfnissen schrittweise angepaßt werden kann“ und den Aufwand (vor allem an Zeit und Mühe) reduziert, die zu jeder Zeit investiert werden müssen. Der Schwerpunkt auf Befragung und Beteiligung der Verwaltung berücksichtigt, daß Vergleichstests mit Wandel zu tun haben, mit Einführung neuer Abläufe, und daß die Durchführung des Tests als Prozess eines Verwaltungswandels zu effektiverem und langandauerndem Wandel führt und zu verbesserter Leistung.

Die an dem Versuch teilnehmenden Organisationen galten bereits als Beispiele für die Anwendung bestmöglicher Verfahren, und die allgemeine Überzeugung lautete hinterher, daß „wenig geändert werden müsse.“ Diese Schlußfolgerung „kein Bedarf an Veränderungen“ war auch eines der Ergebnisse einer Untersuchung von Robbins und Daniels (2001), bei der die Einstellung der Vorgesetzten gegenüber dem Auskunftsdienst in Bibliotheken des Gesundheitswesens erfragt wurde. Robbins und Daniel benutzten eine Variante von SERVQUAL (SERVPERF) als grundlegendes Erhebungs- und Analyseverfahren ihrer Studie - ein Leistungsvergleichstest für Kundenverhalten. Bei beiden Schlußfolgerungen erhebt sich die Frage, ob die Untersuchungen beeinträchtigt wurden durch mangelndes Interesse der Verwaltung am Leistungsvergleich und Veränderungen an der Verwaltungsstruktur; ist das Modell SERVQUAL/SERVPERF für Vergleichstests nicht gut geeignet oder, wie in der Brockman-Studie, wurde der Ablauf des Vergleichstests nicht richtig verstanden oder wurde er nicht richtig durchgeführt?

Kommunikation, Training und mangelndes Verständnis der Verwaltung für den Prozess des Leistungsvergleichstests sind die von Brockman und seiner Gruppe herausgearbeiteten Hauptfaktoren für die verhältnismäßige Erfolglosigkeit des Versuchs. Für ein funktionierendes Testverfahren, so kommentieren sie, müsse die Organisation bereits auf Qualität in der Verwaltung ausgerichtet sein. „Es ist ein Qualitätswerkzeug und daher am sinnvollsten dort einsetzbar, wo Kultur und Praxis bereits auf bestmögliche Verfahren ausgerichtet sind.“ (Brockman 1997, 372). Um das Problem der Partnerwahl zu lösen und Situationen zu vermeiden, in denen der andere Partner hinter dem Initiator in puncto bestmöglichem Verfahren zurückbleibt, wird die Einrichtung einer Datenbank potentieller Testpartner für Leistungsvergleiche vorgeschlagen, „die Angaben dazu macht, welche Organisationen bestimmte Abläufe bewertet haben und welche Ergebnisse herauskamen. Dies würde eine Bestimmung geeigneter Partner ermöglichen und die Ermittlung der ‚Besten der Besten‘“ (Brockman 1997, 357). Foots Definition von Leistungsvergleichstests und gleichwertigen Partnern, die voneinander zu lernen suchen, scheint auf der Strecke geblieben zu sein beim Streben nach Überlegenheit und dem Versuch, Institutionen zu kategorisieren und zu etikettieren auf der Grundlage früherer Leistungen, statt die Ergebnisse in der Zukunft zu verbessern.

Eine Datenbank solcher Vergleichsbibliotheken existiert natürlich bereits bei jenen öffentlichen Bibliotheken im Vereinigten Königreich, die dem IPF Benchmarking Club angehören (der inzwischen über 100 Bibliotheken zählt), allerdings scheinen die Vergleiche sich nur auf vorhandene BVPIs zu beziehen und andere Wertkennziffern im Schema des Jahresbibliotheksplans. Obwohl die Bibliotheken einander konsultieren zwecks Definition und Standardisierung der zu erhebenden Daten, kann doch die CIPFA/Höchstmöglicher-Nutzen-Methode den eigentlichen Nutzen des Vergleichs von Abläufen beeinträchtigen oder den Nutzen von Bewertung durch Fachleute, wie aus dem industriellen Bereich berichtet wird (Camp 1998), wo vor allem, wie schon gesagt, Wert auf Verbesserung der Geschäftsabläufe gelegt wird, nicht Meßwerte für Eingabe und Ergebnis.

Eine sehr erfolgreiche generische Vergleichsstudie zum Thema Fernleihe und Dokumentenlieferungen wurde 2001 von der Nationalen Kooperations-Arbeitsgruppe (National Resource Sharing Working Group) durchgeführt sowie der Australischen Nationalbibliothek, die das ähnlich erfolgreiche System der ARL zur Bewertung von Fernleihvorgängen anwandte. Die australische Studie nutzte statistische Angaben zu Stückkosten, Erfolgsrate (Anzahl der Bestellungen, die ausgeführt werden konnten) und Zeitaufwand für Bestellen und Ausliefern der Dokumente in 97 australischen Bibliotheken. Dramatische Unterschiede bei Kosten und Wirkungseffekten zeigten sich zwischen hohen und niedrigen Durchschnittswerten und zwischen verschiedenen Bibliotheksbereichen. Wie in den Statistiken für amerikanische öffentliche Bibliotheken gab es auch hier keine Bibliotheken, die mit allen ermittelten Wertkennziffern in die oberen zehn Prozent gelangten. Trotz der unverkennbaren Überlegenheit einiger Institutionen scheint es doch bei allen Organisationen Raum für Verbesserungen zu geben, und man kann viel von Anderen, sogar weniger Erfolgreichen, lernen. Die Umfrage war weitreichend und trotzdem streng genug, um die Erforschung der diesen Zahlen zugrundeliegenden Prozesse zu ermöglichen und einen Vergleich zwischen den Institutionen zu ziehen. Gut begründete und auch gut aufgenommene Schlußfolgerungen betreffen eine Reihe von Punkten, die hinsichtlich der Leistungsverbesserung als kritisch angesehen wurden wie Automatisierung, Katalogisierungspraxis, Fortbildung und Kooperationsabkommen zwischen Bibliotheken. Die Studie

benutzte ein erfolgreich erprobtes Verfahren, Fragen von ausreichender Prägnanz, um verschiedene Verfahren aus dem operationellem Bereich und dem Management zu ergründen, und sie wurde durchgeführt unter Berücksichtigung verschiedener Bibliotheksumgebungen und -kulturen - Faktoren, die bei anderen Vergleichsteststudien in dieser Region zu enttäuschenden Ergebnissen führten, wie zumindest anekdotisch vorgestellt wurde. Vielleicht läßt sich hieraus die Lehre ziehen, daß die Bestimmung des bestmöglichen Verfahrens durch Auswertung statistischer Daten eine ausreichend umfangreiche Stichprobe erfordert und große logische Plausibilität, um die Daten sinnvoll auf Stichhaltigkeit hin zu prüfen. Das bestmögliche Verfahren kann nicht erarbeitet werden durch den Vergleich von Eingaben und Ergebnissen zwischen zwei oder drei Einrichtungen. Organisationen, die einen Leistungsvergleich nur mit einer kleinen Zahl anderer Institutionen anstreben, sollten sich auf das „weichere“ Ende des Spektrums konzentrieren und eine Bewertung durch Fachleute durchführen lassen, die ihnen helfen wird, mehr zu erfahren über ihre Vorzüge und Arbeitsabläufe. Solche Beurteilung durch Fachleute wird von den meisten Unternehmen im marktwirtschaftlichen Bereich verwendet, die von weitreichender Zufriedenheit damit berichten (Camp 1998).

Ergebnisse von Leistungsvergleichstests

Die Vereinigung der Forschungsbibliotheken (Association of Research Libraries, ARL), die zur Entwicklung des höchst erfolgreichen australischen Testprogramms beitrug, hat ihren Arbeitsschwerpunkt von Qualität und Bewertung der Bibliotheken verlagert auf das LIBQUAL+ $\hat{\circ}$ -Projekt. Aufbauend auf der Methodologie von SERVQUAL, das seit zwei Jahrzehnten in Management- und Marketingstudien zur Untersuchung der Dienstleistungsqualität benutzt wurde (Zeithaml, Parasuraman und Berry 1990), ist die LIBQUAL+ $\hat{\circ}$ -Methodologie hervorgegangen aus einem Forschungs- und Entwicklungsprojekt, das die ARL im Jahre 2000 in Zusammenarbeit mit der Texas A&M University durchgeführt hat, „um das Mindestniveau, das verspürte und das gewünschte Niveau an Dienstleistungen bei den Nutzern in vier Kernbereichen zu ermitteln: Zugang zur Information, Wirkung der Dienstleistung, Bibliothekambiente und persönliche Betreuung.“⁶ Das Projekt ist von ursprünglich 12 teilnehmenden Institutionen im Jahr 2000 angewachsen auf 164 im Jahr 2002. Obwohl eigentlich entwickelt als Meßverfahren für individuelle bibliothekarische Leistungsqualität und zur Untersuchung von Problemfeldern, die Beachtung erfordern, ist doch die anwachsende Datenbank immer mehr als nationale Datensammlung genutzt worden, bei der einzelne Institutionen ihre Leistungsfähigkeit ermitteln können in einer Art Kundenvergleichstest.

Anders als die australische ILL-Vergleichsteststudie wurde LIBQUAL+ $\hat{\circ}$ kritisiert (Hernon und Nitecki 2001, Cullen 2001), weil es keine Rücksicht nimmt auf die individuellen, kontextuellen und kulturellen Unterschiede von Bibliotheken in stark differierenden Umgebungen und mit verschiedenen Klientelen überall in den Vereinigten Staaten. Zugegebenermaßen wurden Varianten der SERVQUAL bereits mehrfach in verschiedenen bibliothekarischen Umgebungen mit einigem Erfolg eingesetzt (Calvert 2001, Hernon und

⁶ Vgl. die ARL-Internetseite <http://www.arl.org/pr/libqual>

Calvert 1996). Calverts Studie (2001), daß die Dimensionen der von Nutzern erwarteten Dienstleistungsqualität, in diesem Fall Verhalten der Mitarbeiter, das Ambiente der Bibliothek und Einrichtungen, die dem Kunden wirksam beim Auffinden und Auswerten der Medien helfen, bei chinesischen Studenten ebenso anzutreffen sind wie bei neuseeländischen. Doch laufen zur Zeit Versuche, die LIBQUAL+Ö-Methodologie im Vereinigten Königreich anzuwenden, unterstützt von SCONUL aufgrund der Überzeugung, daß das wohlerprobte LIBQUAL+Ö-Untersuchungsinstrumentarium und Methoden der Datenanalyse bei Institutionen im Vereinigten Königreich ebenso anwendbar sind wie schon in Nordamerika. Dies ist vermutlich richtig, doch ob die Daten einfach alle zusammengeführt werden können oder ob die Daten der Studie aus dem Vereinigten Königreich andere Ergebnismuster aufweisen werden aufgrund kultureller Differenzen, ob die stark voneinander abweichenden Bewertungssysteme, die an den Universitäten der Vereinigten Staaten und des Vereinigten Königreiches Verwendung finden, wiederum Bibliotheksnutzer (Studenten und Mitarbeiter) bei der Bewertung der Erhebungsmethoden beeinflussen, das muß noch abgewartet werden. Wenn Berücksichtigung des Umfeldes wichtig ist für Leistungsvergleichstests, werden die neuen LIBQUAL+Ö-Studien im Vereinigten Königreich das hoffentlich bestätigen. Das bedeutet, wenn auch die Dimensionen an Dienstleistungsqualität ein ziemlich festes Gerüst bilden, das in verschiedenen Ländern und Kulturen seine Gültigkeit hat, wäre es doch unangemessen, die Daten, auf denen solche Dimensionen beruhen, für Zwecke des Leistungsvergleichs heranzuziehen.

Einer der interessantesten Aspekte der LIBQUAL+Ö-Methodologie liegt darin, daß keine Eingaben, Abläufe oder Ausgabedaten verglichen werden, sondern die zusätzliche Wirkung der Zufriedenheit mit der bibliothekarischen Dienstleistung, mit ihrer Qualität. Obwohl das „Protokoll zum Höchstmöglichen Nutzen (Best Value Protocol)“ des Vereinigten Königreichs Angaben zum Grad der Zufriedenheit in den Jahresbibliotheksplänen verlangt, weiß man noch zu wenig über die Ursachen der Nutzerzufriedenheit (Hernon und Altman 1998), als daß man diese Zufriedenheit in verschiedenen geographischen, demographischen und kulturellen Umgebungen messen könnte. Das liegt daran, daß es einige Kernpunkte gibt, die für den Leistungsvergleich sehr wichtig sind. Einer davon ist, wie wir gesehen haben, der Kontext. Ein anderer ist die „Plausibilität“ und betrifft die Frage, ob unser Instrumentarium wirklich die vermuteten Faktoren mißt, und ob es andere Variablen gibt, die den Vergleich beeinflussen können und die wir nicht berücksichtigt haben. Weitaus intensivere Untersuchungen unserer Testmethoden sind erforderlich, bevor wir in diesen Dingen Sicherheit erlangen, und die Behauptungen, die die LIBQUAL+Ö-Entwickler hinsichtlich dessen erprobter Plausibilität aufstellten, bleiben umstritten (Hernon und Nitecki 2001). Um die Frage der Plausibilität zu untersuchen, sind sowohl empirische als auch ethnographische Methoden erforderlich, um potentiell irrtümliche Annahmen auszuschalten und die kontextuellen und kulturellen Unterschiede zu erhellen, die sowohl die Nutzererwartungen beeinflussen als auch die Leistung der Bibliothek. Alle Methoden des Leistungsvergleichs und Verfahren der Datenerhebung müssen einer intensiven Prüfung unterzogen werden, bevor wir wirklich Vertrauen in sie setzen können.

Die Frage nach der Plausibilität berührt eine Reihe von Gebieten, auf denen wir Leistungsvergleichstests gerne entwickelt sähen. In diesem Zusammenhang sind verschiedene Dinge zu beachten. Zum einen neigen wir dazu, wie schon früher gesagt, auf Verhältniszahlen zwischen Messungen zu vertrauen, einfach weil sie vorhanden sind, ohne zu prüfen, ob diese

Zahlen irgendeine Gültigkeit haben oder ob das Verhältnis tatsächlich authentisch existiert. Die Daten, ihre Verhältniszahlen untereinander und andere zu Vergleichszwecken benutzte Angaben müssen einem klaren Verständnis der zugrundeliegenden Konstruktion entspringen und nicht der Gewohnheit, gerade zur Hand befindliche Zahlen zu benutzen, wie Ford es nennt.

Zweitens: Wenn wir notwendige Veränderungen durch einen Leistungsvergleichstest herausfinden wollen, müssen wir eine vernünftige Kombination von Datenvergleich (vor allem für diagnostische Zwecke) und Ablaufvergleich anwenden, so daß sinnvolle Urteile gefällt werden können hinsichtlich der Bedeutung der Daten in ihren Zusammenhängen und der Abläufe, die zur Entstehung dieser Daten führen.

Drittens ist es meine feste Überzeugung, wenn Leistungsvergleichstests für Manager sinnvoll sein sollen, kann man sie nicht von anderen Formen der Bewertung und Leistungsmessung isolieren, die in der jeweiligen Institution angewandt werden. Heute liegt der Schwerpunkt bei Leistungsmessung nicht auf Eingabe, Verarbeitung und Ausgabedaten, wie in der Vergangenheit, sondern auf dem Resultat. Resultate sind bisher auf allen bibliothekarischen Gebieten schlecht definiert, und wenn man sie zumindest definieren kann, sind doch keine befriedigenden Instrumente zu ihrer Messung entwickelt worden. Wenn aber Leistungsvergleichstests ein wirklich unabdingbares Werkzeug werden sollen, müssen wir auch die Resultate in die Vergleichsaktivitäten einbeziehen - vielleicht am „weicheren“ Ende des Spektrums durch Fachgutachten.

Ein Vergleichstestmodell für Nationalbibliotheken

Wenn wir uns jetzt den Nationalbibliotheken zuwenden und ihrer Verwendung von Leistungsvergleichstests zwecks Bewertung, so sind einige Besonderheiten zu beachten. Erstens bilden Nationalbibliotheken einen sehr speziellen Sektor mit festgelegten nationalen Funktionen, die von Land zu Land variieren und politischen Vorgaben unterliegen, die sich häufig schneller ändern als bei öffentlichen oder universitären Bibliotheken. Die Bedeutung des nationalen und politischen Umfeldes hat es ihnen erschwert, ebenbürtige Partner für Vergleichstests zu finden. Auch ist es schwierig für sie, zu partizipieren an Datenbanken von hinreichender Größe, um traditionelle Datenvergleiche zu betreiben oder das bestmögliche Verfahren zu bestimmen, obwohl dieses Verfahren, gesehen als allgemeines Verhalten in allen bibliothekarischen Dingen, natürlich weiterhin von Bedeutung sein wird und es gute Möglichkeiten gibt für allgemeinen Leistungsvergleich, wobei man die verschiedenen Dienstleistungen mit denen ähnlich aufgebauten Institutionen vergleicht, wie bei dem oben genannten australischen Beispiel gezeigt wurde. Sie können auch internationale Vergleiche anstellen wie bei Personalentwicklung, Personalführung und Rechnungswesen, oder Vergleiche mit anderen Regierungsbehörden auf den genannten Gebieten betreiben, sowie Vergleichsuntersuchungen unter den Nutzern. Man sollte sich ins Gedächtnis rufen, daß Leistungsvergleichstests im privatwirtschaftlichen Bereich wichtiger sind als Mitarbeiterfortbildung, Personalentwicklung, Informationstechnologie, Dokumentenkontrolle und Aktenverwaltung, interne und externe Kommunikation und daß man auf der Suche nach Fachleuten sich auch jenseits des Bibliothekswesens an jede Organisation wenden kann, die Erfahrungen zu vermitteln hat. Viele führende Unternehmen wären froh über die Gelegenheit, Vergleichstests und Fachgutachten zu geschäftlichen Vorgängen anzustellen mit einer so

großen und prestigeträchtigen Institution wie einer Nationalbibliothek.

Nationalbibliotheken haben weitaus größere Probleme, Leistungsmeßverfahren zu entwickeln und in Übereinstimmung zu bringen mit den von einer Regierung vorgegebenen Zielen wie z.B. einen Beitrag zum Bildungswesen zu leisten, der Bevölkerung in einer bestimmten Weise zu nützen, kulturelles Erbe zu bewahren und Wissen zu verbreiten. Wer sind ihre Kunden, was ist die Zielgruppe, welche Art von Ausgabedaten werden erwartet und auf welchem Niveau? Noch schwieriger: Was werden von ihnen für Resultate erwartet? Es ist schon schwierig genug, die Arbeitsergebnisse universitärer Bibliotheken zu bestimmen mit Hilfe studentische Lernerfolge oder Forschungsbeiträge der Mitarbeiter oder den Nutzen öffentlicher Bibliotheken zu bestimmen anhand der Förderung ständigen Lernens, Geschäfts- oder Berufserfolgen oder Bereicherung durch Freizeitlektüre, so sind die Resultate bei Nationalbibliotheken noch schwieriger einzugrenzen und zu messen. Aber dennoch müssen wir gerade darauf hinzielen, wenn Leistungsvergleichstests mit Erfolgsmessung Schritt halten wollen.

Wenn sie nur bereit sind, Erfahrungen aus anderen Gebieten zu übernehmen, können Nationalbibliotheken einen Prozess der Expertenbegutachtung in Gang setzen, um Gedankenaustausch zu betreiben bei der Entwicklung von Erfolgsmessungen und -ergebnissen auf diesem schwierigen Gebiet in Zusammenarbeit mit anderen staatlichen Dienststellen und Nichtregierungsorganisationen, die mit einem ähnlichen Spektrum an Sozial- und Bildungsthemen arbeiten. Beispielsweise könnte man die verschiedenen Rollen untersuchen, die eine Nationalbibliothek für ihre verschiedenen Klientelen spielt und Vergleichspartner suchen, die sich in ähnlicher Lage befinden. Und wie schon früher gesagt, ist das Thema Plausibilität ein kritischer Punkt, dem Regierungen bisher nicht viel Aufmerksamkeit geschenkt haben. Wenn man die Probleme betrachtet, die Kunden durch Benutzung einer Nationalbibliothek zu lösen suchen, ob in persona oder virtuell, und die Erfolge bei Erreichung ihrer Ziele sowie die Bedeutung, die dieser Erfolg für ihr Leben und ihre Arbeit hat, wird dies zuerst durch ethnographische Methoden verifiziert und durch empirische Studien untermauert werden müssen, was zu einer Grundlage verhilft, auf der man allmählich eine Konstruktion zur Ergebnisbewertung errichten kann. Partnerschaft und Dialog mit Fachleuten im Bibliotheksbereich, in Regierungs- und Nichtregierungsorganisationen und im privatwirtschaftlichen Bereich können alle bei diesem Prozess nützlich sein.

Für eine Nationalbibliothek kann ein Leistungsvergleichstest daher nicht so sehr eine Aufgabe des Vergleichs von Statistiken zu Ergebnisdaten mit denen anderer großer National- und Staatsbibliotheken aus verschiedenen Kulturen sein, die alle verschiedene Vorgaben haben, sondern vielmehr eine Qualitätsanalyse innerhalb eines bekannten kulturellen Kontexts und das Bemühen, durch unterstützende Fachgutachten zu ständiger Offenheit gegenüber Wandel und Verbesserung zu gelangen. Das bestmögliche Verfahren bei Leistungsvergleichstests müßte demnach bieten

- einen Schwerpunkt auf dem Vergleich von Arbeitsprozessen und fachlicher Begutachtung durch andere Behörden und Organisationen derselben Kultur;
- die Erarbeitung, durch Fachgutachten, von Leistungsbewertungsverfahren, die Eingabe, Verarbeitung, Ausgabedaten und Resultate berücksichtigen
- Bereitschaft zur Änderung von Verwaltungsabläufen) zwecks Sicherstellung von Reformen in der Organisation
- eine wissenschaftliche Grundlage, die über Plausibilität verfügt und kulturelle

sowie organisatorische Kontexte berücksichtigt

Bei Einbeziehung dieser Grundsätze können Leistungsvergleichstests ein sehr wirksames Verfahren werden, um Nationalbibliotheken bei der Anpassung an eine sich wandelnde Umwelt zu helfen und gleichzeitig die Qualitätsverbesserung im Blick zu behalten.

Literaturhinweise

- Allen, Ferne. 1993. Benchmarking: practical aspects for information professionals. *Special Libraries* 84(3) Summer, 123-140.
- Brockman, John et al. 1997. *Quality Management and Benchmarking in the Information Sector*. London: Bowker Saur.
- Bullivant, John. 1994. *Benchmarking for continuous improvement in the public sector*. Harlow: Longman.
- Calvert, Philip J. 2001. International variations in measuring customer expectations. *Library Trends* 49(4): 718-731.
- Camp, Robert C. 1998. *Global cases in benchmarking: best practice from organizations around the world*. Milwaukee, Wisc: ASQ Quality Press.
- Favret, Leo. 2000. Benchmarking, annual library plans and best value: the implications for public libraries. *Library Management* 21(7): 340-348.
- Ford, Geoffrey. 2001. Theory and practice in the networked environment: a European perspective. In *Evaluating Networked Information Services: techniques, policies, and issues*. Charles McClure and John Carlo Bertot, eds. 1-22. Medford: NJ: Information Today.
- Foot, J. 1998. *How to do benchmarking: a practitioner's guide*. London: Inter-Authorities Group.
- Henczel, S. 2002. Benchmarking: measuring and comparing for continuous improvement. *Information Outlook* 6(7): 12-14, 17-18.
- Hernon, Peter and Ellen Altman. 1998. *Assessing service quality: satisfying the expectations of library customers*. Chicago: American Library Association.
- Hernon, Peter and Calvert Philip J. 1996. Methods for measuring service quality in academic libraries in New Zealand. *Journal of American Librarianship*. 22(5): 387-391.
- Hernon, Peter and Danuta A. Nitecki. 2001. Service quality: a concept not fully explored. *Library Trends* 49(4): 687-708.
- Marshall, Joanne G. and Holly Shipp Buchanan. 1991. Benchmarking reference services: an introduction. *Medical Reference Services Quarterly* 14. Fall: 59-73.
- National Resource Sharing Working Group. 2001. *Interlibrary Loan and Document Delivery Benchmarking Study*. Canberra: National Library of Australia.
- Polling, N. 2002. Ahead or behind the curve ... where do you stand? *Information Outlook* 6(7): 22-25.
- Pritchard, Sarah. 1995. Library benchmarking: old wine in new bottles. *Journal of Academic Librarianship* 21(6) November: 491-495.
- Pratt, Alan and Ellen Altman. 1997. Live by the numbers, die by the numbers. *Library Journal* April 15: 48-49.
- Robbins, Kathryn and Kathleen Daniels. 2001. Benchmarking reference desk service in academic health science libraries: a preliminary survey. *College and Research Libraries* 62(4): 348-354.

- Todd-Smith, B. and L. G. Markwell. 2002. The value of hospital library benchmarking: an overview and annotated references. *Medical Reference Services Quarterly* 21(3): 85-95.
- Zeithaml, V. A., Parasuraman, A. and L. Berry. 1990. *Delivering service quality: balancing customer perceptions and expectations*. New York: The Free Press.