Economics and usage of digital libraries: byting the bullet

II. Pricing Electronic Access to Knowledge: The PEAK Experiment

6. User cost, usage and library purchasing of electronically-accessed journals

Robert S. Gazzale and Jeffrey K. MacKie-Mason

6.1 Introduction

Electronic access to scholarly journals has become an important and commonly accepted tool for researchers. Technological improvements in, and decreased costs for, communication networks and digital hardware are inducing innovation in digital content publishing, distribution, access and usage. Consequently, although publishers and libraries face a number of challenges, they also have promising new opportunities. [1] Publishers are creating many new electronic-only journals on the Internet, while also developing and deploying electronic access to literature traditionally distributed on paper. They are modifying traditional pricing schemes and content bundles, and creating new schemes to take advantage of the characteristics of digital duplication and distribution.

The University of Michigan operated a field trial in electronic access pricing and bundling called "Pricing Electronic Access to Knowledge" (PEAK). We provided a host service providing access to roughly four and a half years of content (January 1995 - August 1999) including all of Elsevier Science's approximately 1200 scholarly journals. Participating institutions had access to this content for over 18 months. [2] Michigan provided Internet-based delivery to over 340,000 authorized users at twelve campuses and commercial research facilities across the U.S. The full content of the 1200 journals was received, catalogued and indexed, and delivered in real time. At the end of the project the database contained 849,371 articles, and of these 111,983 had been accessed at least once. Over $500,000 in electronic commerce was transacted during the experiment. For further details on this project, including the resources needed for implementation, see Bonn et al. (this volume).

We elsewhere describe the design and goals of the PEAK research project (MacKie-Mason and Riveros (2000)). In MacKie-Mason et al.

(2000) we detail the pricing schemes offered to institutions and individual users. We also report and analyze usage statistics, including some data on the economic response of institutions and individuals to the different price and access options.

In this paper, we focus on an important behavior question: how much does usage respond to various differences in user cost? We pay careful attention to the effect of both pecuniary costs and non-pecuniary costs such as time and inconvenience.

An interesting aspect of the PEAK project is the role of the library as economic intermediary and the effects of its decisions on the costs faced by end users. [3] In the first stage of the decision process, the library makes access product purchasing decisions. These decisions then have a potentially large effect on the costs that users face in accessing particular electronic journal articles, whether it be the requirement that users obtain and use a password or pay a monetary cost. The consumer then decides whether she will pay these costs to access a given article.

The standard economic prediction is that a user will access an article if the marginal benefit she expects from the article (i.e. the incremental value) is greater than her marginal cost. Different users are going to have different valuations for electronic access to journal articles. Furthermore, even the same user will not place the same value on all requested articles. Information regarding users' sensitivity to user cost (known to economists as the elasticity of demand) for various articles is important to an institutional decision-maker who wants to maximize, or at least achieve a minimally acceptable level of user welfare. [4] Demand elasticity information is also vital to firms designing access options and systems because design decisions will affect non-pecuniary costs faced by the users, and thus overall demand for access.

It is well known that the usage of information resources responds to the monetary cost users bear. We find that even modest per article fees drastically suppressed usage. It is also true, but perhaps less appreciated, that non-pecuniary costs are important for the design of digital information access systems. We find that the number of screens users must navigate, and the amount of external information they must recall and provide (such as passwords), have a substantial impact on usage. We estimate the amount of demand that was choked off by successive increases in the user cost of access. Further, we find preliminary evidence that users were more likely to bear these costs when they are expected. Finally, given the access options and prices offered in the PEAK experiment, we calculate the least costly bundles of access options an institution could have purchased to meet the observed usage, and compare this to the actual bundles purchased in each year. From this comparison we learn about the nature of institutional forecasting errors, and the potential cost savings to them from the detailed usage information of the sort provided by PEAK.

6.2 Access options offered

To choose which access products (and their prices) to offer PEAK participants, we balanced a complex set of considerations. These included the desire to study innovative access options, the desire to create substantial experimental variation in the data, and the need to entice institutions to participate. Hunter (this volume) gives a fuller account of these deliberations. In the end, participating institutions in the PEAK experiment were offered packages containing two or more of the following three access products:

1. *Traditional Subscription:* Unlimited access to the material available in the corresponding print journal.

2. *Generalized Subscription:* Unlimited access (for the life of the project) to any 120 articles from the entire database of currently priced content. Articles are added to the generalized subscription package as users request articles that were not already otherwise paid for, until the subscription is exhausted. [5] Articles selected for generalized subscriptions may be accessed by all authorized users at that institution.

3. *Per Article:* Unlimited access for a single individual to a specific article. If an article is not available in a subscribed journal, nor a generalized subscription, nor are there unused generalized subscription tokens, then an individual may purchase access to the article, but only for his or her use (for the life of the project).

The per-article and generalized-subscription options allow users to capture value from the entire corpus of articles, without having to subscribe to all journal titles. Once the content is created and added to the server database, the incremental delivery cost (to the publisher and system host) is approximately zero. Therefore, to create maximal value from the content, it is important that as many users as possible have access. The design of the pricing and bundling schemes affect both how much value is delivered from the content (the number of readers) and how that value is shared between the users and the publisher.

Generalized subscriptions may be thought of as a way to pre-pay (at a discount) for interlibrary loan requests. One advantage of generalized subscription purchases is that the "tokens" cost substantially less per article than the per-article license price. Institutions did, however, need to purchase tokens at the beginning of a year and thus bore some risk. There is an additional benefit: unlike an interlibrary loan, all users in the community have ongoing unlimited access to the articles obtained via generalized subscription token. To the publisher, generalized subscriptions represent a committed flow of revenue at the beginning of each year, and thus shift some of the risk to the users. Another benefit to the publisher, as noted by Hunter (this volume), is that that they open up access to the entire body of content to all users. Generalized subscriptions thus offer one method for the publisher to increase user value from the already produced content, and which creates an opportunity to obtain greater returns from the publication of that content.

Table 6.1: Access models

Institution ID	Group	Traditional	Generalized	Per Article
5, 6, 7, 8	Green		X	X
3, 9, 10, 11, 12	Red	X	X	X
13, 14, 15	Blue	X		X

NOTE: An "X" indicates that this option was available to the institutions listed in that row of the table.

Participating institutions were assigned randomly to one of three different experimental treatments, which we labeled as the Red, Green and Blue groups. Institutions in every group could purchase articles on a per-article basis. Those in the Green group could purchase generalized subscriptions, while those in the Blue group could purchase traditional subscriptions. Institutions in the Red group could purchase all types of access. Twelve institutions participated in PEAK: large research universities, medium and small colleges and professional schools, and corporate libraries. Table 6.1 shows the distribution of access models and products offered to the participating institutions.

6.3 Summary of user costs

The PEAK experiment was designed to assess user response to various pricing and access schemes for digital collections. Since the content was traditional refereed scholarly literature, we implemented access through the traditional intermediary: the research library. The reliance on research libraries affected the design of the experiment and thus the research questions we could investigate. As we noted above, the intermediary, by choosing the combination of access products available to users, determines the costs faced by its users. The individual users then make article-level access decisions. [6] Thus, there are two different decision makers playing a role in access decisions. We must take both into account when analyzing the usage data.

When confronted with the PEAK access options and prices, nearly all of the participating libraries purchased substantial prepaid (traditional or generalized subscription) access on behalf of their users. As a consequence, relatively few users were faced with the decision of whether or not to pay a pecuniary charge for article access. Although we measured over 200,000 unique individual uses of the system, we estimate that a user was asked to pay a pecuniary cost in only about 1200 instances. Therefore we focus as much on user response to non-pecuniary costs as to pecuniary costs.

Access at zero user cost. Substantial amounts of PEAK content were available at zero user cost. This content included:

- all "unmetered" content, which included articles published at least two calendar years prior as well as all non-full-length articles;

- articles in journals to which the institution purchased an electronic traditional subscription; and

- articles which had previously been purchased by a user at the institution with a generalized subscription token.

All such access required authentication, but this was most often accomplished automatically by system verification that the user's workstation IP address was associated with the participating institution. Thus, most such authentications required no user time, effort or payment, and the overall marginal user cost per access was zero. [7]

Access at medium user cost. For some access, users incurred a higher cost because they were required to enter a password. The transactions cost of password entry ranged from small to substantial. In the worst case, the user needed to navigate elsewhere in the system to fill out a form requesting a password, and then wait to receive it via e-mail. Once received, the user had to enter the password. If the user previously obtained a password, then the only cost to her was to find or recall the password and enter it. Content accessible via password entry included:

- articles in journals to which the institution did not have a traditional subscription, assuming that the institution had generalized tokens available;

- subsequent access to an article which an individual previously purchased on a per-article basis.

Access at high user cost. If the institution did not have any unused generalized subscription tokens, then content not available at zero cost could be accessed by payment of a $7 per-article fee. The user who wished to pay the per-article fee would also bear two non-pecuniary costs: (1) password recall and entry, as above for the use of a generalized subscription token, and (2) credit card recall and entry. [8] In many cases, institutions subsidized, either directly or indirectly, the per-article fee. Although subsidized, access of this type still resulted in higher transactions costs. In the indirect subsidy case, a user needed to submit for reimbursement. In the direct case, except at institution 15, users needed to arrange for the request to be handled by the institution's interlibrary loan department.

Exceptions. Several of the access procedures—and thus users' costs —were different at institutions 13 and 14. At both, per-article access for all requests was paid (invisibly to the user) by the institution, so users never faced a pecuniary cost. [9] At institution 14, a user still faced the non-pecuniary cost of finding her password and entering it to access "paid" content. [10] However, all users at institution 13 accessing from associated IP addresses were automatically authenticated for all types

of access. Thus users at institution 13 could access all PEAK content at zero total (pecuniary *and* non-pecuniary) cost. These differences in access procedures were negotiated by the production and service delivery team during the participant acquisition phase, with the approval of the research team. In our analyses below we use the differences in user cost between these two institutions and the others as a source of additional experimental variation.

Complexity. From the description above, it might appear that the PEAK access program was much more complicated than one would expect to find in production services. If so, then our results might not generalize readily to these simpler production alternatives.

In fact, most of the complexity is at the level of the experiment, and as such creates a burden on us (the data analysts), and on readers, but not on users of the PEAK system. Because this was an experiment, we designed the program to have different treatments for different institutions. We had to keep track of these differences, but users at a single institution did not need to understand the full project (indeed, they were not explicitly informed that different variations of PEAK were available elsewhere). In most cases they did not even need to understand all three access options, because most institutions had only two options available to them.

Among our three access options, the traditional subscription and per-article fee options were designed to closely mimic familiar access schemes for printed journals, and as such they did not cause much confusion. The generalized subscription was novel, but the details largely were transparent to end users: they clicked on an article link, and either it was immediately available, or they were required to enter a password, or they were required to pay a per-article fee. Whether the article was available through a traditional or generalized subscription was not relevant to individual users. Thus, to the user the access system had almost identical complexity to existing systems: either an article is available in the library or not, and if not the user can request it via interlibrary loan (and/or with a per-article fee from a document delivery service).

The librarians making the annual PEAK purchasing decisions needed to understand the differences between traditional and generalized subscriptions of course. We prepared written explanatory materials for them, and provided pre-purchase and ongoing customer support to answer any questions. In section 6.6 below we discuss some evidence on how learning about the system changed behavior between the first and second year, but we did not observe any significant effects we could attribute to program complexity.

6.4 Effects of user cost on access

In this section, we measure the extent to which user costs to access PEAK content affected the quantity and composition of articles actually accessed. Clearly the costs and benefits of accessing the same

information via other means, particularly via an institution's print journal holdings, will have an enormous impact on a user's willingness to bear costs associated with PEAK access. We do not explicitly model these costs, although we do control for them at an institutional level. Kingma (this volume) provides estimates of some costs associated with information access via several non-electronic media.

As noted above, user costs for accessing PEAK content depended on a variety of factors. One factor is the type of content requested ("metered" versus "unmetered"). Looking only at metered content, the pecuniary and non-pecuniary costs associated with access depended in large part on the access products purchased by a user's institution. Further, the access costs faced by users within a given institution depended on the specific products selected by an institution (i.e. the specific journals to which an institution holds a traditional subscription, and the number of generalized subscription tokens purchased), individual actions (whether a password had already been obtained) and also on the actions of other users at the institution (whether a token had already been used to purchase a requested article, and how many tokens remain). In the following sections, we estimate the effects of these incremental costs on the quantity and composition of metered access.

Non-pecuniary costs

To gauge the impact of user cost of usage on aggregate institutional access, we compared the access patterns of institutions in the Red group with those in the Blue group. Red institutions had both generalized and traditional subscriptions available; Blue had only traditional. Users at both institutions could obtain additional articles at the per-article price. We constructed a variable we call "Normalized Paid Accesses" to measure the number of "paid" accesses to individual articles (paid by generalized tokens or by per-article fee) per 100 unmetered accesses, normalized to account for the number of traditional subscriptions. Adjusting for traditional subscriptions accounts for the amount of prepaid content provided by the user's institution; adjusting for unmetered accesses adjusts for the size of the user community and the underlying intensity of usage in that community. [11]

Table 6.2:Normalized paid access per 100 unmetered accesses, by institution.

Institution	Access group	Normalized paid accesses per 100 unmetered accesses
3	Red	13.5
9	Red	20.4
10	Red	31.7
11	Red	7.59
12	Red	26.4

Average	Red	15.1
13	Blue	51.0
14	Blue	15.1
15	Blue	4.72

NOTE: Average not reported for Blue institutions because of variations in experimental conditions; see text for details.

We use our statistic, Normalized Paid Accesses, as a measure of relative (cross-institution) demand for paid access. We present the statistic in Table 6.2. Even after controlling for the size of an institution's subscription base and the magnitude of demand for unmetered content, paid demand differed among institutions with the same access products. This suggests that there are institution-specific attributes affecting demand for paid access. It is also possible that we incompletely control for subscription size. One possibility is that the number of traditional subscriptions affects the cost a user expects to have to pay for an article before the actual cost is realized. Users at an institution with a large traditional subscription base, such as institution 3, would have had a lower expected marginal cost for access as a large percentage of the articles are accessible at zero cost. Some users at these institutions might attempt to access articles via PEAK, expecting them to be free, while not willing to pay the password cost when the need arises. This difference between *expected* and *actual* marginal cost may be important; we return to this point later.

We can make some interesting comparisons between institutions in the Red group and those in the Blue group. While institution number 13, as a member of the Blue group, only had traditional subscriptions and per-article access available, users at this institution did not need to authenticate for any content, and thus faced no marginal cost in accessing any paid content. Most users at Red institutions faced the cost of authenticating to spend a token. [12] We would therefore expect a higher rate of paid access at institution 13, and this is in fact the case.

Paid access at institution 14 was similarly subsidized by the institution. However, in contrast to institution 13, authentication was required. Thus the marginal user cost of paid access at institution 14 was exactly the same as at the Red institutions. We therefore expected that demand for paid access would be similar. This is in fact the case: Normalized Paid Access is 15.1 at both. Finally, per-article access for users at institution 15 was not automatically subsidized. Thus, users faced very high marginal costs for paid content. In addition to the need to authenticate with a password, users at this institution needed either to: a) pay the $7.00 per-article fee and enter their credit card information; or b) arrange for the request to be handled via the institution's interlibrary loan department. In either case, the user cost of access was higher than password only, and, as we expected, the rate of paid access was much lower than in

the Red group.

Table 6.3: Estimated effects of user cost on access.

	No month dummies	Month dummies
Constant	87.535*	108.615*
	(10.394)	(14.643)
Blue: Credit Card (Inst. 15)	-280.490*	-270.879*
	(37.627)	(35.508)
Red + Institution 14	-58.999*	-57.764*
	(7.900)	(7.186)
Out of Tokens	-25.070*	-25.665*
	(1.635)	(2.533)
Graduate Students/Faculty Ratio	43.821*	41.748*
	(7.301)	(6.912)
Percentage Engineering, Science and Medicine	-225.913*	-215.767*
	(7.535)	(36.553)
Sample Size	530	530
R^2	0.171	0.229

NOTE: Standard errors are shown in parentheses.
Dependent variable is weekly normalized paid access per 100 free accesses.
* Significant at the 99% level.

Table 6.3 summarizes the results from a multiple regression estimate of the effects of user cost on access. We controlled for differences in the graduate student / faculty ratio and the percentage of users in Engineering, Science and Medicine.[13] The dependent variable, *Paid accesses per 100 unmetered accesses*, controls for learning and seasonality effects. We thus see the extent to which paid access, starting from a baseline of access to paid content at zero marginal user cost, falls as we increase marginal costs. Imposition of a password requirement reduces paid accesses by almost 60 accesses per 100 unmetered accesses (Red and institution 14), while the depletion of (institution-purchased) tokens results in a further reduction of approximately 25 accesses (per 100 unmetered).

We use the distinction between metered and unmetered access to further test the extent to which increased user costs throttle demand. As a reminder, full-length articles from the current year are metered: either the institution or the individual must pay a license fee to gain access. Other materials (notes, letters to the editor, tables of contents, and older full-length articles) are not metered: anyone with institutional access to the system can access this content after the institution pays the institutional participation license fee. Some of the unmetered content comes from journals that are covered by

traditional subscriptions, some from journals not in subscriptions. We calculate the ratio of this *free* content accessed from the two subsets of content. If we make the reasonable assumption that, absent differential user costs, the ratio of metered content from the two subsets would be the same as the ratio of unmetered content, then we can estimate what the demand would be for metered content outside of paid subscriptions if that content were available at zero user cost (e.g., if the institution added the corresponding journals to its traditional subscription base). Our estimate is calculated as:

$$\text{predicted access}_{\text{metered, not subscribed}}$$

$$= \text{access}_{\text{metered, subscribed}} \times \frac{\text{access}_{\text{unmetered, not subscrib}}}{\text{access}_{\text{unmetered, subscribed}}}$$

Table 6.4: Paid access as percentage of average predicted for zero user cost.

Institution	Year	Actual Per Predicted	Percent Free Access Psswd. Authent.	Credit Card Required	Password Entered When Prompted
3	1998	21.1%	11.1%	0	6.69%
10	1998	146.2%	45.4%	0	13.5%
11	1998	16.4%	8.81%	0	2.6%
12	1998	83.3%	51.7%	0	7.14%
13	1998	125.9%	98.8%	0	100.0%
14	1998	79.3%	54.5%	0	44.4%
15	1998	0.00%	22.2%	1	8.06%
3	1999	31.4%	19.1%	0	10.4%
10	1999	123.4%	43.9%	0	13.4%
11	1999	20.8%	18.5%	0	14.1%
13	1999	77.7%	100.0%	0	100.0%
14	1999	56.7%	63.2%	0	17.8%
15	1999	19.5%	12.2%	1	2.39%

"Percent free access password authenticated" indicates the percentage of times that users accessing free material were already password authenticated (which isn't in fact necessary for free accesses).
"Credit card required" means the user was required to pay a per-article fee.

In Table 6.4 we present *actual* paid access (when customers face the actual user cost) as a percentage of predicted access (at zero user cost) for all institutions that had traditional subscriptions in a given year. All observations except three (institutions 10 and 13 in 1998,

and institution 10 in 1999) show actual access substantially below predicted when users bear the actual user cost. We conjecture that the surprising result for institution 10 might be partially due to the fact that they had the fewest traditional subscriptions. Because relatively little was available at zero user cost, users at this institution might have expected to bear the user cost (password recollection and entry in this case) for *every* access. If this were the case, then our method of predicting access at zero user cost is biased and the results for institution 10 are not meaningful. As for institution 13, recall that its users in fact faced no incremental user cost to access paid materials. We thus expect its paid accesses to be closer to that predicted for zero user cost, and are not surprised by this result.

Though not related to our focus on user cost, there are two other statistical results reported in Table 6.4 that bear mention. First, usage is substantially, and statistically significantly higher when the graduate student / faculty ratio is higher. It is not implausible that graduate students make more frequent use of the research literature, reading more articles while taking classes and working on their dissertations, than more established scholars. This may also reflect life cycle differences in effort and productivity. However, it is also possible that a higher graduate student ratio is proxying for the intensity of research (by both graduate students *and* faculty) at the institution, which would be correlated with higher access.

The other, and more surprising result is that the higher is the percentage of engineering, science and medicine (STM) users, the lower is usage, by a large and statistically significant amount. We cannot be sure about the interpretation of this result, either. We were surprised because the Elsevier catalogue is especially strong in STM, reflected in breadth, depth and quality of content. Perhaps the nature of study and research in STM calls for less reading of journal articles, but this conjecture cannot be tested without further data.

For all other institutions we generally see that the user costs associated with paid access caused an appreciable reduction in the number of paid articles demanded. We also present in Table 6.4 factors which we believe help explain this shortfall, namely the percentage of free access that is password authenticated, whether or not a credit card is required for all paid access, and the rate at which passwords were entered for paid access when prompted.

Table 6.5: Estimation results of effects of user cost on actual paid accesses as percent of predicted accesses

Independent variable	Coefficient (standard error)
Percent Free Psswd. Auth.	2.12*
	(.45)
Prompted Login Percent	-1.05**
	(.54)

Credit Card Required	-.213
	(.25)
Sample Size	13
R^2	0.85

NOTE: Standard errors shown in parentheses. Dependent variable is actual paid access as a percentage of predicted.
*Significant at the 99% level; **Significant at the 95% level.

In Table 6.5 we summarize the results from the estimation of the effects of user cost on actual paid access as a percentage of predicted accesses. Despite the small sample size, the results clearly demonstrate that, as we increase the number of individuals who can access paid content without additional marginal costs (proxied by the percent of free access that is password authenticated, which indicates that the password user cost has already been incurred), more paid access is demanded. The dummy variable for credit card required (for per-article payment) is not significant, but there was almost no variation in the sample from which to measure this effect. [14] The coefficient for the percent of prompted users who log in is of the wrong sign to support our hypothesis: we expected that the higher the number of users who are willing to bear the non-pecuniary costs of login, the higher would be the access to paid material.

Pecuniary costs

If an institution did not purchase any, or depleted all of its tokens, a user wanting to view a paid article not previously accessed had three choices. [15] She could pay $7.00 to view the article, and also incur the non-pecuniary cost of entering credit card information and waiting for verification. If the institution subscribed to the print journal, she could use the print journal article rather than the electronic product. She could also request the article through a traditional interlibrary loan, which also involves higher non-price costs (effort to fill out the request form, and waiting time for the article to be delivered) than spending a token. [16]

Due to details of the system design, we are unable to determine the exact number of times that users were faced with the decision of whether or not to enter credit card information in order to access a requested article. We *were* able to identify in the transaction logs events consistent with the credit card decision (hereafter we call these "consistent events"). These consistent events are, however, a noisy signal for the actual number of times users faced this decision.

We used evidence from the experimental variation to estimate the actual rate of requests for credit card payment. In some months some institutions had unused tokens and thus there were *no* credit card (per-article) purchases, since unused tokens are always employed first. For these months we divided the number of

consistent events by the number of access requests handled by the system for that institution, to obtain a measure of the baseline rate of consistent events that are not actual credit card requests. For each institution that did deplete its supply of tokens, we then subtracted this estimated baseline rate from the total number of consistent events to measure requests for credit card payment. For institutions that never had tokens, we use the weighted average of the estimated baseline rates for institutions with tokens.

Table 6.6: Credit card payments as a percent of requests, estimated from transaction log evidence

Institution	Estimated Credit Card Requests	Credit Card Payments	Percent
3	53	13	25.5%
6	260	194	74.6%
9	190	1	0.5%
11	562	61	10.9%
15	137	73	53.3%

In Table 6.6 we present the number of actual payments as a percent of estimated requests for credit card payments. The relative percentages are consistent with our intuition. Institutions 6 and 15 never had any tokens. We thus expect that users at these institutions expected a relatively high cost of article access, and would not bother accessing the system or searching for articles if they were not prepared to pay fairly often.[17] Among the institutions at which tokens were depleted, the payment rate is appreciably higher at institutions 3 and 11, which is consistent with the fact that at these institutions the user could make an interlibrary loan request for articles through PEAK, and the institution would pay the per article charge on behalf of the user.

We gain further understanding of the degree to which differences in user cost affects the demand for paid article access by looking at only those institutions that depleted their supply of tokens at various points throughout the project. There were three institutions in this category: institution 3 ran out of tokens in November 1998 and again in July 1999; institution 11 in May 1999; and institution 9 in June 1999.

For institutions that had tokens available at certain times, we can estimate the number of credit card requests (by PEAK, to the user) based on the number of tokens spent per free access. If we make the assumption that this rate of token expenditure would have remained constant had tokens still been available, we can estimate the number of credit card requests to be equal to the estimated number of tokens that would have been spent had tokens been available.

Table 6.7: Credit card payments as a percent of requests,

estimated from token expenditure rate

Institution	Credit Card Requests	Credit Card Payments	Percent
3	128	13	10.2%
9	366	1	0.3%
11	1128	61	5.4%

In Table 6.7 we present the rate of credit card payments as estimated from the rate of token expenditure. The relative percentages are consistent with our previous estimates for these institutions. The estimated number of *requests* for credit card payment are about twice as high as the estimates in Table 6.6. One possible explanation is that when users know they are going to face a credit card payment request (tokens have run out, which they learn on their first request for an article that is not prepaid) they may make fewer attempts to access material, which would be another measure of the effect of transaction payments on service usage.

Table 6.8: Effect of token depletion on demand for paid content

	Institution 3	Institution 3	Institution 9	Institution 11
	1998	1999	1999	1999
30 days prior	13.6	18.4	20.2	16.0
30 days after	0.25	0.29	0.00	0.35
Percentage Decrease	-98.2%	-98.4%	-100.0%	-97.8%

NOTE: Units: Normalized paid access per 100 unmetered accesses.

To further quantify the decrease in demand for paid access resulting from a depletion of tokens, in Table 6.8 we present the normalized accesses of metered content per hundred accesses of free content at these institutions for the 30 days prior and subsequent to running out of tokens. Usage plummeted after tokens ran out and users were required to pay per article for access to metered content.

Summary: Effects of user costs

The results we presented in this section demonstrate that increases in user costs *substantially* diminish demand for paid content. In particular, the decisions made by thousands of users demonstrate that non-pecuniary costs, such as password use, have an impact on demand that is of the same order of magnitude as direct monetary costs.

6.5 Effects of Expected User Cost on Access

As we showed in Table 6.4, at most institutions actual paid usage when users directly paid the user cost was substantially below predicted usage with zero user costs. Users at institution 10 were notable

exceptions. We hypothesized that users at this institution might have *expected* to bear more cost, and they were willing to pay more often when confronted with costs. We explore this hypothesis in this section.

According to our hypothesis, the frequency with which users are asked to pay for content will affect a user's *ex ante* estimation of how much she will need to pay. This effect on her estimate can stem from either her previous direct experience, or through "word of mouth" learning. It is our hypothesis that the expected access cost affected the probability that a user paid for access when requested.

We have two conjectures about user behavior that would cause willingness to pay to depend on prior expectations about cost. The first concerns an induced selection bias. The higher the expected cost to access an article, the fewer the users who will even attempt to access the information via PEAK. In particular, users with a low expected benefit for an article will generally be less likely to use PEAK at all. The result would be that those who do use PEAK are more likely to pay necessary article access fees. Our second conjecture is that context of the request for payment matters, i.e. there is a "framing" effect. It is possible that if a user is habituated to receiving something for free, she will be resistant to paying for that object, even if her expected benefit is greater than the actual cost. [18] Unfortunately, the data that we have do not permit us to distinguish between these two scenarios.

Table 6.9: Effect of subscription coverage on paid access

Institution	Normalized paid accesses per 100 unmetered	Estimated expected rate of zero cost access	Percent who log in when requested
3	13.5	83.6%	8.48%
10	31.7	6.9%	13.5%
11	7.6	74.2%	2.6 %
12	26.4	11.1%	7.1%
14	15.1	31.4%	29.6%

NOTE: See text for more complete definitions of the variables.
Correlation coefficients: Paid access and % of unmetered in subscription base: -0.87
Prompted login and % of unmetered in subscription base: -0.36

In Table 6.9 we present some evidence that users' expectations do matter. To explore this hypothesis, we rely on the difference in user cost between accesses to traditional subscription material (no password required) and generalized subscription material (password required). Therefore, we report all institutions at which password entry was required in order to spend a generalized subscription token, plus institution 14, at which users faced similar costs. We use accesses of unmetered content—which has zero incremental user cost for all material, whether in traditional subscriptions or not—as our comparison benchmark. In the second column we report the forecast of

unmetered content accesses that were contained within the institution's traditional subscription base. We use this as an estimate of the user's expected user cost of access. For example, if 75% of unmetered access came from traditional subscription material, then we estimate that the user also expects 75% of her demand for metered material to be from traditional subscriptions (with zero incremental user cost), and only 25% of requests for metered material to involve the password user cost (for generalized subscription content).

In the last two columns we present measures of user willingness to bear user cost. The institution's normalized paid access is a scaled measure of the rate at which (metered) generalized subscription material was accrued (and thus how soon the password cost was incurred). The pecent who login when requested is another measure of user willingness to bear the password user cost.

The data are consistent with our hypothesis that users with lower expected access costs (see column 2) will be less likely to bear the user cost of password retrieval and entry. The correlation between the expected rate of zero-cost access and normalized paid access is -0.87. We also see a negative correlation of -0.36 between the expected rate of zero cost access and willingness to enter a password when requested.

6.6 Improving library budgeting with usage information

Librarians are in an unenviable position when they select subscriptions to scholarly journals. [19] They must determine which journals best match the needs and interests of their community subject to two important constraints. The budgetary constraint has become increasingly binding because renewal costs have risen faster than serial budgets Haar (1999). The second constraint is that libraries have incomplete information about community needs. A traditional print subscription forces libraries to purchase publisher-selected bundles of information (the journal), while users are interested primarily in the articles therein. Users only read a small fraction of articles, [20] and the library generally lacks information about which articles the community values. Further compounding the problem, a library makes an *ex ante* (before publication) decision about the value of a bundle, while the actual value is realized *ex post*.

The PEAK electronic access products relaxed these constraints. First, users had low-cost access to articles in journals to which the institution did not subscribe. This appeared to be important: at institutions that purchased traditional subscriptions, 37% of the most accessed articles in 1998 were outside the institution's traditional subscription base. This figure was 50% in 1999. Second, the transaction logs that are feasible for electronic access allowed us to provide libraries with monthly reports not only on which journals their community valued, but also which articles. Detailed usage reporting should enable libraries to provide additional value to their communities. They can better allocate their serials budgets to the most valued journal titles or to other access products.

In this section we present analyses of the extent to which improved information available from an electronic usage system could lead to reduced expenditures and better service.

Improved budgeting with improved usage forecasts

We first estimate an upper bound on how much the libraries could benefit from better usage data. We analyze each institution's accesses to determine what would have been its optimal bundle if it had been able to *perfectly forecast* which material would be accessed. We then calculate how much this bundle would have cost the institution, and compare this perfect foresight cost with the institution's actual expenditures. Obviously even with extensive historical data, libraries would not be able to perfectly forecast future usage, so the realized efficiencies from better usage data would be less. Below we analyze how the libraries used the information from 1998 to change their purchasing decisions in 1999.

We present these results by access product in Table 6.10. We found that actual expenditures were markedly higher than optimal purchases in 1998. In particular, institutions in the Red and Blue groups purchased far more traditional subscriptions than would be justified if they had perfect foresight. Most institutions purchased more generalized subscriptions than would have been optimal with perfect foresight. We believe that much of the budgeting "error" can be explained by a few factors:

- First, institutions overestimated demand for access, particularly for journals for which they purchased traditional subscriptions. [21]

- Second, institutional practices, such as "use it or lose it" budgeting and a preference for fixed, predictable expenditures, might have affected decisions. A preference for predictable expenditures would induce a library to rely more heavily on traditional and generalized subscriptions, and less on reimbursed individual article purchases or interlibrary loan. [22] However, Kantor et. al. (2001) Kantor et al. (this volume) report the opposite: that libraries dislike bundles because they perceive them as forcing expenditures for low-value items.

- Third, because demand foresight is necessarily important, libraries might want to "over-purchase" to provide insurance against higher than expected usage demand. Of course, per-article purchases (possibly reimbursed to users) provide insurance (as does an interlibrary loan agreement), but at a higher cost per article than pre-purchased generalized subscription tokens, or than traditional subscriptions.

Table 6.10: Actual versus optimal expenditures per access product

for 1998-1999

| | | Access Product | | | | | | Totals |
| | | Traditional | | Generalized | | Per Article | | |
Year	Instid	Actual	Optimal	Actual	Optimal	Actual	Optimal	Actual
1998	3	25,000	17,000	2,740	3,836	7	133	27,747
	5	N/A	0	15,344	6,576	0	169	15,344
	6	N/A	0	0	548	672	0	672
	7	N/A	0	24,660	12,604	0	0	24,660
	8	N/A	0	13,700	2,740	0	0	13,700
	9	0	556	13,700	6,576	0	56	13,700
	10	4,960	323	8,220	7,672	0	483	13,180
	11	70,056	5,217	2,192	13,700	0	84	72,248
	12	2,352	107	2,192	1,096	0	98	4,544
	13	28,504	139	N/A	0	952	1,120	29,456
	14	17,671	0	N/A	0	294	504	17,965
	15	18,476	0	N/A	0	0	1,176	18,476
1999	3	12,500	10,528	2,740	1,096	84	0	15,324
	5	N/A	0	8,768	2,740	0	399	8,708
	6	N/A	0	0	548	686	0	686
	7	N/A	0	10,960	9,864	0	511	10,960
	8	N/A	0	6,028	5,480	0	462	6,028
	9	0	278	7,124	6,576	7	182	7,131
	10	2,480	1,401	8,768	6,576	0	210	11,247
	11	0	576	4,384	2,740	427	532	4,559
	12	0	0	1,644	548	0	539	1,644
	13	9,635	7,661	N/A	0	19,964	7,175	29,599
	14	0	0	N/A	0	623	623	623
	15	8,992	1,058	N/A	0	511	1,694	9,502

Table 6.11: Predicted vs. actual direction of expenditure change for traditional and generalized subscriptions (by institution, 1998-99).

| | Change in expenditure 1998-99 | | | |
| | Traditional | | Generalized | |
Institution	Predicted	Actual	Predicted	Actual
3	-	0	+	+
5	N/A	N/A	-	-

6	N/A	N/A	+	o
7	N/A	N/A	-	-
8	N/A	N/A	-	-
9	+	o	-	-
10	-	o	-	+
11	-	-	+	+
12	-	-	-	+
13	-	o	N/A	N/A
14	-	o	N/A	N/A
15	-	+	N/A	N/A

NOTE: Predicted change direction is based on whether institution over- or under-purchased that product in 1999.
"o" indicates no change; "N/A" indicates the access product was not available to that institution; "+" and "-" indicate an increase and decrease, respectively.

We also analyzed changes in purchasing behavior from the first to the second year of the project. The PEAK team provided participating institutions with regular reports detailing usage. We hypothesized that librarian decisions about purchasing access products for the second year (1999) might be consistent with a simple learning dynamic: increase expenditures on products under-purchased in 1998 and decrease expenditures on products they over-purchased in 1998. For each institution we compared the direction of 1998-99 expenditure change for each access product to the change we hypothesized. [23] We present the results in Table 6.11.

Six of the nine institutions adjusted the number of generalized subscriptions in a manner consistent with our hypothesis. [24] Fewer adjusted traditional subscriptions in the predicted direction. Two of the seven institutions that purchased more traditional subscriptions in 1998 than was *ex post* optimal then decreased the number purchased in 1999. Indeed, only three of the eight institutions made any changes at all to their traditional subscription lineup. This suggests an inertia that cannot be explained solely by direct costs to the institution. Perhaps libraries see a greater insurance value in having certain titles freely available through traditional subscriptions than from having generalized subscription tokens available that can be used on articles from any title. Generalized subscription tokens are also more expensive per article than traditional subscription prices, so the libraries are purchasing more *potential* usage with their budgets. Another explanation might be that libraries were more cautious about purchasing generalized subscriptions because it was a less familiar product.

Table 6.12: Estimation results for forecast error

 Independent variable   Coefficient (standard error)

Year 1999	-35.7*
	(9.3)
Green	54.6*
	(10.0)
Red	53.3*
	(8.1)
Blue	85.8*
	(9.2)
Sample Size	24
R^2	0.85

NOTE: Dependent variable is forecast error (in percent).
No constant term is included in the regressions.
Standard errors are shown in parentheses.
* Significant at the 99% level.

We performed a regression analysis to assess the differences between apparent over-purchasing in 1998 and 1999. Our dependent variable was the difference between the perfect forecast expenditure and actual expenditure, which we call the "forecast error". In Table 6.12 we report the effects of learning (the change in the error for 1999) and the average differences across experimental groups. The perfect foresight overspending over the life of the project averaged between 53% (Red) and 86% (Blue). However, the overspending was on average 36 percentage points lower in 1999. This represents a reduction of about one-half in perfect foresight overspending. [25]

We also considered other control variables, such as the institution's level of expenditures, fraction of the year participating in the experiment and number of potential users, but their contribution to explaining the forecast error was not statistically significant. The between-group variation and the 1999 improvement account for about 85% of the variation, as measured by the R^2 statistic.

Decisions about specific titles

In addition to comparing the total number of subscriptions for an institution with the optimal number, we can also identify the optimality for each particular title subscribed. We calculate, based on observed usage and prices, which titles an institution with perfect foresight should have obtained through traditional subscriptions, and call this the *optimal set*. Then we calculate two measures of actual behavior. First, we determine which titles in the optimal set an institution actually purchased. Second, we determine which traditional subscription titles the institution would have been better off foregoing because actual access would have been less expensive using other available access products.

In Table 6.13 we present our analysis of the traditional subscription

titles selected by institutions. There is wide variation both in the percent of purchased subscriptions that *are* in the optimal set, and in the percent of journals in the optimal set to which the institution did not subscribe,[26] Overall, there is substantial opportunity for improvement. This is not a criticism of institutional decisions. Rather, it indicates the opportunity for improved purchasing decisions if libraries obtain the type of detailed usage information PEAK provided.

We do generally see better decisions in 1999. However, in both years a rather large percentage of subscribed journals were not accessed at all.

Table 6.13: Optimality of subscription choices

Institution	Year	Total subscriptions	Percent subscribed that are in optimal set	Percent of optimal set that were not subscribed	Percent of subscriptions accessed at least once	
3	1998	907	53.3%	3.4%	92.5%	
10	1998	23	0.0%	100.0%	65.2%	
11	1998	663	3.6%	0.0%	84.5%	
12	1998	22	0.0%	100.0%	81.8%	
13	1998	205	0.5%	0.0%	12.7%	
14	1998	72	0.0%	N/A	36.1%	
15	1998	102	0.0%	N/A	48.0%	
3	1999	907	75.0%	7.7%	97.0%	
10	1999	23	13.0%	76.9%	65.2%	
13	1999	205	29.8%	62.6%	86.8%	
14	1999	72	0.0%	N/A	20.8%	
15	1998	102	10.8%	8.3%	84.3%	

Dynamic Optimal Choice

Access product purchasing decisions made by institutions have a profound impact on the costs faced by users, and thus on the realized demand for access. Therefore, in deciding what access products, electronic or otherwise, to purchase, an institution must not only consider the demand realized for a particular level of user cost, but also what would be demanded at differing levels of user costs. Likewise, in our determination of the optimal bundle of access products, we should not take the given set of accesses as fixed and exogenous. As a simple example, let us assume that a subscription to a given journal requires 25 accesses in order to pay for itself. Now assume that the institution in question did not subscribe to that journal, and that 20 tokens were used to access articles in the time

period. At first look, it appears as though the institutions did the optimal thing. Let us assume, however, that we know that accesses would increase by 50%, to 30, when no password is required. It now appears as though the institution should have subscribed, since the reduced user costs would stimulate sufficient demand to justify these higher costs.

Table 6.14: Optimal bundles with barrier-free access: Selected institutions

Institution	Year	Trad. Subscriptions		Addit. Articles		Increase	Total
		Actual Optimal	Rescaled Optimal	Actual Optimal	Rescaled Optimal	Optimal Cost	Access Increase
3	1998	500	556	1099	1130	9.39%	12.53%
3	1999	737	805	236	146	4.85%	7.46%
11	1998	24	31	2532	3019	21.11%	21.09%
12	1998	1	1	254	287	17.76%	13.67%
14	1999	0	0	168	249	48.21%	48.21%
15	1999	12	17	242	366	47.56%	60.36%

In Table 6.4 we reported results that allow us to estimate how much usage would increase if no passwords or other user costs were incurred. We now calculate the product purchases that would have optimally matched the usage demand that we estimate would have occurred had the library removed or absorbed all user costs. We report the results in Table 6.14. [27] For most institutions, the optimal number of journal subscriptions increases, because greater usage makes the subscription more valuable. In general, the estimated institution cost of the optimal bundle would not increase greatly to accommodate the usage increase that would follow from eliminating user costs. Although we cannot quantify a dollar value for the eliminated user costs (because they include nonpecuniary costs such as those from requiring a password), we show in the last two columns that the modest institutional cost increase would be accompanied by comparable or larger increases in usage. The greatest cost increase (48%) occurs for the institutions (14 and 15) at which generalized subscription tokens were not available and the institution did not directly subsidize the per-article fee, i.e. at those institutions where users faced the highest user costs. Thus, the higher institutional costs should be weighed against high savings in user costs (including money spent on per-article purchases).

6.7 Conclusion

Experience from the early years of electronic commerce indicates that low user costs—nonpecuniary as well as pecuniary—are critical to the success of electronic distribution systems. In the PEAK experiment, we have evidence that for the information goods in question, these non-pecuniary costs are of the same magnitude as significant pecuniary

costs. In a two-tiered decision problem such as in this project, where intermediaries determine the user costs required to access specific content, both the quantity *and* composition of demand is greatly affected by users' reactions to these costs. Therefore any determination of what the intermediary "ought" to do must take these effects into account. Furthermore, we have initial evidence that suggests that users who come to expect information at zero marginal cost are far less likely to pay these non-monetary costs when requested than their counterparts who expect these costs. This finding is of great import to both those who design electronic information delivery and pricing systems as well as any intermediaries controlling information access and costs.

In the second part of the chapter we investigated the extent to which libraries could have improved their purchasing decisions if they had detailed usage information that provided a reliable basis for forecasting future usage. We found that with perfect foresight about next year's usage, libraries could have substantially reduced their expenditures. They could also have substantially improved the match between what titles they purchased and what articles users want to access.

We then linked the two sets of analyses by showing how much greater usage would be if the library absorbed or removed the pecuniary and non-pecuniary user costs we observed. The result would be substantial increases in usage. The library expenditures would have to increase by comparable percentage amounts; however the institution should recognize that these costs would be offset by the lower user costs incurred by its constituents, and the net cost, if any, would support substantial increases in usage.

Notes

1. See MacKie-Mason and Riveros (2000) for a discussion of the economics of electronic publishing.

2. See Bonn et al. (this volume) and Hunter (this volume) for accounts of the genesis of this project.

3. Kingma (this volume) provides a good discussion of the role of library as intermediary.

4. As we further discuss below, user cost may include several components only one of which is a standard price. The other components may include, for example, time and inconvenience. We expect these user costs, taken together, and not price alone, to determine usage.

5. 120 is the approximate average number of articles in a traditional printed journal for a given year. We refer to this bundle of options to access articles as a set of tokens, with one token used for each article added to the generalized subscription during the year.

6. For example, a Green institution first decides how many generalized

subcriptions to purchse (if any). Users then access articles using generalized subscription "tokens" at zero pecuniary cost until the tokens run out, and thereafter pay a fee per article for additional articles. The library determines how many articles (not which articles) are available at the two different prices.

7. To access PEAK from other IP addresses, users entered a password. Once access was granted, all content in these categories was available without further user cost.

8. In the first eight months of the experiment, users paid with a First Virtual VPIN account, rather than with a credit card. Because a VPIN was an unfamiliar product, the non-pecuniary costs were probably higher than for credit card usage, although formally the user needed to undertake the same steps.

9. When the user accessed an article for which per-article payment was required, the institution was automatically billed by the PEAK service.

10. Paid content is metered content not including article in journals to which an institution purchased a traditional subscription.

11. Formally, *Normalized Paid Access* is equal to $\frac{A_{paid}}{A_{unmetered}} \cdot (Scale)$, where A_{paid} is the total number of paid accesses, $A_{unmetered}$ the total number of unmetered accesses, and Scale is equal total number of free accesses divided by the total number of accesses of free content in journals to which the institution does not have a traditional subscription. We multiply by Scale because the more that accesses are covered by traditional subscriptions, the less likely a user is to require paid access. Scaling by access to unmetered content also controls for different overall usage intensity (due to different numbers of active users, differences in the composition of users, differences in research orientation, differences in user education about PEAK, etc.). Unmetered accesses proxies for the number of user sessions, and therefore our statistic is an estimate of paid accesses per session.

12. Only 28% of unmetered accesses from Red group users were password authenticated. This suggests that a large majority of users attempting to access paid content would not already be password authenticated. For these users, the need to password authenticate would truly be a marginal cost.

13. The Elsevier journal catalogue is especially strong in these subject areas, so we expect differences in usage when the subject area concentration of the user community differs.

14. In only two cases were credit cards are required, and both were at the same institution.

15. Recall that all users at an institution could access, without

password authentication, any article previously purchased by that institution with a generalized token. For articles purchased on a per-article basis, only the individual who purchased the article could view it without further monetary cost.

16. The libraries at institutions 3 and 11 processed these requests electronically, through PEAK, while the library at institution 9 did not and thus incurred greater processing delays.

17. In addition, institution 6 is a corporate institution. It is possible that its users' budgetary constraints were not as binding as those associated with academic institutions.

18. This phenomenon was widely discussed—though not, to our knowledge, sufficiently demonstrated—during the early years of widespread public access on the Internet. Many businesses and commentators asked whether users would pay for any content after being accustomed to getting most Internet-delivered information for free.

19. For an excellent discussion of the collection development officer's problem, see Haar (1999)

20. The percentage of articles read through June 1999 for academic institutions participating in PEAK ranged from .12% to 6.40%. An empirical study by King and Griffiths (1995) found that about 43.6% of users who read a journal read five or fewer articles from the journal and 78% of the readers read 10 or fewer articles.

21. Project implementation delays exacerbated the demand forecasting problem. For example, none of the institutions in the Blue Group started the project until the third quarter of the year.

22. With print publications and some electronic products libraries may be willing to spend more on full journal subscriptions to create complete archival collections. All access to PEAK materials ended in August 1999, however, so archival value should not have played a role in decision making.

23. As 1999 PEAK access is for 8 months, the number of 1999 generalized subscriptions was multiplied by 1.5 for comparison with 1998.

24. One of the institutions that increased token purchases despite over purchasing in 1998 was more foresightful than our simple learning model: its usage increased so much that it ran out of tokens less than six months into the final eight-month period of the experiment.

25. E.g., the Green group had average overspending of about 55% so a 36-point change represents a shift from about 73% in 1998 to about 37% in 1999.

26. The calculations in the two columns are independent and should

not generally sum to one. The first column indicates the percent of titles that were subscribed that *should* have been subscribed (given perfect foresight). A high percent means there were not many specific titles subscribed that should not have been. However, this does not indicate that a library subscribed to most of the titles that it should have. A library that subscribes to zero journals will get 100% on this measure: no journals were subscribed that should not have been. The second column addresses this question: what percent of those titles that should have been subscribed were missed? The two columns correspond to Type I and Type II error in classical statistical theory. The first should be high, and the second low if the institution is forecasting well (and following our simple model of "optimal" practice).

27. We performed the calculation for those institutions for which we have a good estimate of the user cost effect (see Table 6.4), and for which there were enough article accesses for meaningful estimation.