

Evaluating learning in library training workshops: using the retrospective pretest design

Mary McDiarmid and Malcolm Binns

Abstract: The aim of this study was to assess the effectiveness of an evaluation instrument using the retrospective pretest design to measure changes in participants' behaviour after library training. This article focuses on the measurement component of training evaluation — the process of answering the question of how much change has occurred. Participants, who were from a large, public academic geriatric care centre in Toronto, included administrators, researchers, clinical and other staff, and university students doing field placements at the hospital. Participants attended one of four 1-hour sessions on the topic of Effectively Searching Google and Google Scholar that were held over a 3-month period. Sixty days post training, a self-administered retrospective pretest questionnaire, consisting of 10 searching behaviour statements developed using the learning objectives for the training session, was used to measure the impact of library training on participants' behaviour. Participants were asked to indicate their level of frequency of performing a searching behaviour described in the statement before and after training using a five-point, Likert-type scale ranging from 1, almost never; 2, seldom; 3, about half the time; 4, often; to 5, almost always. Summary baseline statistics are reported for respondents who never or rarely exhibited the behaviour prior to training. For the change measure, we report the simple percentage of respondents who improved. The findings of this study showed the potential of using the retrospective pretest to help librarians document the outcomes of library training. The benefits of gathering data using the retrospective pretest are discussed.

Introduction

Hospital librarians spend many hours developing training programs and materials, and delivering training sessions to hospital personnel and physicians. The purpose of this article is to bring the retrospective pretest design, a very valuable and easy-to-use research design, to the attention of librarians evaluating library training programs. For example, this design can be used to evaluate the effectiveness of training in programs such as database training in MEDLINE and other health-related databases or for Internet-searching skills instruction. The effectiveness of training is important and is a subject worthy of research to hospital librarians since it is an indicator evaluated by the Canadian Council on Health Services Accreditation's Information Management Standards. Specifically, it is indicator 6.4: "The education, training, and support are effective" [1]. The big question is how do we know library training is effective, and how do we prove it?

Training is focused on trying to change behaviour or teach new behaviours to individual participants [2]. Kirkpatrick's four-level taxonomy of training outcomes distinguishes between participants' reactions to training (level 1), their acquisition of new knowledge (level 2), changes in on-the-job

behaviour (level 3), and changes in organizational results (level 4) [3]. Training outcomes can be easily measured at Kirkpatrick's level one through little more than a simple satisfaction survey administered at the end of training. The satisfaction with training survey only assesses participants' initial reactions to a course, and participants are asked to rate the level of their satisfaction with the content, handouts and (or) materials, instructor, and overall satisfaction with the training received. For the busy hospital librarian, this is where evaluation of library training usually ends. This does not measure learning [2]. Some librarians ask trainees to complete a pretest and posttest to measure how well the attendees learned the skills and knowledge in the workshop. This assessment corresponds to Kirkpatrick's level 2 assessment. This method has problems since a pretest taken at the beginning of a training program may be invalid because participants have limited knowledge to respond accurately to the items being asked on the pretest. In addition, by the end of a training session, their new understanding of the concepts may have an impact on the responses in their self-assessment [4].

The method in the current study, the retrospective pretest, also known as the post-then-pre evaluation approach, focuses on the process of answering the question of how much change has occurred. Retrospective pretest assessment of training outcomes, corresponding to level three in Kirkpatrick's outcomes of learning, documents changes in trainees' on-the-job behaviour. In the retrospective pretest method, the first question on the posttest asks the participant about behaviour as a result of the training. Then the participant is

M. McDiarmid.¹ Staff Library, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, ON M6A 2E1, Canada.

M. Binns. The Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, ON M6A 2E1, Canada.

¹Corresponding author (e-mail: mmcdiarmid@baycrest.org).

asked to report what the behaviour had been before the program. This second question is really the pretest question, but it is asked after the training when the participant has sufficient knowledge to answer the question validly. The primary reason for the increased reliability of the answers is that participants often do not know what they do or do not know before training; asking them first about what they do now helps to indicate what it is they actually did not know or do prior to the training [4]. The retrospective pretest has been effective in eliminating response shift bias in educational and training programs [5,6]. Response shift bias is avoided because participants are rating themselves with a single frame of reference on both the posttest and retrospective pretest.

Using the retrospective pretest to collect self-reported behavioural changes in trainees may provide substantial scientific evidence for a library-training program's impact. Researchers have generally concluded that differences between retrospective pretests and conventional posttests are adequate indicators of change in behaviours [7–12]. The retrospective pretest evaluation may offer health sciences librarians a way of documenting the value of library training to their organizations. Measurement is the only way of providing hard evidence to senior management of the value and the bottom line impact of training. A literature search of *Library Literature and Information Science Fulltext* and *LISA: Library and Information Science Abstracts* uncovered no articles describing the use of the retrospective pretest to assess the impact and (or) outcome of library training.

The purpose of this study was to assess the effectiveness of an evaluation instrument using the retrospective pretest to measure changes in participants' behaviour after library training.

Program description

Design

The retrospective pretest was chosen because it is a simple, convenient and expeditious method of assessing changes in self-reported knowledge and skills. The retrospective pretest has an added advantage in that it is only administered a single time. It is also flexible because questions can be designed to reflect actual program content as it evolves over the time of a training course [13].

Settings and participants

The setting for this study was a large, public academic geriatric care centre in Toronto. The participants included administrators, researchers, clinical and other staff, and university students doing field placements at the hospital. Four 1-hour sessions on the topic of Effectively Searching Google and Google Scholar were held over a 3-month period. There were 42 participants in the training program; 15 returned the retrospective pretest questionnaire (a participation rate of 35.7%).

Intervention

The self-administered retrospective pretest questionnaire (Fig. 1) was used to measure the impact of library training on participants' behaviour. Ten statements were developed using the learning objectives for the training session. Participants were asked to indicate their level of frequency of per-

forming a searching behaviour described in the statement before and after training using a five-point, Likert-type scale ranging from 1, almost never; 2, seldom; 3, about half the time; 4, often; to 5, almost always. The retrospective pretest questionnaire was distributed 60 days post training to all participants via e-mail with a message inviting voluntary participation. Participants were asked to complete the questionnaire and return it anonymously to library services through interoffice mail, or if they preferred, they could complete it online and return it via e-mail. It was felt that 60 days was sufficient time for participants to have incorporated what they had learned into their practice. Participants were told that the data collected would be used to improve the training. Changes in their search behaviour would be reviewed and associated with the content and teaching method. Two weeks after the initial e-mail, a follow-up e-mail was sent to all participants thanking those who had already responded and encouraging the remaining participants, who had not yet responded, to do so within 2 weeks. Participants were not told at the end of the initial training session that they would be receiving a follow-up questionnaire because it could have influenced their post-training behaviour. At the end of each training session, participants received a training satisfaction questionnaire that is routinely given to participants after all library training. They also received a copy of the presentation handout and training slides.

The analysis of data was designed so it could be used by a librarian in a small hospital library setting without access to sophisticated statistical software programs such as SPSS. Data was analyzed using Microsoft Excel. Training effects are described in two waves: baseline and change across training interval. For the baseline data, the percentage of respondents who fell below a score of three (those who never or rarely performed the behaviour) prior to training is reported. For the change measure, the percentage of respondents who improved (i.e., participants who reported increased frequency of a behaviour 60 days post training) is provided.

Outcomes

The changes in respondents' behaviour pre and post training are shown in Table 1. Before library training, 73% of respondents almost never or seldom used quotes when searching terms as a phrase in Google. After training, 80% of respondents had changed their behaviour and reported using quotes more often. The use of the plus operator before a search term in Google was almost never or seldom performed by 87% of respondents before training. After training, 67% of respondents increased their use of the plus operator. Forty percent of respondents almost never or seldom modified or tried different search strategies in searching Google prior to training. After training, 67% of respondents improved their score on this behaviour, reporting more frequent use of different strategies. A notable majority, 93% of respondents prior to training, almost never or seldom used Google special features such as searching "News" for news stories or "Images" for pictures. After training, 53% increased frequency of usage for these features.

The remaining six learning objectives found less change in respondents' post-training behaviour. Before training, 80% of respondents almost never or seldom used the ad-

Fig. 1. Retrospective pretest questionnaire.

For each statement Select the box under the number that best describes your behaviour
IN BOTH the After and Before columns

1 = almost never, 2 = seldom, 3 = about half the time, 4 = often, 5 = almost always

	After Google Training Do I?					Before Google Training Did I?				
	1	2	3	4	5	1	2	3	4	5
Search terms as a phrase using quotes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Use the + operator	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Modify my search strategy using different words/phrases	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Limit my search to trustworthy sites as needed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Use Google flavours to search, e.g. news, images	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Refer to the Advanced Search Page for tips	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Use the Google translation feature	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Retrieve a lot of irrelevant sites	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Use Google Scholar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Use Google Scholar to access the current scholarly medical information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

vanced search page in Google. After training, 47% of the respondents had improved their score. The translation feature in Google was almost never or seldom used by 87% of respondents before training. Post training, only 27% of respondents reported using it more frequently. The retrieval of irrelevant sites had most people scoring quite high (20%) and showed very few respondents (7%) improved after training. Most people prior to training had almost never or seldom used Google Scholar, and after training, 40% of respondents increased their use of Google Scholar. When asked whether respondents used Google Scholar to retrieve current scholarly medical literature, 93% of respondents almost never or seldom did this. After training, 20% of respondents reported increased use of Google Scholar to retrieve scholarly medical literature.

Discussion

The retrospective pretest allows researchers to reduce response shift because the participants are able to give pretest answers that are based on a more accurate frame of reference. The participants only understand what you are asking them about after training and so are not able to be accurate if asked pretraining. By using the retrospective pretest, response shift bias can be reduced and therefore increase the likelihood that the observable results are due to the intended pro-

gram effects [15]. Limitations of the retrospective pretest must be acknowledged. The level of recall accuracy available from any self-report must be considered. Despite the fact that response shift bias is reduced by retrospective pretests, self-reports remain a form of estimations [5,6]. The retrospective pretest design may be prone to other possible biases that are common to most other survey designs. There is a possibility of selection bias because of the low response rate. The small sample of participants that responded to the post-training evaluation may represent only those participants who were highly motivated to learn about the topic and were therefore more likely to utilize the strategies they learned in the training session. Because of the possibility of selection bias and the small sample size in this study, these data are considered preliminary and perhaps not generalizable to a broader population.

As a result of using the retrospective pretest, the librarian-instructor gained the following insights. With a modest investment of time, this self-administered evaluation tool provided rich data (see Table 2). Data gathered were relatively easy to analyze and communicate as change in behaviours. Among the advantages of using the retrospective pretest to evaluate participants' learning is that it aided in clarifying where the training was successful, where content was redundant, or where the content needed to be revised. For example, the training was successful in promoting the use of

Table 1. Percentage of respondents who scored less than 3 on the pretest and the percentage of respondents who performed the behaviour with increased frequency after training ($n = 15$).

Learning objective	Pretest % of respondents who performed behaviour almost never or seldom	Posttest % of respondents who performed behaviour with increased frequency
Search terms as a phrase using quotes	73	80
Use the "+" operator	87	67
Modify the search strategy using different words or phrases	40	67
Limit the search to trustworthy sites as needed	27	33
Use Google flavours to search, e.g., news, images	93	53
Refer to the Advanced Search Page for tips	80	47
Use the Google translation feature	87	27
Retrieve a lot of irrelevant sites	20	7
Use Google Scholar	93	40
Use Google Scholar to access the current scholarly medical information	93	20

Table 2. Frequency of response ($n = 15$).

Learning objective	Test	Almost never	Seldom	About half the time	Often	Almost always
Search terms as phrase using quotes	Pretest	8 (53.3)	3 (20.0)	2 (13.3)	–	2 (13.3)
	Posttest	1 (6.7)	1 (6.7)	3 (20.0)	5 (33.3)	5 (33.3)
Use the + operator	Pretest	12 (80.0)	1 (6.7)	–	–	2 (13.3)
	Posttest	3 (20.0)	2 (13.3)	3 (20.0)	4 (26.7)	3 (20.0)
Modify search strategy	Pretest	1 (6.7)	5 (33.3)	4 (26.7)	3 (20.0)	2 (13.3)
	Posttest	–	–	5 (33.3)	5 (33.3)	5 (33.3)
Limit search to trustworthy sites	Pretest	2 (13.3)	2 (13.3)	3 (20.0)	5 (33.3)	3 (20.0)
	Posttest	1 (6.7)	3 (20.0)	1 (6.7)	5 (33.3)	5 (33.3)
Use Google flavours such as news, images	Pretest	9 (60.0)	5 (33.3)	1 (6.7)	–	–
	Posttest	6 (40.0)	3 (20.0)	5 (33.3)	1 (6.7)	–
Refer to advanced search page	Pretest	12 (80.0)	–	2 (13.3)	1 (6.7)	–
	Posttest	7 (46.7)	3 (20.0)	–	3 (20.0)	2 (13.3)
Use Google translation	Pretest	12 (80.0)	1 (6.7)	2 (13.3)	–	–
	Posttest	8 (53.3)	3 (20.0)	3 (20.0)	1 (6.7)	–
Retrieve a lot of irrelevant sites	Pretest	2 (13.3)	1 (6.7)	4 (26.7)	4 (26.7)	4 (26.7)
	Posttest	2 (13.3)	6 (40.0)	6 (40.0)	–	1 (6.7)
Use Google Scholar	Pretest	13 (86.7)	1 (6.7)	1 (6.7)	–	–
	Posttest	8 (53.3)	3 (20.0)	1 (6.7)	2 (13.3)	1 (6.7)
Use Google Scholar to access current medical information	Pretest	13 (86.7)	1 (6.7)	1 (6.7)	–	–
	Posttest	10 (66.7)	2 (13.3)	1 (6.7)	1 (6.7)	1 (6.7)

Note: The values in parentheses represent the percentage of responses for each search behaviour.

quotes, the use of the plus operator, and modifying a searching strategy using different words or phrases. Only two of 15 respondents did not improve their score after training in using quotes, and only four of 15 respondents failed to change their behaviour regarding the plus operator and modifying a search strategy after training. Another example that demonstrates where training was successful was the number of respondents after training who used Google Scholar to retrieve scholarly medical information. At first glance the low rate of 20% of respondents who improved their score post training by using Google Scholar to find scholarly medical literature might seem disappointing, unless one knows that

during the training session, the limitations of Google Scholar were emphasized. Google Scholar was not promoted as a source for scholarly medical information during training because it was still just a beta test site. The low number of respondents who improved their score after training was evidence of success because it demonstrated that trainees had absorbed the knowledge that Google Scholar was not the place to look for scholarly medical information. The fact that 20% increased their use of Google Scholar after training informed the librarian-instructor that perhaps training sparked an interest in Google Scholar. The slight increase does not necessarily mean that training was unsuccessful in

conveying the unreliability of Google Scholar. Another way to interpret this increase is that people may do a quick and convenient search in Google Scholar and then check what they find in a more reliable source such as MEDLINE.

The librarian-instructor realized from the findings that perhaps there is little demand for the Google translation feature since 10 of 15 respondents did not change their behaviour after training. Of these 10 respondents, eight appeared to have little use for the translation feature because they almost never used it either before or after training. Another area of concern regarding content was in the section dealing with strategies for reducing the amount of irrelevant sites retrieved. The current study showed that after training, respondents still reported a high frequency of retrieving irrelevant sites. This finding prompted the librarian-instructor to change the content of the training session to increase the time spent in the session on strategies that could be used to reduce the number of irrelevant sites.

The data gathered through a self-administered retrospective pretest instrument is beneficial to librarian trainers in five specific ways. First, it is effective as a way to quantify or measure changes in participants' behaviour after library training. Second, it helps to identify training content that needs to be reduced or enhanced. Third, it can be used to demonstrate the impact of library training on workers' behaviour in the workplace. Fourth, it can also be used as a quality improvement initiative by setting a quality improvement goal for each learning behaviour (e.g., at least 50% of all trainees will report that after training they retrieve fewer irrelevant sites). Fifth, another unexpected benefit of the retrospective pretest questionnaire was that it helped to promote the library as a client-focused service. During the writing of this paper, our hospital underwent an accreditation process, and one of the questions posed to the information management team was to give an example of how we changed a procedure or process based on feedback from clients. We described the retrospective pretest evaluation used by library services in assessing training outcomes, and the accreditation surveyors cited this as an excellent example.

Using retrospective pretest evaluation aids in the librarian's never-ending effort to document for senior management how library training is of value, is effective, and impacts on trainees' behaviour. It is easy for trainees to complete, data can be quickly analyzed using simple spreadsheet software, and, if appropriate software is available, comparison of means using statistics such as the *t*-test can be used to identify significant changes in specific behaviours. The challenge in constructing a retrospective pretest evaluation instrument is to identify specific behaviours or learning objectives from your training content that may affect change in trainees. Then you need to specify an appropriate measurement scale such as

the Likert five-point scale used in the current study, which tests the amount of self-perceived behaviour change.

A future step that may be undertaken as an addendum to the current study is to attempt to satisfy Kirkpatrick's level 4 evaluation that documents changes in organizational results. One could revisit the trainees 1 year after training to ask them to specify any benefits they attribute directly to the training they received, such as saved time, increased efficiency, improved decision-making, or increased retrieval of higher quality Internet-based literature.

References

1. The Canadian Council on Health Services Accreditation. *AIM: Achieving Improved Measurement*. Ottawa: The Council; 2005.
2. Goldstein IL, Gilliam P. Training system issues in the year 2000. *Am Psychol*. 1990;45(2):134–43.
3. Kirkpatrick DL. *Evaluating Training Programs: Four Levels*. San Francisco: Berrett-Koehler Publishers; 1994.
4. Rockwell SK, Kohn H. Post-then-pre evaluation. *Journal of Extension*. 1989 [accessed 2005 June 16];27(2). Available from <http://www.joe.org/joe/1989summer/a5.html>.
5. Robinson EAR, Doueck HJ. Implications of the pre/post/then design for evaluating social group work. *Res Soc Work Pract*. 1994;4:224–40.
6. Sprangers M, Hoogstraten J. Pretesting effects in retrospective pretest-posttest designs. *J Appl Psychol*. 1989;74:265–72.
7. Terborg JR, Howard GS, Maxwell SE. Evaluating planned organizational change: a method for assessing alpha, beta and gamma change. *Acad Manage Rev*. 1980;5(1):109–21.
8. Sprangers M, Hoogstraten J. Pretesting effects in retrospective pretest-posttest designs. *J Appl Psychol*. 1989;74:265–72.
9. Sprangers M. Subject bias and the retrospective pretest in retrospect. *Bull Psychon Soc*. 1989;27(1):11–14.
10. Mann S. Implications of the response-shift bias for management. *Journal of Management Development*. 1997;16:328–36.
11. Schwartz CE, Sprangers M. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research: a theoretical model. *Soc Sci Med*. 1999;48:1531–38.
12. Wilson IB. Clinical understanding and clinical implications of response shift. *Soc Sci Med*. 1999;48:1577–88.
13. Pratt CC, McGuigan WM, Katzev AR. Measuring program outcomes using retrospective pretest methodology. *American Journal of Evaluation*. 2000;21(3):341–49.
14. Rossi PH, Lipsey MW, Freeman HE. *Evaluation: a systematic approach*. 7th ed. Thousand Oaks, Calif.: Sage; 2004.
15. Lamb TA, Tschillard R. Evaluating learning in professional development workshops: using the retrospective pretest. *Journal of Research in Professional Learning*. 2005 [accessed 2005 July 28]. Available from <http://www.nsd.org/library/publications/research/lamb.pdf>.