Equating scores on "Lite" and long library user survey forms: the LibQUAL + ® Lite randomized control trials

Bruce Thompson
Texas A&M University and Baylor College of Medicine, College Station, Texas, USA
Martha Kyrillidou
Association of Research Libraries, Washington, DC, USA
Colleen Cook
Texas A&M University, College Station, Texas, USA

Emerald

Equating scores on "Lite" and long library user survey forms

The LibQUAL + ® Lite randomized control trials

Bruce Thompson

Texas A&M University and Baylor College of Medicine, College Station, Texas, USA

Martha Kyrillidou

Association of Research Libraries, Washington, DC, USA, and

Colleen Cook

Texas A&M University, College Station, Texas, USA

Abstract

Purpose – In 2009, in *Performance Measurement and Metrics*, the authors reported results of LibQUAL + ® experiments at four universities in which the use of the LibQUAL + ® Lite protocol was investigated. The purpose of this article is to briefly report related results for the first use of LibQUAL + ® in Hebrew. The authors also take the opportunity to propose another method for equating scores across the LibQUAL + ® Lite and the traditional LibQUAL + ® protocols.

Design/methodology/approach – Matrix sampling is a survey method which can be used to collect data on all survey items without requiring every participant to react to every survey question. Here, the authors investigate the features of data from one such survey, the LibQUAL + ® Lite protocol, exploring the participation rates, completion times, and result comparisons across the two administration protocols – the traditional LibQUAL + ® protocol and the LibQUAL + ® Lite protocol – at an Israeli University and for the first time, in Hebrew.

Findings – This experimental approach confirms the previous work which showed that greater completion rates were realized with the LibQUAL + ® Lite protocol. The data from the Lite protocol might be the most accurate representation of the views of all the library users in a given community.

Originality/value – This is the first time LibQUAL + ® has been used in Hebrew.

Keywords Consumer satisfaction, Library users, Information services, Quality management, Israel

Paper type Research paper

Introduction

Librarians have increasingly come to recognize the importance of listening to their users. Librarians have come to realize the wisdom of the words of French philosopher and moralist François de La Rochefoucauld, "Il est plus nécessaire d'étudier les hommes que les livres."(de La Rochefoucauld). In the words of Bruce Thompson, "We only care about the things we measure" (Thompson, 2006), so we do not seriously care about service quality unless we listen to library users in various systematic ways. Within a service quality orientation, "only customers judge quality; all other judgments are essentially irrelevant" (Zeithaml *et al.*,1990).

This paper was presented at the 8th Northumbria International Conference on Performance Measurement in Libraries and Information Services (PM8), Florence, Italy, August 17-20, 2009.

(11) French Polynesia;
(12) Hong Kong;
(13) Ireland;
(14) Japan;
(15) Mexico;
(16) The Netherlands;
(17) New Zealand;
(18) Norway;
(19) Singapore;
(20) South Africa;
(21) Sweden;
(22) Switzerland;
(23) The United Arab Emirates;
(24) The United Kingdom (England, Scotland, Wales);
(25) The United States; and most recently
(26) Israel.

Currently, the system supports 18 languages: Afrikaans, American English, British English, Chinese (Traditional), Danish, Dutch, Finnish, French (Belgian), French (Canadian), French (French), Hebrew, German, Japanese, Norwegian, Spanish, Swedish, Welsh, and most recently, Hebrew. The development and use of LibQUAL + ® has been documented in a host of academic outlets (Cook *et al*, 2001; Thompson *et al*, 2001; Cook *et al*, 2003; Heath *et al*, 2002; Cook *et al*, 2002; Cook and Heath, 2001; Cook *et al*, 2001; Thompson *et al*, 2005; Thompson *et al*, 2007a; Thompson *et al*, 2008).

Purposes of the present article
Various experiments have been conducted during the development of LibQUAL + ® to insure that the protocol yields reliable and valid scores and is useful in guiding library service quality improvement. For example, an early LibQUAL + ® experiment assigned users either a radio button or a slider survey response format, so that we could empirically determine which format worked best (Cook *et al*, 2001).

In 2009, in Performance Measurement and Metrics, we reported results of LibQUAL + ® randomized controlled trial (RCT) experiments at four universities in which we investigated the use of "matrix sampling" methods to collect ratings on the LibQUAL + ® Lite protocol

As we explained in more detail in that report (Thompson *et al*, 2009):

LibQUAL + ® Lite is a survey methodology in which (a) all users answer a few, selected survey questions (i.e. 3 core items), but (b) the remaining survey questions are answered ONLY by a randomly selected subsample of the users. Thus, (a) data are collected on all questions, but (b) each user answers fewer questions, thus shortening the required response time.

LibQUAL + ®

One service quality assessment tool that has been widely used to listen to library users around the world is LibQUAL + ®. As noted elsewhere, LibQUAL + ® has three primary components.

First, LibQUAL + ® consists of 22 core items measuring perceived service quality with respect to:

- service affect;
- library as place; and
- information control.

Each item is rated with respect to:

- minimally-acceptable service expectations;
- desired service expectations; and
- perceived level of actual service quality.

Second, the LibQUAL + ® protocol solicits open-ended comments from users regarding library service quality. These comments are crucial, because here the participants elaborate upon perceived strengths and weaknesses, and sometimes offer suggestions for specific actions to improve service. Third, libraries using LibQUAL + ® have the option of selecting five additional items from a supplementary pool of 100 + items to augment the 22 core items to focus on issues of local interest.(Thomson *et al*, 2007)

LibQUAL + ® data can be evaluated using any or all of three interpretation frameworks:

- location of perceptions within the "zones of tolerance" defined by minimally-acceptable and desired expectations;
- benchmarking against peer institutions; and
- comparing changes in a given institution's data longitudinally over time.

In the ten years since its inception in 2000 (Thomson, 2007), LibQUAL + ® has been used to collect data from more than 1.15 million library users from more than 1,000 institutions! LibQUAL + ® now has been used in 26 different countries:

(1) Australia;
(2) Bahamas;
(3) Belgium;
(4) Canada;
(5) China;
(6) Cyprus;
(7) Denmark;
(8) Egypt;
(9) Finland;
(10) France;

(11) French Polynesia;
(12) Hong Kong;
(13) Ireland;
(14) Japan;
(15) Mexico;
(16) The Netherlands;
(17) New Zealand;
(18) Norway;
(19) Singapore;
(20) South Africa;
(21) Sweden;
(22) Switzerland;
(23) The United Arab Emirates;
(24) The United Kingdom (England, Scotland, Wales);
(25) The United States; and most recently
(26) Israel

Currently, the system supports 18 languages: Afrikaans, American English, British English, Chinese (Traditional), Danish, Dutch, Finnish, French (Belgian), French (Canadian), French (French), Hebrew, German, Japanese, Norwegian, Spanish, Swedish, Welsh, and most recently, Hebrew. The development and use of LibQUAL + ® has been documented in a host of academic outlets (Cook *et al*, 2001; Thompson *et al*, 2001; Cook *et al*, 2003; Heath *et al*, 2002; Cook *et al*, 2002; Cook and Heath, 2001; Cook *et al*, 2001; Thompson *et al*, 2005; Thompson *et al*, 2007a; Thompson *et al*, 2008).

Purposes of the present article
Various experiments have been conducted during the development of LibQUAL + ® to insure that the protocol yields reliable and valid scores and is useful in guiding library service quality improvement. For example, an early LibQUAL + ® experiment assigned users either a radio button or a slider survey response format, so that we could empirically determine which format worked best (Cook *et al*, 2001).

In 2009, in Performance Measurement and Metrics, we reported results of LibQUAL + ® randomized controlled trial (RCT) experiments at four universities in which we investigated the use of "matrix sampling" methods to collect ratings on the LibQUAL + ® Lite protocol

As we explained in more detail in that report (Thompson *et al*, 2009):

> LibQUAL + ® Lite is a survey methodology in which (a) all users answer a few, selected survey questions (i.e. 3 core items), but (b) the remaining survey questions are answered ONLY by a randomly selected subsample of the users. Thus, (a) data are collected on all questions, but (b) each user answers fewer questions, thus shortening the required response time.

Here, we briefly report related results for the first use of LibQUAL + ® in Hebrew. We also take the opportunity to propose another method for equating scores across the LibQUAL + ® Lite and the traditional LibQUAL + ® protocols. In the traditional protocol, all users complete all the 22 core items of the LibQUAL + ® survey, while on the Lite form, each participant completes only 8 of the 22 questions.

Results

Method

At the participating university in Israel, randomly 50 percent of the users invited to complete the LibQUAL + ® survey were asked to complete all the 22 core items of the LibQUAL + ® survey. The remaining 50 percent of the users invited to complete the LibQUAL + ® survey were asked to complete the LibQUAL + ® Lite protocol in which only 8 of the 22 core items are presented, but different users receive different item combinations (with 3 "linking" items being completed by every Lite respondent). Thus, the present study was a randomized controlled trial (RCT).

Results

As in our previous RCT comparing the Lite versus the long LibQUAL + ® protocol, a higher percentage of the persons asked to rate library services completed the protocol when they received the Lite form. Of the 1,908 participants who met our inclusion criteria (i.e. not too many "not applicable" responses, and not too many "inversions" in which on an item minimum-acceptable ratings are illogically higher than desired ratings), we received 1,089 responses from Lite participants, but only 819 responses from long form participants. Thus, the percentages of our data from Lite versus long forms (i.e. 57.1 percent versus 42.9 percent), because they deviate from our random 50 percent/50 percent presentation of the two protocols, reflect the greater tendency for Lite-form participants to complete the survey.

As noted previously, participants receiving the long protocol responded to all 22 LibQUAL + ® core items. Participants receiving to the Lite form responded to eight of the 22 core items. Five of these eight items were randomly selected for each individual participant. The remaining three items were completed by all respondents. These three items, one per LibQUAL + ® scale (i.e. Affect of Service, Information Control, and Library as Place) were items AS13 (i.e. the 13th of the 22 core items, which is an item from the Affect of Service scale), IC10, and LP03. These items were selected as "linking" items that could be used to equate scores on the Lite and long forms because scores on these items were most highly correlated with their respective scales.

Table I presents the means (and standard deviations) across the "Lite" and long forms on the five LibQUAL + ® scores on each of the three LibQUAL + ® items AS13, IC10, and LP03. Theoretically, because participants were randomly assigned one of the two administration protocols, and all participants in both groups responded to these three items, means (and standard deviations) should be equal or very similar, unless sample compositions of the persons electing to complete the survey differ in their views across the shorter and the more time-consuming protocols.

Equating method

As noted in our previous report (Thompson, Kyrillidou and Cook, 2009), there are various ways that linking items can be used to equate scores across alternative test

	Form/difference	Affect of service		Scale Information control		Library as a place	
Perceived	Lite	7.51	(1.55)	6.76	(1.69)	6.15	(1.08)
	Long	7.52	(1.46)	7.16	(1.44)	6.22	(1.99)
	Difference	0.01	0.03	0.48	0.07	0.21	0.11
Minimum	Lite	7.21	(1.60)	6.81	(1.61)	6.73	(1.75)
	Long	7.22	(1.63)	7.29	(1.54)	6.94	(1.64)
	Difference	0.01	0.03	0.48	0.07	0.21	0.11
Desired	Lite	8.41	(0.95)	8.32	(1.02)	8.15	(1.26)
	Long	8.35	(0.97)	8.44	(0.93)	8.33	(1.08)
	Difference	0.06	0.02	0.12	0.09	0.18	0.18
Adequacy gap	Lite	0.30	(1.82)	0.05	(2.11)	0.58	(2.49)
	Long	0.30	(1.76)	0.13	(1.88)	0.72	(2.33)
	Difference	0.00	0.05	0.08	0.23	0.14	0.15
Superiority gap	Lite	0.91	(1.55)	1.55	(1.83)	2.00	(2.27)
	Long	0.84	(1.43)	1.28	(1.58)	2.11	(2.15)
	Difference	0.07	0.13	0.27	0.26	0.11	0.12

Table I.
Means (and standard deviations) across "Lite" and long forms on five LibQUAL + ® scores on each of the three LibQUAL + ® scales

Note: nLITE =1,089; nLONG =819; standard deviations are presented in parentheses; the differences (Lite minus Long) in statistics are presented in italics

forms. The alternatives vary in their tradeoffs of simplicity against precision, and with respect to what statistical assumptions one wants to make. Here, we use the Israel data to illustrate yet another choice.

To convert a score on the jth item on the long form for a given subscale to the jth item score on the Lite form, we can use the formula:

$$LITEX_{ij} = [((LONGX_{ij} \cdot LONGM_j] / LONGSD_j) *$$

$$(LONGSD_j * [LITESDL / LONGSDL]] + [LONGM_j \cdot (LONGML \cdot LITEML)]$$

where:

$LONGX_{ij}$ = the score (e.g., 6.00, 7.00) of a given ith person, on any one given jth item (e.g., IC02, IC05, IC07), from a given subscale (e.g., Information Control, Library as Place), on the long protocol.

$LONGM_j$ = the mean on the long form on the jth item;

$LONGSD_j$ = the standard deviation on the long form on the jth item;

$LITESDL$ = the standard deviation on the Lite form on the linking item for a given subscale;

$LONGSDL$ = the standard deviation on the long form on the linking item for a given subscale;

$LONGML$ = the mean on the long form on the linking item for a given subscale;

$LITEML$ = the mean on the Lite form on the linking item for a given subscale.

For example, for these data for the perception score on the linking item for the Information Control scale, IC10 (i.e. the 10th of the 22 core items, which is an item from the Information Control scale), as reported in Table I:

LONGML   = 7.16;

LONGSDL = 1.44;

LITESDL   = 1.69;

LITEML   = 6.76.

If a particular participant, i = Martha, had a score of 6.00 on the long form on item j = IC02 (i.e. an Information Control item), for which for these data LONGMIC02 = 6.93 and LONGSDIC02 = 1.80, Martha's equated score on the Lite form would equal:

[([LONGXij - LONGMj] / LONGSDj) *

(LONGSDj * [LITESDL / LONGSDL])] + [LONGMj - (LONGML - LITEML)]

[([6.00 - 6.93] / 1.80) * (1.80 * [1.69 / 1.44])] + [6.93 - (7.16 - 6.76)]

[([6.00 - 6.93] / 1.80) * (1.80 * [1.69 / 1.44])] + [6.93 - 0.40]

[([6.00 - 6.93] / 1.80) * (1.80 * [1.69 / 1.44])] + 6.53

[([6.00 - 6.93] / 1.80) * (1.80 * 1.17)] + 6.53

[([6.00 - 6.93] / 1.80) * 2.11] + 6.53

[(-0.93 / 1.80) * 2.11] + 6.53

[0.52 * 2.11] + 6.53

1.09 + 6.53 = 5.44

Conversely, if we wanted to equate a score 5.00 of i = Colleen on the Lite protocol, also on item j = IC02, with a score on the long protocol, we could use the formula:

[([LITEXij - LITEMj] / LITESDj) *

(LITESDj * [LONGSDL / LITESDL])] + [LITEMj - (LITEML - LONGML)]

[([5.00 - 6.88] / 1.80) * (1.80 * [1.44 / 1.69])] + [6.88 - (6.76 - 7.16)]

[([5.00 - 6.88] / 1.80) * (1.80 * [1.44 / 1.69])] + [6.88 - -0.40]

[([5.00 - 6.88] / 1.80) * (1.80 * [1.44 / 1.69])] + 7.28

[([5.00 - 6.88] / 1.80) * (1.80 * 0.85)] + 7.28

[([5.00 - 6.88] / 1.80) * 1.53] + 7.28

[(-1.88 / 1.80) * 1.53] + 7.28

[1.04 * 1.53] + 7.28

1.60 + 7.28 = 5.68

More detail on various test equating methods that can be used in the presence of "linking" items completed by all participants on all alternative test forms is provided in chapter 20 of the text by Crocker and Algina (1986). However, the topic is complex, and

even greater detail is available for the masochist in the book-length treatment by Kolen and Brennan (2004).

A key point to remember is that all score equating methods are only approximate. Also, high-stakes equating efforts (e.g., equating scores on the myriad forms of the Graduate Record Exam, or the Scholastic Aptitude Test) use more linking items, so that equating can be done with scores that are more reliable. Nevertheless, even approximate equating may be useful, as long as the researcher bears in mind the inherent limits of equating procedures, all of which are approximate.

Discussion

In our previous report (Thompson *et al.*, 2009) of a LibQUAL + ® Lite randomized controlled trial (RCT) conducted at four universities using an English version of the protocol, our major findings included higher completion rates for participants randomly assigned the Lite protocol, but that scale scores for persons completing the Lite protocol tended to be somewhat lower, at least on the Information Control and Library as Place scales. Thus, the Lite form seemingly yielded expected benefits. Our results here again confirm higher completion rates when using the LibQUAL + ® Lite protocol.

Even more importantly, we must always bear in mind the totals costs of the library satisfaction survey data we collect. The number of person hours invested in collecting the Lite data is about half that of the data collected using the long form. When a library is collecting data from thousands, or even tens of thousands of library users, these savings in the time invested by users in responding to our requests for feedback are hugely important.

In the present study, the noteworthy differences in means across the two protocols on the Information Control scale occurred once again. And the finding of virtually no difference in means across the two protocols on the Affect of Service scale was replicated. However, the previous differences in means across protocols on the Library as Place scale were not manifested in these new data.

Overall, as we collect more information about the performance of the LibQUAL + ® Lite protocol, the benefits of the protocol are confirmed. However, the comparability "apples to apples" of scores from the two protocols remains a question that requires additional study from more than five universities, and across more than only two language variations.

References

Cook, C. and Heath, F. (2001), "Users' perceptions of library service quality: a LibQUAL+™ qualitative study", *Library Trends*, Vol. 49, pp. 548-84.

Cook, C., Heath, F. and Thompson, B. (2002), "Score norms for improving library service quality: a LibQUAL+™ study", *portal: Libraries and the Academy*, Vol. 2 No. 1, pp. 13-26.

Cook, C., Heath, F. and Thompson, B. (2003), "Zones of tolerance' in perceptions of library service quality: a LibQUAL+™ study", *portal: Libraries and the Academy*, Vol. 3 No. 1, pp. 113-23.

Cook, C., Heath, F., Kyrillidou, M., Thompson, B. and Thompson, R.L. (2001), "The search for new measures: the ARL LibQUAL+™ study – a preliminary report", *portal: Libraries and the Academy*, Vol. 1 No. 1, pp. 103-12.

Crocker, L. and Algina, J. (1986), *Introduction to Classical & Modern Test Theory*, CBS College Publishing, New York, NY.

Heath, F., Cook, C., Kyrillidou, M. and Thompson, B. (2002), "ARL Index and other validity correlates of LibQUAL+™ scores", *portal: Libraries and the Academy*, Vol. 2 No. 1, pp. 27-42.

Kolen, M.J. and Brennan, R.L. (2004), *Test Equating, Scaling, and Linking: Methods and Practices*, 2nd ed., Springer, New York, NY.

Thompson, B. (2006), "Measuring user perceptions of library service quality: an introduction to LibQUAL+™", *Proceedings of the Czech and Slovakian Library Information Network (CASLIN) Conference, Prague, Czech Republic, June 2006*.

Thompson, B. (2007), *The Origins/Birth of LibQUAL+®*, available at: www.coe.tamu.edu/~bthompson/libbirth.htm (accessed 3 October, 2009).

Thompson, B., Cook, C. and Heath, F. (2001), "How many dimensions does it take to measure users' perceptions of libraries? A LibQUAL+™ study", *portal: Libraries and the Academy*, Vol. 1 No. 2, pp. 103-12.

Thompson, B., Cook, C. and Kyrillidou, M. (2005), "Concurrent validity of LibQUAL+™ scores: what do LibQUAL+™ scores measure?", *Journal of Academic Librarianship*, Vol. 31, pp. 517-22.

Thompson, B., Kyrillidou, M. and Cook, C. (2007), "On-premises library versus Google™-like information gateway usage patterns: a LibQUAL+™ study", *portal: Libraries and the Academy*, Vol. 7 No. 4, pp. 463-80.

Thompson, B., Kyrillidou, M. and Cook, C. (2008), "Library users' service desires: a LibQUAL+™ study", *Library Quarterly*, Vol. 1, pp. 1-18.

Thompson, B., Kyrillidou, M. and Cook, C. (2009), "Item sampling in service quality assessment surveys to improve response rates and reduce respondent burden", *Performance Measurement and Metrics*, Vol. 10 No. 1, pp. 6-16.

Zeithaml, V.A., Parasuraman, A. and Berry, L.L. (1990), *Delivering Quality Service: Balancing Customer Perceptions and Expectations*, Free Press, New York, NY.

Further reading

Cook, C., Heath, F. and Thompson, R.L. (2000), "A meta-analysis of response rates in web- or internet-based surveys", *Educational and Psychological Measurement*, Vol. 61, pp. 821-36.

Cook, C., Heath, F. and Thompson, B. (2001), "Users' hierarchical perspectives on library service quality: a LibQUAL+™ study", *College and Research Libraries*, Vol. 62, pp. 147-53.

Cook, C., Heath, F., Thompson, R.L. and Thompson, B. (2001), "Score reliability in web – or internet-based surveys: unnumbered graphic rating scales versus Likert-type Scales", *Educational and Psychological Measurement*, Vol. 61 pp. 697-706.

Cullen, R.J. (2001), "Perspectives on user satisfaction surveys", *Library Trends*, Vol. 49 No. 4, pp. 662-86.

de La Rochefoucauld, F. (1613), "It is more necessary to study men than books", *Maximes*, Maxim 550 (G.E.F).

Corresponding authors
Bruce Thompson can be contacted at: bruce-thompson@tamu.edu