greece.wp1 5/29/08

Using National and International Score Norms as a Library Service Quality Benchmarking Tool

Bruce Thompson

Texas A&M University
and
Baylor College of Medicine (Houston)

C. Colleen Cook

Texas A&M University

Martha Kyrillidou

Association of Research Libraries

Bruce Thompson is distinguished professor of educational psychology and CEHD distinguished research fellow, and distinguished professor of library science, Texas A&M University, College Station, TX, and adjunct professor of family and community medicine, Baylor College of Medicine, Houston, TX. He may be contacted via e-mail at: bruce-thompson@tamu.edu.

Colleen Cook is dean of libraries and holder of the Sterling C. Evans Chair, Texas A&M University, College Station, TX. She may be contacted at: ccook@tamu.edu.

Martha Kyrillidou is director of statistics and service quality programs at the Association of Research Libraries, Washington, DC. She may be contacted at: martha@arl.org.

Paper presented at the 17th meeting of the Greek Academic Libraries, Ioannina, Greece, September 24-26, 2008.

Abstract

There are three primary mechanisms for interpreting library service quality assessment scores. First, we can compare our scores for one year with the scores we obtained on the same measure in previous years. This is the ultimate form of benchmarking, because no one is more like us than us. Second, we can compare the Perception scores we received against ratings for both Minimally-Acceptable service scores and Desired service scores. The second set of two scores creates a "zone of tolerance" against which we can interpret users' service quality assessment ratings of Perceived service quality. Third, we can interpret the scores we receive against the scores of similar, peer institutions. Some library service quality assessment protocols (e.g., LibQUAL+®) allow libraries to use any combination or all of these three interpretation frameworks. The present paper explains the use of a powerful benchmarking mechanism called norms tables. Norms tables can be created for protocols involving either large numbers of users or large numbers of libraries.

LibQUAL+® is a library service quality assessment and improvement protocol that uses the internet to measure library users' perceptions of library service quality. The history of LibQUAL+®'s birth has been detailed by Thompson (2007). To date, LibQUAL+® has been used to collect data from roughly 1,000,000 library users, from more than 1,000 institutions!

LibQUAL+® has been used in geographically diverse locations, including the United States, Canada, Mexico, Australia, New Zealand, the United Kingdom (England, Scotland, Wales), France, Ireland, the Netherlands, Switzerland, Germany, Denmark, Finland, Norway, Sweden, Egypt, the United Arab Emirates, and South Africa. A Chinese version was implemented in Hong Kong in the Fall of 2007. Currently, the system supports 12 languages: Afrikaans, American English, British English, Chinese (Traditional), Danish, Dutch, Finnish, French (Canadian), French (European), German, Norwegian, and Swedish. The various editions of LibQUAL+® have been used over a period of eight years.

LibQUAL+® has been an important tool for service quality improvement from the local perspective of a given library (Cook, 2002a; Heath, Kyrillidou, & Askew, 2004; Kyrillidou, 2006). However, the LibQUAL+® database is important in its own right, given the sheer scale, the cultural diversity (Kyrillidou & Persson, 2006), and the longitudinal features of the data (Thompson, Cook & Kyrillidou, 2006b; Thompson, Kyrillidou & Cook, 2007a, 2007b, 2008). The roots of LibQUAL+® within the long-standing culture of assessment characterizing the Association of Research Libraries programmatic efforts has been extensively

documented (cf. Kyrillidou & Cook, in press; Kyrillidou & Heath, 2001).

The grounding of LibQUAL+® within the perceptions and interests of library users is documented by Cook (2002b) and Cook and Heath (2001). Investigations of the reliability and validity of LibQUAL+® scores have been extensively documented (Cook, Heath, Thompson & Thompson, 2001; Cook & Thompson, 2001; Heath, Cook, Kyrillidou & Thompson, 2002; Thompson & Cook, 2002; Thompson, Cook & Heath, 2001, 2003a, 2003b; Thompson, Cook & Kyrillidou, 2005; Thompson, Cook & Thompson, 2002). The properties of LibQUAL+® scores have even been investigated with sophisticated methods such as taxonometric analysis (Arnau, Thompson & Cook, 2001) and higher-order factor analysis (Cook, Heath & Thompson, 2001).

In addition to (a) providing open-ended comments, which historically roughly 40% of participants do, and (b) rating five service quality items selected by institutions or consortia from an optional item pool consisting of more than 100 ancillary items, LibQUAL+® participants (c) rate each of the 22 LibQUAL+® core items on a 1-to-9 scale (9 is highest) with respect to (a) perceived levels of actual service quality, (b) the minimally-acceptable level of service with respect to a given item, and (c) the desired level of service with respect to a given item. The difference between the desired rating and the minimally-acceptable rating for a given item is the "*zone of tolerance*" for that item. Ideally, perception ratings will fall within the zone.

The 22 core LibQUAL+® items create three subscales: (a) Affect of Service ("AS"), (b) Information Control ("IC"), and (c) Library

as Place ("LP"). There are respectively 9, 8, and 5 items constituting these three LibQUAL+® subscales.

LibQUAL+® is *unique* in that LibQUAL+® perception scores can be interpreted in any combination of three ways: (a) by benchmarking against peer institutions, (b) by benchmarking against oneself longitudinally, or (c) by interpreting perception ratings in the context of the zones of tolerance for given items (Cook, Heath & Thompson, 2003; Thompson, Cook & Heath, 2000). **If the aggregated perception ratings for a given institution that cannot be interpreted using one or more of these three interpretation frameworks, then the ratings simply cannot be interpreted** (i.e., are meaningless)!

Philosophical Underpinnings of LibQUAL+®

Three quotations have come to be frequently used in introducing LibQUAL+® to new participants. The first quotation is used to emphasize that, in the digital era, collection counts are no longer viable as the sole metric for measuring library quality. On this point, we often cite French philosopher and moralist François de La Rochefoucauld (1680, p. 51, line 106), who noted that "Il est plus nécessaire d'étudier les hommes que les livres."

Second, we emphasize that, within a service quality orientation, "only customers judge quality; all other judgments are essentially irrelevant" (Zeithaml, Parasuraman & Berry, 1990, p. 16). Third, we cite Bruce Thompson (2006), who noted that "We only care about the things we measure" (p. 1). We do not really care about user perceptions of library service quality, unless we periodically measure these perceptions in various systematic ways.

Purpose of the Present Paper

Score norms are one vehicle for benchmarking against peer institutions. The use of LibQUAL+® norms has been explained in Cook, Heath and Thompson (2002). Score norms allow the conversion of a library's mean (i.e., average) scores into percentile ranks, which indicate *what percentage of either (a) individual protocol respondents, or (b) institutions, generated lower LibQUAL+® scores.*

The present paper was written in service of two purposes. First, some key norms concepts are explained. Second, some comparisons of LibQUAL+® norms for total perceptions averages (i.e., library users' mean ratings on perceived library quality across the 22 LibQUAL+® core questions) are presented for 2007 data across (a) Association of Research Libraries (ARL) participants in North America, (b) Society of College, National, and University Libraries (SCONUL) participants in the United Kingdom, (c) the Canadian academic libraries, most of which participated in LibQUAL+® in 2007, and (d) Hong Kong libraries who presented the survey to their users in 2007 in Traditional Chinese.

Key Norms Concepts

Norms and Benchmarking

Norms tell us how scores "stack up" within a particular user group. For example, on the 1-to-9 ("9" is highest) scale, users might provide a mean "perceived" rating of 6.5 on an item, "complete run of journal titles."

A total market survey administered to hundreds of thousands of users, as is LibQUAL+®, affords the unique opportunity to ask normative questions such as, "How does a mean 'perceived' score of

6.5 stack up among all individual users who completed the survey?", or "How does a mean 'perceived' score of 6.5 stack up among all institutions at which users completed the survey?" For example, if 70% of 100,000 individual users generated "perceived" ratings lower than 6.5, 6.5 might not be so bad.

Only norms give us insight into this comparative, benchmarking perspective. And a local user-satisfaction survey (as against a total market survey) can never give us this insight!

For several years, both (a) individual and (b) institutional LibQUAL+® score norms have been published for certain types of users (e.g., undergraduate students, postgraduate students, faculty) and certain types of libraries (e.g., major research libraries, medical school libraries). These norms conversion tables are reported at:

http://www.coe.tamu.edu/~bthompson/libq2005.htm

http://www.coe.tamu.edu/~bthompson/libq2004.htm

http://www.coe.tamu.edu/~bthompson/libq03b.htm

http://www.coe.tamu.edu/~bthompson/libq2003.htm

http://www.coe.tamu.edu/~bthompson/libq2002.htm

http://www.coe.tamu.edu/~bthompson/servnorm.htm

Thompson, Cook and Kyrillidou (2006a) presented evidence that LibQUAL+® score norms tend to be quite stable.

Common Misconception Regarding Norms

An unfortunate and incorrect misconception is that norms make value statements. Norms do not make value statements! Norms make fact statements. If you are a forest ranger, and you earn $25,000 a year, a norms table might inform you of the fact statement that

you make less money than 85% of the adults in the United States.

But if you love the outdoors, you do not care very much about money, and you are very service-oriented, this fact statement might not be relevant to you. Or, in the context of your values, you might interpret this fact as being quite satisfactory.

Substantive Uses of Norms

Norms can also be used for substantive research purposes, and not only for benchmarking. For example, percentile equivalents might be compared across (a) years or (b) user groups or (c) locations to determine whether ratings are similar at different points in the ratings distributions. Even if the score means or medians are the same across user groups, the score distributions are not necessarily the same. Here are some illustrative patterns that might be detected when making such comparisons in such cases:

1. The users in the two groups (e.g., time, roles, locations) have similar mean or median ratings, but the users in one group are more homogeneous in their ratings (i.e., ratings in the lower percentile ranks are higher, while ratings in the higher percentile ranks are lower).
2. The users in the two groups (e.g., time, roles, locations) have similar mean or median ratings, but Group "B" users in the lower percentile ranks have more positive ratings than users in Group "A", while Group "B" users in the higher percentile ranks have less positive ratings than users in Group "A".

Comparison of Four International Cohorts

Table 1 presents percentile equivalents for Total LibQUAL+®

Perception ratings respectively for ARL (n=28,430), SCONUL (n=17,648), Canadian (n=48,085), and Hong Kong (n=8,196) LibQUAL+® 2007 participants. Here are some patterns that emerge from these comparisons.

INSERT TABLE 1 ABOUT HERE

1. The median ratings (i.e., the 50th percentile) are similar (i.e., 7.14, 6.86, 7.18, 6.73), with the rank ordering of ratings from most to least positive being Canadian, ARL, SCONUL, and Hong Kong users.
2. The ranges or dispersions of the ratings were comparable, with the rank ordering of score dispersions from most to least dispersed/heterogeneous SCONUL (3.32-9.00), ARL (3.67-9.00), Canadian (3.82-9.00), and Hong Kong (4.00-8.77) users.
3. In the lower ends of the score distributions, the Hong Kong users tended to generate the highest scores (even though the median was lowest for these users), while the SCONUL users tended to generate the lowest scores.
4. In the higher ends of the score distributions, the Hong Kong users tended to generate the lowest scores, while the ARL and the Canadian users tended to generate the highest scores.

Discussion

In the aggregate, the users from all four cohorts had reasonably similar perceptions of their libraries. However, in the lower parts of the ratings distributions, the Hong Kong users were somewhat more positive than users in the other cohorts. In the higher parts of the distributions, the Hong Kong users were less

positive than users in the other cohorts. Of course, other patterns might be revealed within different user subgroups (e.g., undergraduate students, postgraduate students, faculty), or on other LibQUAL+® scores (e.g., subscale scores, or superiority or adequacy gap scores).

These results illustrate the use of norms for substantive inquiry purposes. However, the greatest practical utility of norms lies in benchmarking against the results of peer institutions, as discussed in more detail by Cook, Heath and Thompson (2002). As more and more institutions of higher education realize that they are players in a larger global economy of educational assets, international benchmarking is becoming increasingly important.

References

Arnau, R.C., Thompson, R.L., & Cook, C. (2001). Do different response formats change the latent structure of responses?: An empirical investigation using taxonometric analysis. Educational and Psychological Measurement, 61, 23-44.

Cook, C. (Ed.). (2002a). The maturation of assessment in academic libraries: The role of LibQUAL+™ (special issue). Performance Measurement and Metrics, 3(2), 34-112.

Cook, C.C. (2002b). A mixed-methods approach to the identification and measurement of academic library service quality constructs: LibQUAL+™. (Doctoral dissertation, Texas A&M University, 2001). Dissertation Abstracts International, 62, 2295A. (University Microfilms No. AAT3020024)

Cook, C., & Heath, F. (2001). Users' perceptions of library service quality: A "LibQUAL+™" qualitative study. Library Trends, 49, 548-584.

Cook, C., Heath, F., & Thompson, B. (2001). Users' hierarchical perspectives on library service quality: A "LibQUAL+™" study. College and Research Libraries, 62, 147-153.

Cook, C., Heath, F. & Thompson, B. (2002). Score norms for improving library service quality: A LibQUAL+™ study. portal: Libraries and the Academy, 2, 13-26.

Cook, C., Heath, F. & Thompson, B. (2003). "Zones of tolerance" in perceptions of library service quality: A LibQUAL+™ study. portal: Libraries and the Academy, 3, 113-123.

Cook, C., Heath, F., Thompson, R.L. & Thompson, B. (2001). Score reliability in Web- or Internet-based surveys: Unnumbered

graphic rating scales versus Likert-type scales. Educational and Psychological Measurement, 61, 697-706.

Cook, C., & Thompson, B. (2001). Psychometric properties of scores from the Web-based LibQUAL+™ study of perceptions of library service quality. Library Trends, 49, 585-604.

de La Rochefoucauld, F. (1680). Maximes posthumes.

Heath, F., Cook, C., Kyrillidou, M., & Thompson, B. (2002). ARL Index and other validity correlates of LibQUAL+™ scores. portal: Libraries and the Academy, 2, 27-42.

Heath, F., Kyrillidou, M., & Askew, C. (Eds.) (2004). Libraries act on their LibQUAL+™ findings: From data to action (special issue). Journal of Library Administration, 40, 1-239.

Kyrillidou, M. (2006). Service quality: A perceived outcome for libraries. In P. Hernon, R. E. Dungan, & C. Schwartz (Eds.), Revisiting outcomes assessment in higher education (pp. 331-366). Westport and London: Libraries Unlimited.

Kyrillidou, M., & Cook, C. (in press). The evolution of measurement and evaluation of libraries: A perspective from the Association of Research Libraries. Library Trends, 57.

Kyrillidou, M., & Heath, F. (2001). Measuring service quality: Introduction. Library Trends, 49, 541-547.

Kyrillidou, M., & Persson, A. (2006). The new library user in Sweden: A LibQUAL+™ study at Lund University. Performance Measurement and Metrics, 7, 45-53.

Thompson, B. (2006, June). Measuring user perceptions of library service quality: An introduction to LibQUAL+™. Paper presented at the Czech and Slovakian Library Information

Network (CASLIN) conference, Prague, Czech Republic.

Thompson, B. (2007). The origins/birth of LibQUAL+®. **<http://www.coe.tamu.edu/~bthompson/libbirth.htm>.**

Thompson, B., & Cook, C. (2002). Stability of the reliability of LibQUAL+™ scores: A "Reliability Generalization" meta-analysis study. Educational and Psychological Measurement, 62, 735-743.

Thompson, B., Cook, C., & Heath, F. (2000). The LibQUAL+™ gap measurement model: The bad, the ugly, and the good of gap measurement. Performance Measurement and Metrics, 1, 165-178.

Thompson, B., Cook, C., & Heath, F. (2001). How many dimensions does it take to measure users' perceptions of libraries?: A "LibQUAL+™" study. portal: Libraries and the Academy, 1, 129-138.

Thompson, B., Cook, C., & Heath, F. (2003a). Structure of perceptions of service quality in libraries: A LibQUAL+™ study. Structural Equation Modeling, 10, 456-464.

Thompson, B., Cook, C. & Heath, F. (2003b). Two short forms of the LibQUAL+™ survey assessing users' perceptions of library service quality. Library Quarterly, 73, 453-465.

Thompson, B., Cook, C., & Kyrillidou, M. (2005). Concurrent validity of LibQUAL+™ scores: What do LibQUAL+™ scores measure? Journal of Academic Librarianship, 31, 517-522.

Thompson, B., Cook, C., & Kyrillidou, M. (2006a, April). Stability of library service quality benchmarking norms across time and cohorts: A LibQUAL+™ study. Paper presented at the Asia-Pacific Conference of Library and Information Education and

Practice, Singapore.
<http://www.coe.tamu.edu/~bthompson/libq2005.htm>

Thompson, B., Cook, C., & Kyrillidou, M. (2006b). Using localized survey items to augment standardized benchmarking measures: A LibQUAL+™ study. portal: Libraries and the Academy, 6, 219-230.

Thompson, B., Cook, C., & Thompson, R.L. (2002). Reliability and structure of LibQUAL+™ scores: Measuring perceived library service quality. portal: Libraries and the Academy, 2, 3-12.

Thompson, B., Kyrillidou, M., & Cook, C. (2007a). On-premises library versus Google™-like information gateway usage patterns: A LibQUAL+® study. portal: Libraries and the Academy, 7, 463-480.

Thompson, B., Kyrillidou, M., & Cook, C. (2007b). User library service expectations in health science versus other settings: A LibQUAL+® study. Health Information and Libraries Journal. 24(Supplement 1), 38-45.

Thompson, B., Kyrillidou, M., & Cook, C. (2008). Library users' service desires: A LibQUAL+® study. Library Quarterly, 1, 1-18.

Zeithaml, V.A., Parasuraman, A., & Berry, L.L. (1990). Delivering quality service: Balancing customer perceptions and expectations. New York: Free Press.

Table 1
Percentile Equivalents for Total Scores Respectively for ARL (leftmost, n=28,430), SCONUL (n=17,648), Canadian (n=48,085), and Hong Kong (rightmost, n=8,196) LibQUAL+® 2007 Participants

Percentiles	1	3.67136	3.31818	3.81564	3.99864
	2	4.31818	3.95455	4.36364	4.40909
	3	4.63636	4.27883	4.70588	4.63636
	4	4.85714	4.50000	4.90909	4.81818
	5	5.00000	4.68182	5.05993	4.95455
	6	5.14286	4.81818	5.22727	5.00000
	7	5.27273	4.92857	5.35000	5.09524
	8	5.40000	5.00000	5.45455	5.18182
	9	5.50000	5.10000	5.57143	5.27273
	10	5.59091	5.19905	5.66667	5.31818
	11	5.68182	5.28087	5.72727	5.40909
	12	5.77273	5.36364	5.81818	5.45455
	13	5.83360	5.45000	5.90000	5.50000
	14	5.90909	5.50000	5.95455	5.54545
	15	5.95455	5.57895	6.00000	5.59591
	16	6.00000	5.63636	6.05804	5.63636
	17	6.06667	5.68182	6.13636	5.68299
	18	6.13333	5.73684	6.18182	5.72727
	19	6.17647	5.80000	6.22727	5.77273
	20	6.22222	5.85714	6.27273	5.81818
	21	6.27273	5.90476	6.31818	5.86364
	22	6.31250	5.95186	6.36364	5.90909
	23	6.35000	6.00000	6.40909	5.95455
	24	6.38095	6.00000	6.45000	6.00000
	25	6.41176	6.05000	6.47619	6.00000
	26	6.45455	6.09524	6.50000	6.04545
	27	6.50000	6.13636	6.54545	6.09091
	28	6.54545	6.18182	6.59091	6.13636
	29	6.57143	6.22727	6.61905	6.13636
	30	6.59091	6.25000	6.63636	6.18182
	31	6.63636	6.28571	6.68182	6.21429
	32	6.66667	6.31818	6.71429	6.22727
	33	6.68421	6.36364	6.72727	6.27273
	34	6.72727	6.40600	6.77273	6.27273
	35	6.76190	6.42105	6.80000	6.31818
	36	6.77273	6.45455	6.81818	6.35000
	37	6.81818	6.50000	6.85714	6.36364
	38	6.85000	6.52381	6.86364	6.40909
	39	6.86364	6.54545	6.90909	6.44444
	40	6.90476	6.57895	6.94118	6.45455
	41	6.90909	6.59091	6.95455	6.50000
	42	6.95455	6.63636	7.00000	6.50000
	43	6.95455	6.66667	7.00000	6.54545
	44	7.00000	6.68182	7.04545	6.57143
	45	7.00000	6.72727	7.04545	6.59091
	46	7.04545	6.73333	7.09091	6.61905
	47	7.09091	6.77273	7.09091	6.63636
	48	7.09524	6.80952	7.13636	6.68182
	49	7.13636	6.81818	7.13636	6.68182

Percentiles	50	7.14286	6.86364	7.18182	6.72727
	51	7.18182	6.88235	7.18182	6.72727
	52	7.19048	6.90909	7.22727	6.77273
	53	7.22727	6.95000	7.22727	6.77273
	54	7.25000	6.95455	7.27273	6.80952
	55	7.27273	7.00000	7.27778	6.81818
	56	7.30000	7.00000	7.31818	6.85714
	57	7.31818	7.04545	7.31818	6.86364
	58	7.35000	7.09091	7.36364	6.90909
	59	7.36364	7.09091	7.36842	6.90909
	60	7.40000	7.13636	7.40909	6.95455
	61	7.40909	7.14286	7.42105	6.95455
	62	7.45000	7.18182	7.45455	7.00000
	63	7.45455	7.19048	7.47368	7.00000
	64	7.50000	7.22727	7.50000	7.00000
	65	7.50000	7.26316	7.52381	7.04545
	66	7.54545	7.27273	7.54545	7.09091
	67	7.57143	7.31818	7.57143	7.09091
	68	7.59091	7.33333	7.59091	7.13636
	69	7.61905	7.36364	7.61905	7.13636
	70	7.63636	7.40000	7.63636	7.18182
	71	7.66667	7.40909	7.66667	7.18676
	72	7.68182	7.45455	7.68182	7.22727
	73	7.72727	7.47619	7.72222	7.25000
	74	7.75000	7.50000	7.72727	7.27273
	75	7.77273	7.54545	7.77273	7.31818
	76	7.80000	7.57143	7.78947	7.31818
	77	7.81818	7.59091	7.81818	7.36364
	78	7.85714	7.63636	7.85000	7.38095
	79	7.88235	7.66667	7.86364	7.40909
	80	7.90909	7.68182	7.90476	7.45455
	81	7.95238	7.72727	7.90909	7.47619
	82	7.95455	7.77273	7.95455	7.50000
	83	8.00000	7.80952	8.00000	7.54545
	84	8.04545	7.81818	8.00000	7.58341
	85	8.05556	7.86364	8.04545	7.59091
	86	8.09524	7.90909	8.09091	7.63636
	87	8.13636	7.95455	8.11111	7.68182
	88	8.18182	8.00000	8.15000	7.72727
	89	8.22727	8.04545	8.20000	7.77273
	90	8.27273	8.09091	8.25000	7.81818
	91	8.31818	8.13636	8.30000	7.86364
	92	8.38095	8.19048	8.36364	7.90909
	93	8.45455	8.27273	8.40909	8.00000
	94	8.52381	8.33333	8.47619	8.04545
	95	8.59091	8.42857	8.54545	8.13636
	96	8.68182	8.54545	8.63636	8.22727
Percentiles	97	8.77273	8.63158	8.73684	8.36364
	98	8.90909	8.77273	8.86364	8.52381
	99	9.00000	9.00000	9.00000	8.77273