Score Norms for Improving Library Service Quality: A LibQUAL+™ Study[1]

Colleen Cook, Fred Heath, and Bruce Thompson

abstract: Based on data from 20,416 LibQUAL+™ respondents from forty-three universities, the authors developed norm tables to allow librarians to interpret LibQUAL+™ scores with respect to typical profiles at other universities. Norms were developed for both "perceived" service scores and "gap" scores (e.g., "perceived" performance minus "minimally acceptable" performance). Norms such as these will assist library managers in decision-making by identifying (a) specific areas for needed improvement, (b) specific areas of needed additional service quality information (e.g., focus groups), and (c) peer institutions from which superior service practices can be modeled.

Librarians, library users, and university administrators increasingly have recognized that library input measures, such as counts of collections, are not sufficient to measure all aspects of library quality. Users also expect to receive quality service in accessing both physical and virtual collections. Even libraries with superb collections are likely to receive poor assessments from their users when service quality is perceived to be lacking.

Librarians, library users, and university administrators increasingly have recognized that library input measures, such as counts of collections, are not sufficient to measure all aspects of library quality.

Particularly in the contemporary milieu of increasing technological change and demands for accountability, "a measure of library quality based solely on collections has become obsolete."[2, 3] These various considerations have prompted the Association of Research Libraries (ARL) to sponsor a number of "New Measures" initiatives that focus on *outcome* measures, such as assessments of service quality and satisfaction. One such initiative has been the "LibQUAL+™" project.[4]

portal: Libraries and the Academy, Vol. 2, No. 1 (2002), pp. 13–26.

Phase Zero and Phase One Initiatives

In the phase zero 1999–2000 academic year the pilot web-based LibQUAL+™ tool consisting of forty-one items (22 SERVQUAL plus 19 trial items) was completed by 4,407 participants from thirteen ARL institutions. Various analyses were conducted to investigate the integrity of LibQUAL+™ scores. These studies were wide reaching. For example, R.C. Arnau, R.L. Thompson and Colleen Cook determined in a taxonometric study that library users think about library service quality by making refined continuous judgments, rather than by employing crude categorical judgments (e.g., merely "good" versus "bad").[5]

In the phase one 2000–2001 academic year, the LibQUAL+™ survey was further reformed based on phase zero results.[6] Data from this year afford a unique opportunity to elaborate LibQUAL+™ norm conversion tables, because (a) the sample of participants was so large (n = 20,416) and (b) represented participants from many (i.e., forty-three), diverse universities.

Purpose of the Study

Cook and Thompson argued that one of the LibQUAL+™ applications of greatest potential was the creation of LibQUAL+™ norms to allow librarians to place individual library performance in a context of perceived performance of selected peers.[7] As defined by Sanders and the Joint Committee on Standards for Educational Evaluation, a norm is "a single value, or a distribution of values, constituting the typical performance of a given group".[8]

As A. Anastasi observed,

> To say that an individual has correctly solved 15 problems on a mathematical reasoning test, or identified 34 words in a vocabulary test, or successfully assembled a mechanical object in 57 seconds conveys *little or no information about... standing* on any of these functions. [Thus scores] . . . are most commonly interpreted by reference to *norms*, which represent the test performance of the standardization sample.[9]

As L. Crocker and J. Algina explained,

> . . . normative scores provide information about an examinee's performance in comparison to the score distribution of some norm sample or reference group... [T]he meaningfulness of these scores depends on (1) the extent to which the test user is interested in comparing the examinee to the normative population and (2) the adequacy of the norming sample in representing that population.[10]

In summary, the present study was conducted to address three research questions:

1. Can LibQUAL+™ score norms at the level of the *individual user* be developed for use in the research library context?
2. How comparable are the conversion tables for LibQUAL+™ total scores across three user groups: undergraduate students, graduate students, and faculty?
3. Can LibQUAL+™ score norms at the *institutional level* be developed for use in the research library context?

Types and Units of Analysis

Norm Types

There are numerous choices with regard to the type of norm scores into which raw scores (e.g., number of right answers, the sum of item scores on an attitude measure) can be converted (e.g., IQ, grade-equivalent scores). In the present study, norms were developed to convert LibQUAL+™ scale and total raw scores into two types of norm scores.

First, norm tables were developed to convert LibQUAL+™ raw scores into "percentile ranks." Percentile ranks indicate the percentage of participants (or institutions, in the case of group norms) scoring less than a given raw score. A percentile rank (i.e., a percentage of *people*) is *not* the same as a percentage score, which is a percentage of *items* (e.g., percentage of items correct), although the meaning and correct use of percentile ranks can be confusing.[11]

Second, norm tables were developed to convert LibQUAL+™ raw scores into what are called "*T* scores" that can provide an easily interpretable score metric. *T* scores in the normative or standardization sample have means of 50.0 and standard deviations of 10.0. Although they are scaled differently, *T* scores are not unlike IQ scores, which have means of 100.0 and often standard deviations of 15.0.

Units of Analysis

Norm conversion tables can be developed at different levels of analysis. For example, a librarian may know that a particular individual user generated a total score of 5.44 on the twenty-five LibQUAL+™ total item set, or that the mean of individual users on the campus was 5.44. The librarian may want to know how many individual users nationally rated their libraries lower than this 5.44. This application may require norm conversion tables created at the unit of analysis of the *individual*.

Or a librarian may know that users on his/her campus generated a mean total score on the twenty-five LibQUAL+™ total item set of 6.176. The librarian may want to know how many university libraries nationally were rated lower by their users. This application may require norm conversion tables created at the unit of analysis of the *institution*.

Results

Individual-level Norms

Table 1 presents the norm conversion table for transforming LibQUAL+™ total scores, computed by summing an individual's responses to the twenty-five "perceived service quality" items, and then dividing the summated score by twenty-five, into (a) percentile rank and (b) standardized *T* scores. Similar norms might also be developed for "perceived" ratings for each of the LibQUAL+™ subscale scores: (a) Service Affect, (b) Library as Place, (c) Personal Control, and (d) Information Access.

Table 1

Norms for Converting LibQUAL+™ Raw Scores for the *Total* Scale ($v = 25$) into Scale Percentile Ranks and T Scores ($n = 16{,}918$ *all ARL* Participants)

Raw Score	%ile	T	Raw Score	%ile	T	Raw Score	%ile	T
3.302	1	20.369	3.880	2	25.464	4.200	3	28.284
4.404	4	30.084	4.600	5	31.809	4.738	6	33.022
4.880	7	34.277	4.992	8	35.267	5.051	9	35.783
5.140	10	36.571	5.229	11	37.353	5.287	12	37.866
5.366	13	38.559	5.440	14	39.212	5.516	15	39.884
5.560	16	40.269	5.633	17	40.911	5.680	18	41.327
5.753	19	41.968	5.800	20	42.384	5.840	21	42.737
5.887	22	43.153	5.925	23	43.487	5.970	24	43.881
6.006	25	44.198	6.058	26	44.654	6.098	27	45.007
6.137	28	45.357	6.160	29	45.557	6.200	30	45.909
6.240	31	46.262	6.280	32	46.614	6.313	33	46.903
6.342	34	47.164	6.374	35	47.445	6.403	36	47.699
6.440	37	48.024	6.480	38	48.377	6.513	39	48.666
6.550	40	48.997	6.562	41	49.095	6.600	42	49.434
6.633	43	49.723	6.655	44	49.917	6.680	45	50.139
6.720	46	50.492	6.750	47	50.759	6.760	48	50.844
6.800	49	51.197	6.822	50	51.394	6.840	51	51.549
6.874	52	51.852	6.888	53	51.969	6.920	54	52.254
6.953	55	52.541	6.963	56	52.635	7.000	57	52.959
7.024	58	53.171	7.040	59	53.312	7.077	60	53.640
7.090	61	53.752	7.120	62	54.017	7.153	63	54.306
7.165	64	54.415	7.200	65	54.722	7.218	66	54.877
7.245	67	55.115	7.280	68	55.427	7.306	69	55.653
7.320	70	55.779	7.360	71	56.132	7.389	72	56.387
7.400	73	56.487	7.440	74	56.837	7.473	75	57.126
7.490	76	57.281	7.520	77	57.542	7.558	78	57.878
7.578	79	58.049	7.600	80	58.247	7.640	81	58.599
7.673	82	58.888	7.695	83	59.082	7.726	84	59.357
7.760	85	59.657	7.800	86	60.009	7.840	87	60.362
7.880	88	60.714	7.920	89	61.067	7.960	90	61.419
8.000	91	61.772	8.073	92	62.413	8.120	93	62.829
8.178	94	63.337	8.242	95	63.901	8.320	96	64.592
8.440	97	65.649	8.570	98	66.799	8.791	99	68.745

Note. In a normal distribution, the (1) mean equals the median and (2) 95% of scores are between the mean plus and minus two times the standard deviation. We can see that the data were not perfectly normally distributed, because the median or the 50th percentile (i.e., $T = 51.394$) did not exactly equal the T score mean of 50.0. Second, roughly 4% of the scores (rather than 2.5%) were below the T score of 50.0 - 2(10) or 30.0, and less than 1% of the scores (rather than 2.5%) were above the T score of 50.0 + 2(10) or 70.0.

Illustrative Use of Individual-level Norms

A concrete heuristic illustration of the application of the Table 1 norms may be instructive. Presume that at some future date a faculty member on an ARL campus provided answers to the twenty-five LibQUAL+™ total item set such that the user's mean rating on the twenty-five "perceived" items was 5.44. Absent norms, the score provides relatively little information.

Is 5.44 very good, good, or bad? The LibQUAL+™ items are scaled from "1" (low) to "9" (high). Thus, the scale midpoint is 5.0, and so 5.44 might be interpreted as moderately good, because 5.44 is greater than 5.0.

But would 5.44 be good (or moderately good) if 5.44 was the lowest score ever generated on the LibQUAL+™ items? Clearly, norms are useful in understanding the meaning of scores in context.

Because the hypothetical user is from an ARL campus, Table 1 might be used to convert the observed score (i.e., 5.44) into (a) a percentile rank and/or (b) a T score. The use of the norm conversion table presumes that (a) the present sample is reasonably representative of a meaningful population (e.g., library users at ARL institutions) and (b) the new user is reasonably similar to the participants in the present study who were used to create the normative tables.

In this case the *percentile rank* equivalent of a raw score of 5.44 is 14. This means that in the normative sample (i.e., the present study) this user had a LibQUAL+™ total score that was higher than only 14% of the 16,918 members of the normative sample.

The raw score of 5.44 can also be converted into a *T score*, using Table 1. Based on the data from the 16,918 ARL participants yielding the Table 1 conversion table, we learn that this faculty member's normative T score was 39.212. We now know that 5.44 was below the mean (i.e., 50.0) in the normative sample. Indeed, because the SD in the normative sample was 10.0, and 50.0 minus 39.212 is slightly more than 10.0, we also know that 5.44 was slightly more than one standard deviation below the mean in the normative sample.

The norm tables permitted much greater insight by putting a raw score into a normative context.

From both the percentile rank and the T score perspective, this faculty member had relatively unfavorable perceptions of the library. The norm tables permitted much greater insight by putting a raw score into a normative context. We are no longer reduced merely to interpreting the observed raw score of 5.44 against the scale midpoint (i.e., 5.0) on a "1" to "9" response scale.

User-group Norms

Alternatively, one might instead want to compare this raw score of 5.44 within the normative sample consisting only of ARL faculty. Norms can also be developed within discrete user groups. Tables 2 through 4 present norms for LibQUAL+™ total scores for (a) undergraduate students, (b) graduate students, and (c) faculty, respectively. According to Table 4, in this case the percentile rank of 5.44 would be slightly higher: 16.

It is noteworthy that the norms for the three user groups at the ARL institutions in the present sample were so similar. For example, the medians (or 50th percentiles) of the three user groups were similar (i.e., 6.880, 6.800, and 6.745 from Tables 2 thru 4,

Table 2

Norms for Converting LibQUAL+™ Raw Scores for the *Total* Scale ($v = 25$) into Scale Percentile Ranks and *T* Scores (n = 5,154 *undergraduate student ARL* Participants)

Raw Score	%ile	T	Raw Score	%ile	T	Raw Score	%ile	T
3.508	1	22.184	4.012	2	26.631	4.343	3	29.548
4.558	4	31.442	4.706	5	32.746	4.866	6	34.157
4.960	7	34.982	5.009	8	35.411	5.120	9	36.392
5.200	10	37.097	5.273	11	37.738	5.320	12	38.154
5.400	13	38.859	5.453	14	39.324	5.520	15	39.917
5.584	16	40.479	5.640	17	40.974	5.705	18	41.547
5.760	19	42.032	5.831	20	42.656	5.880	21	43.089
5.920	22	43.442	5.960	23	43.794	6.008	24	44.219
6.060	25	44.674	6.118	26	45.184	6.156	27	45.525
6.193	28	45.851	6.225	29	46.127	6.269	30	46.519
6.313	31	46.903	6.344	32	47.176	6.374	33	47.444
6.400	34	47.672	6.440	35	48.024	6.480	36	48.377
6.512	37	48.660	6.537	38	48.881	6.570	39	49.168
6.600	40	49.434	6.633	41	49.725	6.659	42	49.958
6.685	43	50.186	6.720	44	50.492	6.753	45	50.780
6.766	46	50.896	6.800	47	51.197	6.833	48	51.486
6.845	49	51.595	6.880	50	51.902	6.908	51	52.152
6.922	52	52.273	6.960	53	52.607	6.973	54	52.717
7.000	55	52.959	7.025	56	53.181	7.040	57	53.312
7.073	58	53.601	7.086	59	53.721	7.120	60	54.017
7.132	61	54.126	7.160	62	54.369	7.177	63	54.518
7.200	64	54.722	7.233	65	55.009	7.250	66	55.166

Raw Score	%ile	T	Raw Score	%ile	T	Raw Score	%ile	T
7.280	67	55.427	7.313	68	55.716	7.323	69	55.810
7.360	70	56.132	7.393	71	56.421	7.405	72	56.530
7.440	73	56.837	7.470	74	57.103	7.488	75	57.260
7.520	76	57.542	7.560	77	57.894	7.582	78	58.092
7.600	79	58.247	7.640	80	58.599	7.673	81	58.888
7.698	82	59.107	7.729	83	59.383	7.764	84	59.693
7.800	85	60.009	7.840	86	60.362	7.880	87	60.714
7.920	88	61.067	7.960	89	61.419	8.000	90	61.772
8.058	91	62.279	8.105	92	62.697	8.160	93	63.182
8.215	94	63.670	8.280	95	64.239	8.391	96	65.221
8.519	97	66.343	8.646	98	67.468	8.880	99	69.527

Table 3

Norms for Converting LibQUAL+™ Raw Scores for the *Total* Scale ($v = 25$) into Scale Percentile Ranks and *T* Scores ($n = 5{,}566$ *graduate student ARL* Participants)

Raw Score	%ile	T	Raw Score	%ile	T	Raw Score	%ile	T
3.160	1	19.119	3.840	2	25.111	4.120	3	27.579
4.360	4	29.694	4.594	5	31.752	4.736	6	33.008
4.880	7	34.277	5.000	8	35.334	5.073	9	35.977
5.140	10	36.568	5.228	11	37.341	5.303	12	38.001
5.400	13	38.859	5.457	14	39.362	5.524	15	39.949
5.600	16	40.622	5.640	17	40.974	5.720	18	41.679
5.775	19	42.160	5.840	20	42.737	5.880	21	43.089
5.920	22	43.442	5.960	23	43.794	6.000	24	44.147
6.040	25	44.499	6.080	26	44.852	6.113	27	45.142
6.156	28	45.523	6.191	29	45.833	6.218	30	46.065
6.240	31	46.262	6.280	32	46.614	6.320	33	46.967
6.355	34	47.279	6.387	35	47.559	6.420	36	47.850
6.458	37	48.179	6.480	38	48.377	6.520	39	48.729
6.560	40	49.082	6.573	41	49.193	6.600	42	49.434
6.633	43	49.723	6.655	44	49.917	6.680	45	50.139

Table 3 Cont.

Raw Score	%ile	T	Raw Score	%ile	T	Raw Score	%ile	T
6.720	46	50.492	6.743	47	50.698	6.760	48	50.844
6.791	49	51.116	6.800	50	51.197	6.839	51	51.536
6.850	52	51.641	6.880	53	51.902	6.906	54	52.127
6.927	55	52.320	6.960	56	52.607	6.991	57	52.883
7.000	58	52.959	7.040	59	53.312	7.048	60	53.382
7.080	61	53.664	7.109	62	53.924	7.131	63	54.117
7.160	64	54.369	7.200	65	54.722	7.217	66	54.869
7.240	67	55.074	7.273	68	55.363	7.280	69	55.427
7.320	70	55.779	7.330	71	55.871	7.360	72	56.132
7.400	73	56.484	7.417	74	56.632	7.440	75	56.837
7.480	76	57.189	7.498	77	57.344	7.520	78	57.542
7.560	79	57.894	7.600	80	58.247	7.610	81	58.338
7.640	82	58.599	7.680	83	58.952	7.713	84	59.238
7.738	85	59.459	7.778	86	59.812	7.810	87	60.101
7.859	88	60.527	7.898	89	60.869	7.946	90	61.297
8.000	91	61.772	8.040	92	62.124	8.120	93	62.829
8.160	94	63.182	8.240	95	63.887	8.320	96	64.592
8.440	97	65.649	8.560	98	66.707	8.715	99	68.074

Table 4

Norms for Converting LibQUAL+™ Raw Scores for the *Total* Scale ($v = 25$) into Scale Percentile Ranks and T Scores ($n = 4{,}420$ *faculty ARL* Participants)

Raw Score	%ile	T	Raw Score	%ile	T	Raw Score	%ile	T
3.280	1	20.176	3.675	2	23.654	4.000	3	26.521
4.240	4	28.636	4.400	5	30.046	4.540	6	31.277
4.664	7	2.376	4.777	8	33.366	4.908	9	34.519
4.995	10	35.290	5.073	11	35.976	5.154	12	36.692
5.233	13	37.390	5.302	14	37.996	5.361	15	38.519
5.455	16	39.345	5.520	17	39.917	5.560	18	40.269

Raw Score	%ile	T	Raw Score	%ile	T	Raw Score	%ile	T
5.623	19	40.826	5.680	20	41.327	5.720	21	41.679
5.760	22	42.032	5.818	23	42.539	5.872	24	43.016
5.898	25	43.244	5.921	26	43.448	5.960	27	43.794
6.000	28	44.147	6.040	29	44.499	6.080	30	44.852
6.120	31	45.204	6.160	32	45.557	6.200	33	45.909
6.218	34	46.064	6.250	35	46.353	6.280	36	46.614
6.320	37	46.967	6.360	38	47.319	6.400	39	47.672
6.434	40	47.973	6.452	41	48.128	6.480	42	48.377
6.520	43	48.729	6.560	44	49.082	6.600	45	49.434
6.632	46	49.719	6.655	47	49.916	6.680	48	50.139
6.717	49	50.462	6.745	50	50.716	6.760	51	50.844
6.800	52	51.197	6.840	53	51.549	6.859	54	51.714
6.880	55	51.902	6.909	56	52.161	6.930	57	52.346
6.960	58	52.607	6.991	59	52.884	7.000	60	52.959
7.040	61	53.312	7.068	62	53.562	7.086	63	53.719
7.120	64	54.017	7.154	65	54.320	7.165	66	54.411
7.200	67	54.722	7.211	68	54.814	7.251	69	55.173
7.280	70	55.427	7.320	71	55.779	7.352	72	56.065
7.379	73	56.296	7.400	74	56.484	7.440	75	56.837
7.473	76	57.126	7.496	77	57.330	7.520	78	57.542
7.555	79	57.854	7.576	80	58.036	7.600	81	58.247
7.640	82	58.599	7.673	83	58.886	7.680	84	58.954
7.720	85	59.304	7.760	86	59.657	7.800	87	60.009
7.833	88	60.298	7.868	89	60.608	7.913	90	61.003
7.960	91	61.419	8.000	92	61.772	8.078	93	62.462
8.120	94	62.829	8.200	95	63.534	8.275	96	64.193
8.360	97	64.944	8.520	98	66.354	8.695	99	67.895

respectively). The score ranges (roughly 3.508 to 8.880, 3.160 to 8.715, and 3.280 to 8.695 from Tables 2 thru 4, respectively) were comparable. And the scores at various percentile ranks (e.g., at the 25th percentile of 6.060, 6.040, and 5.898) were comparable.

These results suggest that the total scores were distributed in a strikingly similar way with regard to location (e.g., median), spread (e.g., range), and shape. Overall, the result suggests that, at least with regard to LibQUAL+™ total scores, a single aggregated norm conversion table (i.e., table 1) for all user groups may be sufficient.

At the level of the four LibQUAL+™ subscales, however, there may be noteworthy differences across user groups. For example, our qualitative data suggested that Library as Place may be particularly important to undergraduate users.[12] At the dimension level it may be desirable to pay close attention to nuances and to those aspects of service quality in which groups may be differentially vested.

Raw Score	%ile	T	Raw Score	%ile	T	Raw Score	%ile	T
5.623	19	40.826	5.680	20	41.327	5.720	21	41.679
5.760	22	42.032	5.818	23	42.539	5.872	24	43.016
5.898	25	43.244	5.921	26	43.448	5.960	27	43.794
6.000	28	44.147	6.040	29	44.499	6.080	30	44.852
6.120	31	45.204	6.160	32	45.557	6.200	33	45.909
6.218	34	46.064	6.250	35	46.353	6.280	36	46.614
6.320	37	46.967	6.360	38	47.319	6.400	39	47.672
6.434	40	47.973	6.452	41	48.128	6.480	42	48.377
6.520	43	48.729	6.560	44	49.082	6.600	45	49.434
6.632	46	49.719	6.655	47	49.916	6.680	48	50.139
6.717	49	50.462	6.745	50	50.716	6.760	51	50.844
6.800	52	51.197	6.840	53	51.549	6.859	54	51.714
6.880	55	51.902	6.909	56	52.161	6.930	57	52.346
6.960	58	52.607	6.991	59	52.884	7.000	60	52.959
7.040	61	53.312	7.068	62	53.562	7.086	63	53.719
7.120	64	54.017	7.154	65	54.320	7.165	66	54.411
7.200	67	54.722	7.211	68	54.814	7.251	69	55.173
7.280	70	55.427	7.320	71	55.779	7.352	72	56.065
7.379	73	56.296	7.400	74	56.484	7.440	75	56.837
7.473	76	57.126	7.496	77	57.330	7.520	78	57.542
7.555	79	57.854	7.576	80	58.036	7.600	81	58.247
7.640	82	58.599	7.673	83	58.886	7.680	84	58.954
7.720	85	59.304	7.760	86	59.657	7.800	87	60.009
7.833	88	60.298	7.868	89	60.608	7.913	90	61.003
7.960	91	61.419	8.000	92	61.772	8.078	93	62.462
8.120	94	62.829	8.200	95	63.534	8.275	96	64.193
8.360	97	64.944	8.520	98	66.354	8.695	99	67.895

respectively). The score ranges (roughly 3.508 to 8.880, 3.160 to 8.715, and 3.280 to 8.695 from Tables 2 thru 4, respectively) were comparable. And the scores at various percentile ranks (e.g., at the 25th percentile of 6.060, 6.040, and 5.898) were comparable.

These results suggest that the total scores were distributed in a strikingly similar way with regard to location (e.g., median), spread (e.g., range), and shape. Overall, the result suggests that, at least with regard to LibQUAL+™ total scores, a single aggregated norm conversion table (i.e., table 1) for all user groups may be sufficient.

At the level of the four LibQUAL+™ subscales, however, there may be noteworthy differences across user groups. For example, our qualitative data suggested that Library as Place may be particularly important to undergraduate users.[12] At the dimension level it may be desirable to pay close attention to nuances and to those aspects of service quality in which groups may be differentially vested.

Table 5

Institutional Norms Based on Faculty and Graduate and Undergraduate Student Responses ($n = 18{,}165$) at 43 Universities on the LibQUAL+™ Total **Perceptions** Scale (25 items)

Raw Score	%ile	T
6.176	10.00	30.595
6.455	20.00	42.479
6.502	25.00	44.491
6.564	30.00	47.129
6.665	40.00	51.007
6.687	45.00	52.344
6.699	50.00	52.859
6.717	55.00	53.611
6.729	60.00	54.127
6.751	70.00	55.098
6.773	75.00	55.998
6.812	80.00	57.672
6.866	90.00	59.989

of users. The purpose of LibQUAL+™ is to help libraries realize these potentials by identifying specific areas for needed improvement and/or specific areas of needed additional service quality information (e.g., focus groups).

Discussion

Norms are helpful, and indeed may even be essential, in interpreting LibQUAL+™ results. Norms can be used to make comparative statements about performance from either an individual or an institutional perspective. When deficiencies are noted, plans for improvement must be formulated, or additional diagnostic information must be sought through other ways of listening to users.[15]

One powerful application of LibQUAL+™ would involve a library performing at less than desired levels on the total scale, or on one of the four subscales (i.e., Service Affect, Library as Place, Personal Control, and Information Access), then isolating a peer institution performing better in the target area. The library in question might then improve service by using that peer institution as a model. The current LibQUAL+™ protocol provides library administrators with what the service quality literature calls a

Table 6

Institutional Norms Based on Faculty and Graduate and Undergraduate Student Responses ($n = 9{,}181$) at 43 Universities for LibQUAL+™ Total **Service Adequacy Gap** Scores (25 items)

Raw Score	%ile	T
-0.080	10.00	35.216
0.187	20.00	43.766
0.239	25.00	45.417
0.266	30.00	46.284
0.327	40.00	48.236
0.340	45.00	48.682
0.371	50.00	49.652
0.412	55.00	50.990
0.421	60.00	51.269
0.563	70.00	55.817
0.623	75.00	57.726
0.681	80.00	59.606
0.743	90.00	61.575

Note. "Adequacy" is a cumulative function of "Perceived" - "Minimum" service scores; higher scores are more desirable.

"total market survey" involving twenty-five questions, each of sufficient granularity to be useful in local resource allocation decisions. These twenty-five questions and the four dimensions they delineate characterize the overarching concept of service quality delivery measured by the total LibQUAL+™ score.

While a negative gap score is a source of concern, this normative information may help a library director place a negative score in a national context.

Now information at the item, subscale, and total score level can be placed in context using norms for individuals or institutions. An example at the question level may be illuminating. One LibQUAL+™ question asks users to rate the quality of runs of journal titles. Among the 9,293 respondents who were randomly assigned all three response frameworks (i.e., minimum, perceived, desired), this item had a negative service adequacy gap score (M = -.21), thus indicating that on the average users perceive journal runs to be below what they deem minimally acceptable.

Table 7

Institutional Norms Based on Faculty and Graduate and Undergraduate Student Responses ($n = 9{,}181$) at 43 Universities for LibQUAL+™ Total **Service Superiority Gap** Scores (25 items)

Raw Score	%ile	T
-1.831	10.00	32.665
-1.498	25.00	44.845
-1.379	30.00	47.166
-1.337	40.00	48.502
-1.316	45.00	49.200
-1.273	50.00	50.566
-1.206	55.00	52.701
-1.164	60.00	54.075
-1.105	70.00	55.938
-1.087	75.00	56.524
-1.057	80.00	57.507
-0.942	90.00	61.184

Note. "Superiority" is a cumulative function of "Perceived" - "Desired" service scores. Higher (positive or at least less negative) scores are more desirable, although a large positive value may reflect undue investment of resources in pursuing a service objective.

While a negative gap score is a source of concern, this normative information may help a library director place a negative score in a national context. On this particular item a negative mean gap score could still place a given institution's performance above the national mean. At each level—item, subscale, total score—normative information enables library personnel to better understand local performance issues, to place them in the context of peer institutions, and to seek best practices that can improve local service delivery.

Colleen Cook is Professor of Library Science and Executive Associate Dean of the Sterling C. Evans Library at Texas A&M University; she may be contacted via e-mail at: ccook@tamu.edu.

Fred Heath is Dean of Library Services and holder of the Sterling Evans Chair at Texas A&M University; he may be contacted via e-mail at: fheath@tamu.edu.

Bruce Thompson is Professor of Educational Psychology and Distinguished Research Scholar, Texas A&M University, and Adjunct Professor of Community Medicine, Baylor College of Medicine; he may be contacted via Internet URL: <http://www.coe.tamu.edu/~bthompson>.

Notes

1. This research was supported in part by a grant from the U.S. Department of Education Fund for the Improvement of Postsecondary Education (FIPSE). Colleen Cook may be contacted via e-mail at Education (FIPSE).
2. D. A. Nitecki, "Changing the Concept and Measure of Service Quality in Academic Libraries," *The Journal of Academic Librarianship* 22 (1996): 181–190.
3. R. Cullen, "Perspectives on User Satisfaction Surveys," *Library Trends* 49, 4 (2001): 662–686.
4. Colleen Cook and Fred Heath, "The Association of Research Libraries LibQUAL+™ Project: An Update," *ARL Newsletter: A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC* 211 (2000): 12–14; Cook, Heath, and Bruce Thompson, "LibQUAL+™: One Instrument in the New Measures Toolbox," *ARL Newsletter* 212 (October 2000): 4–7; Cook, Heath, and B. Thompson, "Users' Hierarchical Perspectives on Library Service Quality: A 'LibQUAL+™' Study," *College & Research Libraries* 62 (2001): 147–153; Cook and others, "The Search for New Measures: The ARL 'LibQUAL+™' Study—A Preliminary Report," *portal: Libraries and the Academy* 1, 1 (2001): 103–112.
5. R. C. Arnau, R. L. Thompson, and C. Cook, "Do Different Response Formats Change the Latent Structure of Responses? An Empirical Investigation Using Taxonometric Analysis," *Educational and Psychological Measurement* 61 (2001): 23–44.
6. B. Thompson, C. Cook, and R. L. Thompson, "Reliability and Structure of LibQUAL+™ Scores: Measuring Perceived Library Service Quality," *portal: Libraries and the Academy* (THIS ISSUE).
7. C. Cook and B. Thompson, "Psychometric Properties of Scores From the Web-Based LibQUAL+™ Study of Perceptions of Library Service Quality, *Library Trends* 49, 4 (2001): 585–604.
8. J. R. Sanders and the Joint Committee on Standards for Educational Evaluation, *The Program Evaluation Standards*, 2nd ed. (Thousand Oaks, CA: Sage, 1994), 207.
9. A. Anastasi, *Psychological Testing*, 6th ed. (New York: Macmillan, 1988), 71–72.
10. L. Crocker and J. Algina, *Introduction to Classical and Modern Test Theory* (New York: Holt, Rinehart and Winston, 1986), 431.
11. B. Thompson, "GRE Percentile Ranks Cannot Be Added or Averaged: A Position Paper Exploring the Scaling Characteristics of Percentile Ranks, and the Ethical and Legal Culpabilities Created By Adding Percentile Ranks in Making 'High-Stakes' Testing Decisions" (paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans, 1993).