2014

The Suitability of Web Analytics Key Performance Indicators in the Academic Library Environment

Jody C. Fagan
James Madison University, faganjc@jmu.edu

Follow this and additional works at: http://commons.lib.jmu.edu/letfspubs

Part of the Library and Information Science Commons

Recommended Citation

The Dimensions of Library Service Quality:

A Confirmatory Factor Analysis of the LibQUAL+ Instrument

Jody Condit Fagan

Carrier Library MSC 1704

James Madison University

Harrisonburg, VA 22807

540-568-4265

faganjc@jmu.edu[1]

---

[1] Correspondence may be addressed to Jody Condit Fagan, faganjc@jmu.edu, Carrier Library MSC 1704, James Madison University, Harrisonburg, VA 22807, 540-568-4265

1. Introduction

Academic libraries contribute to their universities by acquiring, organizing, and disseminating information; providing space for research activities; and supporting users in finding and using information. Libraries are generally customer-centered and engage in numerous evaluation activities. For many years, the lion's share of library effectiveness studies focused on specific programs or services at single institutions. In the past ten years, interest has surged in assessing the entire library unit: Searching the Library, Information Science & Technology Abstracts database for "assessment or evaluation or 'user study'" produced 5,401 results for the decade beginning in 1990 and 17,598 results for the decade beginning in 2000.

This surge of interest in assessment has been encouraged by the larger trend in higher education for more measurement and accountability (U.S. Department of Education, 2006). Academic units, including libraries, feel greater pressure to demonstrate their contribution to the academic enterprise in quantitative ways. The LibQUAL+ survey was designed by the Association of Research Libraries (ARL) and library partners to measure perceptions of library service quality (Saunders, 2007). One of the benefits of LibQUAL+ is the standardization of a measure across libraries, facilitating peer comparison and large-group studies. LibQUAL+ has been used at more than 1,164 institutions around the world and administered to over 1.2 million subjects (Lane, Anderson, Ponce, & Natesan, 2012, p. 22). ARL has access to LibQUAL+ data across its users, which supports their ongoing efforts to improve the instrument. The common ground that libraries have had to converse about service quality has supported collaborative developments in response to shared trends and challenges (e.g., Cook, 2002; Heath, Kyrillidou, & Askew, 2004).

2. Problem statement

Research involving LibQUAL+ has focused on the global fit of the LibQUAL+ model and largely overlooked the areas of local misfit in the model. Most studies of LibQUAL+ psychometric properties have correlated error residuals before proposing a final model without explaining the theoretical justification for so doing. In addition to the methodological objections to post-hoc correlation of error residuals (MacCallum, Roznowski, & Necowitz, 1992), researchers lose valuable information when correlating error terms. This study tests the factor structure of LibQUAL+ with one population of undergraduates and demonstrates how investigating areas of local misfit, using residuals and factor structure and pattern coefficients, can reveal areas for further investigation. If local misfit or weak discriminant validity is repeated across institutions it points to a potential need for instrument development; if local misfit or discriminant validity is specific to a population it could make libraries' LibQUAL+ results more locally meaningful – for example, by suggesting areas where programmatic changes could have the most effect on improving library service quality.

The purpose of this study was to assess the structure of LibQUAL+ perceived service performance scores using data from undergraduate students, and to carefully examine any areas of local misfit. Confirmatory factor analyses produce residuals for each measured item, indicating specific areas where the gathered data does not fit the model. Thus, residuals can reveal problems with how constructs are defined, or can suggest the presence of additional, unmeasured constructs or systematic group differences that may affect a specific population's response. Until Keiftenbeld and Natesan's (2013) measurement invariance study, studies of the current LibQUAL+ instrument tended to pool respondents across distinct groups (e.g., faculty and students). For this reason, the present study focused on the perceived service performance scores of one user group—undergraduates—in order to control for any confounding due to group

differences in responding to LibQUAL+. No one has yet isolated a particular group assessing the factor structure of the current LibQUAL+ instrument.[2] Based on suggestions from Lane et al. (2012), three models were tested: the three-factor model of Affect of Service, Information Control, and Library as Place; a two-factor model of Library as Place and a second factor composed of Affect of Service and Information Control; and a one-factor model combining all of these.

3. LibQUAL+

3.1. Model of library service quality

[Insert Figure 1]

ARL's (2012a) explanation of what LibQUAL+ intends to measure focuses on quality of service: "Service quality is the customers' assessment of how good/bad, or pleasant/unpleasant their experiences are. ... The LibQUAL+ instrument service quality 'measurements' are snapshots or discrete summaries of customers' evaluation of their experiences" ("What is the Difference Between", para. 1)). A concurrent validity study of the current LibQUAL+ instrument by Thompson, Cook, and Kyrillidou (2005) found LibQUAL+ scores relate more to satisfaction than academic outcomes, although there may still be a distal relationship between library service quality and academic outcomes (p. 521).

---

[2] Thompson, Cook, & Heath (2003) fit a post-hoc revised model (with five correlated error terms) to separate groups of undergraduates, graduate students, and faculty, and demonstrated similar fit across the groups. Unfortunately, this tells us little about the instrument's factorial invariance; the factor structure for the groups was not discussed separately, nor were constraints applied to test for metric or scalar invariance (Bontempo & Hofer, 2007).

The current model of user perceptions of library service quality, upon which LibQUAL+ is based, consists of three dimensions: Affect of Service, Information Control, and Library as Place (ARL, 2012a). Assessing the dimensionality of LibQUAL+ scores to see if they align with this structure is important because scores on the three dimensions are used in longitudinal studies (e.g., Greenwood, Watson, & Dennis, 2011) and to support peer comparisons (ARL, 2011, p. 48). In addition, analyzing the areas where data does not fit the model (i.e., residuals) is important for determining whether error is related to the measurement instrument or to variations between sample data sets. For example, Horst, Finney, and Barron (2006) diagnosed model misfit by examining standardized covariance residuals and detected similar patterns among four independent samples representing two different populations of students, indicating evidence that items were not functioning well. The full list of the 22 core items[3] can be found in order in Figure 1. It is important both to define the constructs and test the ability of the instrument to discriminate between those constructs in order for results to have interpretable meaning.

3.2. Affect of service

Affect of Service concerns "the human dimension of service quality" (ARL, 2012b) and is operationalized with nine questions about user interactions with staff. Aspects of this dimension include user perceptions of staff helpfulness, competency, dependability, and care for

---

[3] The LibQUAL+ instrument has 22 items intended to represent the underlying factors of service quality. When libraries distribute LibQUAL+, they may add five additional items from a fixed pool (Thompson, Cook, & Kyrillidou, 2006), which are integrated with the 22 core items into one survey section. Appended to this in a separate survey section are eight additional items, which address information literacy/outcomes and general satisfaction; these have been used, among other things, to try to support concurrent validity (Thompson, Cook, & Kyrillidou, 2005). Demographics questions and a comment box make a total of 44 questions on the distributed survey. Only the 22 core items were included in this study.

users. Today, many types of staff provide library customer service; however, for many years, the performance of reference librarians was the focus of research. Rimland (2007) reviewed research about the evaluation of reference effectiveness from 1994 to 2006 and found a lack of consistency in the field with respect to selecting variables, operationalizing their measurement, and reporting research findings. She found that most studies have been qualitative; instruments have not generally been evaluated for psychometric integrity; most study participants have been drawn from a convenience sample; and usually faculty and students are grouped together in the same study. Even though there have been some methodological problems, it is fruitful to uncover what library practitioners have considered relevant to evaluating this aspect of service.

Initially, many libraries measured reference service quality in terms of whether questions were answered correctly. A landmark study found the accuracy rate at reference desks to be only 55% (Hernon & McClure, 1986). The Wisconsin-Ohio Reference Evaluation Program (WOREP) instrument provided a different, more multidimensional view of a reference interaction and was adopted at many libraries (Murfin & Gugelchuk, 1987). The WOREP was noteworthy in that it separated satisfaction from "correctness," although it was criticized for not measuring the instructional qualities of reference interactions.

In the mid-1980s, there was increased interest in behavioral aspects of service and in patron satisfaction. Schwartz and Eakin (1986) conceptualized good reference service as attitude and demeanor; interviewing, listening, and referring; search strategy skills; and knowledge of library collections. Durrance (1989, 1995) included interpersonal skills; interviewing and listening skills; and the effectiveness of approach, including search strategy, accuracy, and ability to provide the correct answer on her willingness to return instrument. Recent research has suggested perceptions of customer service quality vary depending on staff type differences (Vilelle & Peters, 2008) and between online and in-person environments (Desai & Graves, 2008;

Nilsen & Ross, 2006). These studies suggest different behaviors may contribute to the patron's perception of the helpfulness, competency, dependability, and care for users that compose the Affect of Service construct.

3.3. Information control

The developers of LibQUAL+ have defined Information Control as "whether users are able to find the required information in the library in the format of their choosing, in an independent and autonomous way" (ARL, 2012b). The eight questions created to represent this construct involve having the right print and electronic materials in the collections, being able to access desired resources independently, and the extent to which access tools are modern and intuitive.

Having the right collections appears to be central to student perceptions of library service quality. Part of a nationwide survey by the Online Computer Library Center (OCLC, Inc., 2010) studied student perceptions of library resources and their access to them. When asked what the most important role of the library was for them, college students' top response (35%) was "to provide books, videos, and music" (p. 59). These findings support the idea that having relevant collections is important to overall perceptions of library service quality.

The ability for users to find information independently has been growing since the mid-1980s, when libraries began experimenting with end-user searching. Early research (Friend, 1985) questioned "whether individuals would really want to do their own searches" (p. 136) but discovered users were "anxious to enter as active participants" (p. 141). With the advent of the Web, users' expectations of self-sufficiency in retrieving information skyrocketed and libraries responded by commissioning numerous usability tests, user surveys, and focus groups to improve the user's ability to find information unaided (Blummer, 2007; Freund, Nemmers, & Ochoa, 2007; Letnikova, 2003). These investigations usually concerned specific systems such as

the library catalog or the library web site, not holistic investigations of independence in using the library.

The third component of Information Control is the idea of modern and intuitive access tools, including the LibQUAL+ item "A library Web site enabling me to locate information on my own." Results from a national OCLC survey (2010) showed more than half of 18-24-year-old college students used the library web site, with almost 40% indicating they used electronic journals and 30% online databases (p. 55). These figures suggest potential room for libraries to improve their web sites, as well as access to and promotion of electronic resources. This aspect of information control could use further research. Some librarians may think of the library web site as the primary portal to library information; yet, many users access electronic resources via search engines, personal bookmarks, or journal publisher web sites.   Thus, the library web site might not deserve more focus than other ways of accessing online content. Additionally, although the Information Control items appear relevant to library service quality, they may be more different from one another than the other factors' items.

3.4. Library as place

Finally, Library as Place is defined as "the physical environment of the library as a place for individual study, group work, and inspiration" (ARL, 2012b). The five LibQUAL+ questions assess the availability of both quiet and community space, the comfort and welcoming feel of space, and the suitability of space for study, learning, and research. Many academic libraries have been changing their spaces radically to keep pace with the rapid increase in online information, including creating "learning commons" spaces to support document and media production (Accardi, Cordova, & Leeder, 2010; Seeholzer & Salem, 2010) and adding technology, group study spaces, and coffee shops.

New and renovated buildings have been shown to correlate with higher building usage (Shill & Tonner, 2004). However, in a survey of 182 libraries with major space improvement projects, Shill and Tonner (2004) found only certain improvements were associated with usage gains, suggesting not all improvements to library space are equal.

The need for both quiet and group study space was reflected in survey results from Vaska, Chan, and Powelson (2009, p. 224) with 170 faculty and student users. The respondents' favorite study area in the library was the study carrels, followed by group spaces, suggesting a multidimensional aspect to space preferences. On two separate questions about desired overall changes and changes in technology, the top response was more electrical outlets (Vaska, Chan, & Powelson, 2009, p. 227, 229), suggesting the considerations underlying users' satisfaction with space may be very practical.

Students are still interested in using library buildings. At James Madison University (JMU), both major libraries are regularly crowded. In a national study, when students who did not use the library web site were asked why, 21% said they preferred to use the physical library (OCLC, 2010, p. 55). Also, when asked what the most important role of the library was to them, the second-most popular response was a "place to learn," shared by 32% of respondents. Finally, when reporting on library activities overall, 66% of students reported using the library as a place to do homework or to study (OCLC, 2010, p. 59). Similar to Affect of Service, the Library as Place items appear to represent the breadth of this aspect of library service quality.

Because this study investigates residuals, which can illuminate differences particular to a sample, characteristics of the JMU environment are relevant. JMU is a residential, student-centered school of just over 18,000 undergraduates (JMU, 2013a; JMU, 2013b); it is classed as a "Master's Colleges and University" by the Carnegie Foundation (2013); and boasts a retention rate of 91.4% and a student-faculty ratio of 16:1 (JMU, 2013a). Although JMU does have several

graduate programs of distinction, its focus on undergraduates can be seen through extensive

orientation programs, student services, extracurricular activities for student engagement, and

attention to general education (JMU, 2013c). In student reviews of JMU, the "community" and

"people" of JMU are frequently mentioned (Unigo, 2013).

4. Literature review

The LibQUAL+ instrument was partially inspired by the SERVQUAL instrument, which

aimed to measure five dimensions of service quality (Tangibles, Reliability, Responsiveness,

Assurance, and Empathy) using the Gap Theory of Service Quality (Parasuraman, Berry, &

Zeithaml, 1991). The purpose of both SERVQUAL and LIBQUAL+ was to identify gaps in

service quality by asking respondents to rate (a) the minimum level of service they are willing to

accept, (b) their desired level of service, and (c) their perception of actual service rendered for

each item (Cook, Heath, & Thompson, 2002). Some libraries have administered SERVQUAL,

but the instrument did not yield the theoretical five dimensions (Thompson, Cook, & Heath,

2001). Additionally, researchers suspected the SERVQUAL might be limited to use in the for-

profit environment (Lane et al., 2012, p. 23; Thompson et al., 2001). Thus, in October 2000,

ARL held a two-day conference to begin revising SERVQUAL for academic libraries and

conducted pilot studies with 13 institutions (Cook, Heath, & Thompson, 2002). In 2001, they

conducted an additional study with 43 institutions (Thompson, Cook, & Thompson, 2002). In

2002, 25 questions spanning four dimensions of service quality (Affect of Service, Personal

Control, Access to Information, and Library as Place) were piloted with 146 institutions

(Thompson, Cook, & Heath, 2003). Model-data fit was tested using confirmatory factor analysis

(CFA); however, the researchers used modification indexes to identify correlated error terms

when the original model did not fit particularly well (Thompson, et al., 2003). Although there

may have been theoretical justification to support these modifications, it was not explained in the

study. MacCallum, et al. (1992) have shown that empirically-based modification-of-fit indices capitalize on chance, may not replicate, and require theoretical justification. In addition, analyzing and discussing error terms can provide useful information.

Because of the high correlation between two of the original dimensions (r = .75) (Thompson et al., 2003), the developers combined Personal Control and Information Access into Information Control and deleted three items (Lane et al., 2012, p. 28), resulting in a 22-item instrument. The current three-factor design is purported to represent Affect of Service (9 items), Information Control (8 items), and Library as Place (5 items) (Thompson, Kyrillidou, & Cook, 2008), still asking participants to rate these items three times to represent their minimum, desired, and perceived service performance levels (see Appendix A).

Little research has been conducted on the validity of the "gap scores" generated by the difference between the minimum and desired service levels. One study found the perceived service performance scores to be better predictions of satisfaction with library services than gap scores (Roszkowski, Baky, & Jones, 2005).[4] Until more research is conducted related to gap scores, when one is interested in perceptions of library service quality, it may be prudent to use "perceived" scores. Recent studies of LibQUAL+ have used only perceived scores (e.g., Lane et al., 2012; Kieftenbeld & Natesan, 2013); therefore, this study also focused on just the perceived service performance scores.

To date, reliability of LibQUAL+ scores has been measured using only Cronbach's alpha (Lane et al., 2012, p. 27; Kieftenbeld & Natesan, 2013, p. 145; Thompson, Kyrillidou, and Cook,

---

[4] Roszkowski, Baky, and Jones (2005) discuss the theoretical difference between service quality and satisfaction, and framed their discussion in terms of satisfaction, which is why that term is used here instead of "user perceptions of library service quality."

2007, p. 42). Reliabilities computed for scores for the three subscales have been generally quite good (i.e., above .85); Kieftenbeld and Natesan (2013) reported reliabilities for undergraduates of .93 for Affect of Service, .90 for Information Control, and .87 for Library as Place (p. 145). Potential problems with using Cronbach's alpha to calculate reliability include the inherent assumptions of unidimensionality and uncorrelated errors (Yang & Green, 2011). Lane et al. (2012) noted their reliabilities would be affected by the errors they specified to correlate. Therefore, this study used McDonald's (1999, p. 205-206) omega for calculating reliability.

Lane et al. (2012) collected LibQUAL+ data in 2005, 2007, and 2009 at a large public university and conducted a multigroup CFA where year was the group; as with Hunter and Perret's (2011) study, the data were not separated into student and faculty groups. The researchers concluded the data from each year "adequately fit" the three-factor model, but chose to add four error covariances among items representing Affect of Service and Library as Place to their model to improve the fit. Unfortunately, these error covariances were not examined in detail or discussed with respect to theory. The multigroup analysis supported global model fit using multiple fit indices on the revised model. Correlations between Affect of Service and Information Control were high across the three samples, ranging from .83 - .86, and structure coefficients (i.e., the total relationship between an item and a factor) were sufficiently large ($r_s >$ .60) to prompt the researchers to consider collapsing Affect of Service and Information Control. However, after testing a two-factor model, they retained the three-factor model with correlated errors.

In 2013, Kieftenbeld and Natesan conducted an investigation into the measurement and structural invariance of LibQUAL+ across user groups. They found the relationships between items and constructs were equivalent across their samples of undergraduate students, graduate students, and faculty. Prior to their work, investigations of differences between students and

faculty examined differences in mean scores without demonstrating measurement invariance (Lippincott & Kyrillidou, 2004; Thompson et al., 2005; Thompson, Kyrillidou, & Cook, 2008; Wei, Thompson, & Cook, 2005). Without establishing measurement invariance, across-group comparisons of subscale means may be confounded by measurement bias (Bontempo & Hofer, 2007). Although it would be impossible to conduct an invariance study for every possible definition of groups, it is important to do so for groups with logical differences if the comparison is the target of a research question of interest. Faculty and students are often thought of as logically different groups. For example, Kayongo and Jones (2008) isolated faculty LibQUAL+ Information Control scores to increase awareness of library services to faculty. Additionally, research into student and faculty perceptions of library resources and the library building has shown these groups differ on what drives satisfaction (e.g., Antell & Engel, 2006; Ithaka, 2010; OCLC, 2010). Therefore, it is important to know whether LibQUAL+ can discriminate between actual differences in the groups or if the instrument measures constructs differently for each group. By demonstrating full measurement invariance in one sample, Keiftenbeld and Natesan (2013) laid a foundation that, if replicated, could greatly improve researcher's abilities to compare LibQUAL+ data for faculty and student user groups.

5. Participants and Procedures

[INSERT TABLE 1]

Sample data came from James Madison University's administration of the LibQUAL+ instrument in spring 2011 via a web survey sent to all students, faculty, and staff. The population of interest was undergraduates with informed opinions about library services. That is, undergraduates who did not answer the 22 core items, who answered "NA" for an item, or who

answered more than two items illogically[5] were removed, because they did not provide informed or valid responses regarding library service quality (see Table 1). For example, cases with the "NA" option were removed because "NA" overrode all three ratings (minimum, desired, and perceived performance service levels) for each item, indicating the respondent had insufficient information about or experience with that aspect of library service; respondents who answered illogically once or twice were left in to accommodate simple mistakes; but those who answered illogically more than two times were assumed to misunderstand or be inattentive to the survey.

The sample data demographics were similar in proportion to JMU's undergraduate student population. Sixty-three percent of the sample was female, which is just 3.5% more than in the student population (JMU, 2011). With respect to academic discipline, students could select only broad categories (e.g., "Social sciences/psychology"), making comparisons difficult; however, mapping these broad categories to JMU colleges suggested a generally proportionate sample in terms of academic discipline.

The data were screened for outliers. There were no univariate outliers, however four cases were flagged as potential multivariate outliers (detected using Mahalanobis distance) and removed: One case had taken over 20 hours to submit the survey; the other three had response patterns suggesting lack of attention to the questions. Thus, the final sample for this study was composed of 1,096 undergraduate students.

The sufficient sample size necessary for a structural equation modeling (SEM) study is determined by a combination of factors. Weston and Gore (2006, p. 734) noted that low-

---

[5] An illogical response was one where a respondent rated the "desired" level higher than the "minimum" score. This showed the respondent did not understand the rating system or was not putting much thought into his/her answers.

reliability measures, more complex models, or both, require a higher sample size. Bentler and Chou (1987, p. 91) proposed a 5:1 ratio of subjects to parameters when there are many indicators of latent variables and factor loadings are large. Because LibQUAL+ has been demonstrated to have good reliability, the models in this study were simple. Also, there have been moderate-to-large factor loadings in previous studies, thus the number of subjects was deemed sufficient.

Structural equation modeling was used because it permits estimation of measurement error separate from estimation of parameters among latent constructs and their indicators (Maruyama, 1998). CFA, rather than exploratory factor analysis, was appropriate because considerable effort was spent developing the instrument and its factor structure. Nested CFA models could be easily compared to further inform the use and interpretation of LibQUAL+ scores.

6. Results

[INSERT TABLE 2]

Descriptive statistics and item correlations are presented in Table 2. All items were moderately and positively correlated with one another, which was a good sign they shared variance due to one or more underlying constructs. Multicollinearity was assessed using the regression procedure in SPSS 20 to produce the tolerance level of each variable. Because all tolerance values were larger than .10, multicollinearity of the 22 items did not prohibit further analysis. That is, the items appeared as if they were not redundant and provided unique information (Weston & Gore, 2006, p. 735). The correlation matrix showed higher correlations among indicators representing the same factor; therefore, the ability of the three-factor model to reproduce these relationships was promising.

LibQUAL+ has nine possible response options per item (1 = low; 9 = high). For items 17 and 18, responses ranged from 3 to 9, and for items 19 and 20, responses ranged from 2 to 9.

Other items had just a handful of students selecting the lowest response categories. Although all items had a good number of cases for at least five of the nine response options, the limited number of cases with responses at the ends of the scale made the instrument look more like a five-category scale than a nine-category scale. Research suggests scales with more than four choices per indicator can be treated as continuous, especially if maximum-likelihood estimation is used with the Satorra-Bentler scaling method (Finney & DiStefano, 2013, p. 45). Therefore, the items in this study were evaluated as if they were continuous. However, future researchers should examine the range of LibQUAL+ responses in a particular sample before assuming the nine response levels ensure the data can be treated as continuous.

Studies have suggested large values of skewness or kurtosis can bias standard errors, parameter estimates, and fit indexes; thus, the data was examined for skewness greater than |2| and kurtosis greater than |7| (West, Finch, & Curran, 1995). All items were negatively skewed, and all but one item had positive kurtosis (see Table 2). The largest values found were |1.163| for skewness and |2.430| for kurtosis. However, the normalized Mardia's coefficient for multivariate kurtosis was 116.4, much greater than the critical value of 20 suggested by Bentler (1998) for data to be considered normal.

To guard against effects of non-normality, maximum-likelihood (ML) estimation with the Satorra-Bentler adjustment (S-B) to standard errors and fit indexes was used to estimate the parameters and fit between the observed and model-implied covariance matrices. ML has been shown to be less sensitive to sample size and kurtosis than generalized least squares and asymptotically distribution free methods: further, it produces unbiased estimates given correct model specification, and demonstrates higher accuracy in terms of fit indexes (Finney & DiStefano, 2013; Olsson, Foss, Troye, & Howell, 2000). Another option could have been using weighted-least-squares with a robust asymptotic covariance matrix; however, this method was

less familiar to the researchers, may still be affected by non-normality, and introduces additional complexity without an obvious benefit over robust ML (Flora & Curran, 2004; Yang-Wallentin, Jöreskog, & Luo, 2010). PRELIS 2.80 was used to prepare the covariance matrix (Jöreskog & Sörbom, 2007b).; LISREL 8.80 was used for the CFA procedures (Jöreskog & Sörbom, 2007a).

Because linearity is an underlying assumption of standard SEM techniques (Quintana & Maxwell, 1999, p. 497), the linear relationships among the 22 items were confirmed by examining bivariate scatterplots. Heteroscedasticity was examined by plotting standardized residuals against predicted values for each variable. Variability in scores was consistent for the values with more than a handful of responses; the restriction of range meant the highest and lowest predicted values had much less variability in residuals than values near the center. While heteroscedasticity makes it more difficult to predict variables from each other because the relationship is not consistent across all values, the real problem appeared to be the few responses using the ends of the scale.

Model fit was examined using x^2; the absolute fit indices root-mean-square-error-of-approximation (RMSEA) and standardized root-mean-square residual (SRMR); and the comparative fit index (CFI), an incremental index. The x^2 tests a model's exact fit to the data and can be affected by sample size (Kelloway, 1995), thus additional fit indexes should be used to provide additional information about the degree of fit or misfit. The RMSEA estimates the proportion of specification error per degree of freedom, thus addressing the parsimony of the model (Raykov & Marcoulides, 2000). The SRMR provides a standardized summary of the average covariance residuals. The CFI provides a measure of incremental fit by comparing the fit of the hypothesized model against an independence model where all indicators are uncorrelated. These model-data fit indices were chosen because they are sensitive to both misspecified factor correlations and misspecified factor loadings (Hu & Bentler 1998, 1999). To test for significant

difference between nested models, the S-B x^2 difference test was conducted using a critical value of $p < .05$ (Bryant & Satorra, 2012); however, the other fit indexes and the correlation residuals were also examined to evaluate the degree of difference in fit between models.

Researchers have offered a variety of approaches for interpreting fit indexes. Browne and Cudeck (1993) suggested RMSEA cutoff values of .05 or less indicate close fit, whereas Hu and Bentler (1999) suggested RMSEA values of .06 or less. For the CFI, Hu and Bentler (1999) recommended values of .95 or above. Studies by Hu and Bentler (1998, 1999) also suggested using an SRMR cutoff value of .08 or less. Marsh, Hau, and Wen (2004) suggested these guidelines were too extreme and would over-reject acceptable models. For this reason, the following values were used as rough guidelines for close model-data fit: RMSEA $\leq .06$, CFI $\geq$ .95, SRMR $\leq .08$. Of note, standardized covariance residuals were examined to assess how well the model-implied covariances reproduced model-implied item covariances. High positive residuals indicate the model is underestimating the relationship between items, meaning they share some construct-irrelevant variance. Negative values suggest the items are less related than the model suggests.

[INSERT TABLE 3]

As shown in Table 3, all three models failed the x^2 test, meaning they fit significantly worse than a just-identified model. All models met the close-fit guidelines chosen for the CFI and SRMR and were close to the chosen RMSEA values, indicating at least approximate model-data fit. The Satorra-Bentler x^2 difference test indicated each model fit significantly worse than the one in which it was nested, even though each model had similar values for the approximate fit indices. Thus, standardized covariance residuals were examined to better understand the comparative fit of the three models.

Raykov and Marcoulides (2000) suggested covariance residuals of |3| or more could be considered "large" (p. 43). The three-factor model had ten residuals greater than |3|, although most were close to |3|. The two-factor model naturally had more large residuals, with 22 greater than |3|, and seven of these greater than |4|. Thus, although the two-factor model met the criteria for global fit, it added 12 large standardized covariance residuals. Four of the large, positive standardized covariance residuals involved two Information Control items, suggesting they shared something beyond the factor they were set to represent. There were also two large, negative standardized covariance residuals between Affect of Service and Information Control items, suggesting the model was overestimating their relationship and meaning the items had less relationship with one another than specified by the model. Thus, there was less empirical support for the two-factor model, especially with regard to local misfit. Also, theoretically it would be difficult to interpret this two-factor model because no one has defined a meaningful construct that incorporates the items from both Affect of Service and Information Control. Such a theoretical definition would have to explain, for example, why AS18 (Willingness to help others) and IC05 (A library web site enabling me to locate the information on my own) are thought to measure a common underlying construct.

[INSERT TABLE 4]

The one-factor model had 30 standardized residuals greater than three, with 21 of these relating to Library as Place. Seventeen standardized covariance residuals were greater than |4|. As shown in Table 4, the correlation residuals for the one-factor model showed a positive relationship between seven pairs of Library as Place items greater than |.10|, indicating the presence of the unspecified Library as Place factor. Of note, these high correlation residuals disappeared when examining the two-factor model, where Library as Place was given its own factor. Therefore, although the one-factor model met the lower guidelines chosen for overall fit, the differences in

residuals between the three models suggested the data does not fit the one-factor model. In summary, the three-factor model fit well globally and locally; the two-factor model, while fitting globally, had a notable amount of local misfit; and the one-factor model did not meet all the cutoffs for global nor did it demonstrate local fit. For this reason, the three-factor model will be championed and areas of misfit will be examined in more detail.

7. Examination of the Three-Factor Model

Substantial amounts of variance in each item were explained by their factors, as shown by R^2 values, which can be calculated by subtracting the standardized error variances in Figure 1 from one. The proportion of variance accounted for (Fornell & Larcker, 1981) was over .50 for Affect of Service (.573) and Library as Place (.551), suggesting the items (as a group) were explained adequately well by their associated factors; however, the Information Control items were less well explained by their factor (.498). Inter-item correlations (see Table 2) and factor pattern coefficients (see Figure 1) for the three-factor model also suggested the instrument was generally performing as theorized for undergraduates.

Correlations were highest for Affect of Service items, with many correlations above .500 and some in the .600-.700 range. Standardized factor pattern coefficients for these items were generally quite high, ranging from .63 to .84. However, eight of the ten large standardized covariance residuals for the model were related to Affect of Service, suggesting some problematic items. The correlation residual (.183) between AS01 ("Employees who instill confidence in users") and AS04 ("Giving users individual attention") suggested these two items may share something not represented by Affect of Service. These items were also the least correlated with other items in the group and had the lowest standardized factor pattern

coefficients. The standardized error variance of both AS01 and AS04 was fairly large (.60), indicating Affect of Service did not explain much variance in these items for undergraduates.

It is not clear why these two items were so problematic. AS01 ("Employees who instill confidence in users") appears quite similar in content to AS11 ("Employees who have the knowledge to answer user questions"); thus, it seems logical that employees the user perceives as knowledgeable could stimulate confidence. However, whereas AS01 and AS04 had a standardized covariance residual of 8.09 (the largest in the study), AS01 and AS11 had a standardized covariance residual of only -.77. Perhaps the fact that AS01 was the first question where users assigned minimum, desired, and perceived ratings introduced some measurement error. However, the correlation residuals between AS01 and AS 04 were not large enough in either Lane et al.'s (2012) study of the three-factor model or Thompson et al.'s (2003) study of a previous instrument version for them to specify it in their revised models. Because they did not report the value of their residuals, it could be the residuals were comparable to this study but were just not large enough to suggest post-hoc model re-specification. The problem could, however, demonstrate something specific to this sample: for example, a systematic group difference could be introducing error. JMU may want to consider interviewing students in detail about their feelings of confidence in library services and impressions of the individualized attention received.

Correlations among Information Control items were less impressive, with many values less than .500, and only two values greater than .600. However, the standardized factor pattern coefficients still indicated a large influence from the factor on the items (.53 to .78). In addition, a number of standardized error variances among Information Control items were over .50, including IC02, IC10, and IC14. The construct of Information Control explained less about the variance in these items. The correlation residual (.130) between IC02 ("Making electronic

resources accessible from my home or office") and IC05 ("A library Web site enabling me to locate information on my own") suggested that these two items share something beyond the construct of "Information Control" for undergraduates. The shared variance here was possibly due to the frustration users experience with online access and authentication to library resources, which has been expressed by JMU users anecdotally and at library service desks. Neither residual was used in previous studies to support model re-specification (Lane et al., 2012; Thompson et al., 2003), leaving no point of comparison for this study.

Library as Place had correlations ranging from .454 to .615 and factor pattern coefficients ranging from .67 to .81, suggesting generally strong relationships between the items and this construct. However, LP03 ("Library space that inspires study and learning") was associated with five of the model's highest standardized covariance residuals, all with residual correlations between |.07| and |.08|, suggesting this item may be multidimensional, at least for undergraduates. Indeed, satisfaction with space could be very different depending on whether one was hoping to work in a group (a common activity in libraries) or alone. At JMU Libraries, frequent complaints have been heard from students about both the need for group study space and the need for quiet study space in the libraries. The error covariance between LP03 and LP08 ("Quiet space for individual activities") was found to be significant enough in both Lane et al. (2012) and Thompson et al. (2003) that they chose to include an error correlation in their revised models.

More than half the variance in item LP21 ("Community space for group learning and group study") was not explained by Library as Place. It is unknown how this compares with other studies, because this item did not appear in the 24-item version studied by Thompson et al., (2003); because Lane et al. (2012) did not report error variances in their study; and Kieftenbeld and Natesan (2013) reported them for only their factor variance invariance model. This also

suggested students' ideas about community space may contain some elements not strictly related to the Library as Place construct.

Both Information Control and Affect of Service are less distinct than Library as Place. This can be seen by looking at the correlation residuals in Table 4: while high correlation residuals emerged for Library as Place items when the data was forced to fit a one-factor model, Information Control or Affect of Service items did not pair up to form residuals. Even with the two-factor model, there was only one correlation residual between two Information Control items greater than |.10|. This was related to the high, positive correlation between Affect of Service and Information Control (.89), which provided additional evidence that these constructs may not be very distinct for undergraduates. This correlation was also the highest in Lane et al.'s (2012) three samples combining faculty and students, ranging from .83-.86 (p. 24-26).

[INSERT TABLE 5]

To further explore the distinctiveness of Affect of Service and Information Control, structure coefficients were calculated (Mels, 2012) and compared with the standardized factor pattern coefficients (Table 5). Structure coefficients represent the total relationships between an item and a factor. For items not directly representing a factor, the structure coefficient is an indirect effect of the factor on the item via the factor correlations for items for which they were not serving as indicators. There were at least moderate relationships between most items and each of their non-associated factors despite the pattern coefficient being fixed to zero in the model. For most items, however, the factor pattern coefficient was high, and it was higher than the structure coefficients associated with the non-associated factors. However, two areas of weakness could be seen. IC02 had a fairly weak relationship to Information Control (.53), and the structure coefficients associated with Affect of Service and Library as Place were not that much lower, making IC02's contribution to each construct's definition fairly equal. In Lane et

al.'s 2012 study, this pattern was also evident, although less extreme; however, Kieftenbeld and Natesan (2013) found a stronger factor pattern coefficient for IC02[6] than structure coefficients for undergraduates, graduates, and faculty on the other dimensions. Of greater note, three items written to represent Affect of Service (AS13, AS15, and AS18) had an equal or greater relationship to the Information Control factor than four of the eight Information Control items written to represent Information Control. Thus, although the high correlation between Affect of Service and Information Control suggested a lack of distinction, the structure coefficients provided unique evidence that the Information Control factor was the more problematic of the two because four of its items were less relevant than were other items not written to represent Information Control. This pattern did not emerge in any of Lane et al.'s (2012) three samples. Kieftenbeld and Natesan (2013) found AS13, AS15, and AS18 to have a greater relationship than IC02 and IC05, demonstrating a similar, albeit weaker, pattern than did this study. Kieftenbeld and Natesan (2013) did not report separate pattern coefficients for undergraduates, making it unclear whether the pattern difference between their study and this one was because this study focused on just undergraduates or if there is something specific to this sample causing this pattern. Finally, comparing the factor pattern and structure coefficients supported the distinction between Library as Place and both Affect of Service and Information Control for undergraduates, similar to Lane et al. (2012) and Kieftenbeld and Natesan (2013).

Reliabilities were calculated using McDonald's (1999, p. 205-206) omega because items were not tau-equivalent (tau-equivalence is assumed by Cronbach's alpha). Reliabilities for each

---

[6] Natesan confirmed the factor pattern coefficients were equivalent across groups, and therefore reported the factor pattern coefficients for all groups combined (personal communication, July 20, 2013)

of the three subscale scores were very good (Affect of Service = 0.918; Information Control = 0.879; Library as Place = 0.848).

8. Discussion

A three-factor model of user perceptions of library service quality was supported by responses from undergraduates to the LibQUAL+ model. Cross-validation studies would provide evidence of this finding's generalizability (MacCallum & Austin, 2000, p. 211-212). Because local areas of misfit have not been previously examined and discussed, it is unknown whether this study's findings related to local areas of misfit would reproduce in independent samples of undergraduates.

Although model-data fit for the three-factor model was quite good, several of the areas of local misfit suggested the need to continue instrument development. The measurement of Information Control construct seems problematic. In addition to the empirical issues discussed above, the construct contains three theoretical ideas: the idea of remote or electronic access; the idea of having the right print and electronic collections; and the idea of independence when using the library. The creation of Information Control from two previous factors, Personal Control and Information Access (Thompson, et al., 2003), may be part of the cause. Although the instrument appears to be measuring this construct reliably, it is not clear how to interpret the results of "Information Control" scores. Studies of specific library systems have often been grounded in the idea that users want to be self-sufficient (Blummer, 2007; Freund, Nemmers, & Ochoa, 2007; Letnikova, 2003), and it seems logical that "the right collections" are important drivers of user perceptions of library service quality. But with one subscore representing both these aspects together, a library wanting to improve its Information Control scores may not be sure where to invest their resources. It may be that a holistic assessment of library service quality such as

LibQUAL+ is not the place to discriminate between these aspects; however, this issue does seem to introduce ambiguity into the Information Control construct. This study only examined misfit in one data sample; although the specific misfit may be local to JMU, other studies have found ambiguity with the Information Control construct (Kieftenbeld & Natesan, 2013). It is also important to determine if the local areas of misfit with respect to Information Control are present in other groups (e.g., faculty, graduate students), or if they are specific to undergraduates; so far, results from Kieftenbeld and Natesan (2013) appear to indicate consistency across these groups in factor pattern coefficients and structure coefficients.

Examining the residuals also provided a few areas for JMU to investigate more closely with undergraduates. Public services may want to explore the extent to which undergraduates feel confident in and attended to by library staff. Are there some student groups who feel more confident and perceive greater individualized attention than others? A disturbing study by Shachaf and Horowitz (2006) suggested library service behavior may differ by perceived patron race or ethnicity. Although JMU does not have a highly diverse population, it is still important to consider and investigate potential group differences. Results from groups such as number of years at JMU, academic departments, or gender could be compared; if such analyses are not fruitful, additional, more targeted surveys could be conducted. Likewise, access to electronic resources and the library web site offer areas where the library could explore options for improvement and use a future iteration of LibQUAL+ to measure progress. Finally, the library has anecdotally heard from students that noise policies need to be clearer and better enforced, which may be a cause for the residual between LP03 and several other items. While examining the means of these items from later LibQUAL studies could show improvement, reproducing this factor analysis and studying the residuals could illustrate if the constructs are better measuring the JMU population.

In order to avoid distraction, the means on subscales and items were not presented in the results. However, looking at them briefly illustrates how residuals can reveal potential problem areas for investigation that means do not uncover. Residual problems were detected related to each construct, but the associated items may or may not have had means that would catch a researcher's interest. For example, the means on AS01 and AS04 were almost a point lower than for other Affect of Service items. The overall mean for Affect of Service in this sample was 7.43, while the means on AS01 ($M = 6.86$) and AS 04 ($M = 6.69$) were the lowest for the subscale. This could have raised questions about these items without residual analysis. However, the focus of investigation would be "why is perception of quality lower on these two items," not "what construct is it that is affecting responses on these items in a systematic way?" In other cases, the means for items were not remarkably different from their subscale mean: The mean for IC05 ($M = 7.41$) was close to the mean for Information Control overall ($M = 7.44$), and the mean for LP03 ($M = 7.19$) was not that much lower than the average for Library as Place ($M = 7.37$). The residual analysis was what identified the opportunity for further investigation.

9. Future Research

Some areas of overall library service quality may not be represented by LibQUAL+. The modern academic library provides teaching and research support through varied programs. Many libraries, including JMU's, have an extensive instruction program for students and a library liaison program to support faculty and student research needs. A study of which optional questions libraries chose to include after the 22 core items (Thompson, Cook, & Kyrillidou, 2006) showed three of the most popular related to teaching and awareness of library resources. These items correlated most highly with Affect of Service, although their correlations with Information Control were not much lower. While libraries consider such items relevant to service quality, it is unclear whether users do. A research agenda exploring the relevance of library

instruction to user perceptions of service quality could provide important information for both instrument and program development.

LibQUAL+ may illuminate gaps in other areas. Library assessment research increasingly focuses on connecting library data to institutional-level outcomes (Hinchliffe, Oakleaf, & Malenfant, 2012; Schwieder & Hinchliffe, 2012; Stemmer & Mahan, 2012), often using data from surveys such as Association for College and Research Libraries (ACRL, 2013) Library Trends & Statistics and the National Center for Education Statistics (NCES, 2013) Academic Libraries Survey. A problem with these surveys is the limited information on virtual interactions with the library and on electronic resources. It could be fruitful to consider how LibQUAL+ data might be used in combination with ACRL and NCES survey data to reveal a fuller picture of library assessment. Vice versa, the ALS and ACRL data may be explored for possible explanations for LibQUAL+ residuals at specific institutions. Kayongo and Jones (2008) correlated numerous ARL variables with LibQUAL+ gap scores to investigate relationships between institutional characteristics and faculty perceptions of library service quality. They found a strong correlation between Total Materials Expenditures and Faculty Adequacy Gap Scores, but no correlation among institutional characteristics such as expenditures for electronic serials. Hunter and Perret (2011) correlated LibQUAL+ Minimum, Perceived, and Desired scores (apparently for all groups together) with data from the ACRL Library Trends and Statistics database. They found 17 statistically significant correlations related to Information Control, and also raised additional questions about the utility of gap scores. It would be interesting to follow these methods for undergraduates; further, it would be interesting to attempt correlations between institutional variables and perceived scores as well as between institutional variables and gap scores.

In addition to local and demographic questions, participants must provide three scores for the 22 LibQUAL+ items (minimum, desired, and perceived), making the instrument quite long, which can suppress response rates (Stanton, Sinar, Balzer, & Smith, 2002). After developing and testing a shorter version of LibQUAL+ using eight of the 22 core items, Thompson, Kyrillidou, and Cook (2009) noted that scores were lower for the short version, leading them to believe "persons with more positive views are disproportionately likely to respond to the [long] survey" (p. 11). Therefore, a shorter instrument could not only increase the response rate, but reduce the positive skew. The LibQUAL+ Lite developers have found scores on the shorter version to be comparable with the longer LibQUAL+. In her dissertation, Kyrillidou (2009) examined participation rates and completion times and compared results between LibQUAL+ and LibQUAL+ Lite using data from more than 10,000 library users across 14 institutions in a randomized trial. Overall, participation increased by 13% when people used the Lite form (Kyrillidou, 2009, p. 134). Kyrillidou noted that having a higher response rate improves the ability to analyze more granular user groups and to garner more qualitative comments. Completion times improved by an average of 4.01 minutes for the Lite version (p. 138). While Kyrillidou found statistically significant differences on scores between long and Lite LibQUAL+ versions, she noted the effect size, or practical difference, was very small ($\eta^2 <= 1.16\%$) (2009, p. 141). Finally, Kyrillidou noted that undergraduate students showed the least difference between long and Lite LibQUAL+ forms (p. 150). An additional avenue to explore for reducing the instrument length would be to gather only perceptions scores, especially considering the questionable usefulness of the "gap score" concept (Roszkowski et al., 2005).

The previous methodological problems with LibQUAL+ research should be avoided in future studies. Previous LibQUAL+ researchers have often correlated indicator errors to improve model fit with no discussion of empirical rationale (Lane et al., 2012; Thompson et al., 2003).

The goal of SEM is to test a theory and to explain patterns of covariance among variables, not to improve model fit (Kelloway, 1995; Mueller, 1997). Correlating errors based solely on software recommendations conflicts with this goal, can produce ungeneralizable findings, and prevents analysis of local model misfit (MacCallum et al., 1992). If researchers choose to perform post-hoc modifications, should avoid the appearance of doing so based solely on software recommendations by clearly stating any theoretical basis for the modifications, acknowledging that modifications improve fit to a single sample only, and noting the importance of testing the new (modified) model in subsequent studies using independent samples (MacCallum et al., 1992). A superior approach would be to specify multiple alternative models *a priori* based on emerging theory (MacCallum et al. 1992)

Correlating errors (and not reporting residual values) also limits the ability to compare areas of local misfit across independent studies. High residuals can identify the presence of an additional variable or a method effect (Quintana & Maxwell, 1999, p. 493). In this study, the researchers conceived of frustration as a possible influence on items IC02 and IC05. Because past studies did not report residuals, it is somewhat unclear if this was a repeated finding, or specific to this sample. If areas of local misfit are repeated across independent samples, this could suggest important changes in the instrument. If they are not replicated, the misfit could point out ways in which LibQUAL+ does not measure a particular library's services effectively. That library might wish to conduct a targeted study on the area of misfit. For example, the misfit between IC02 and IC05 in this study's data may suggest improvements to reduce frustration with the local authentication system rather than in the web site itself. On the other hand, the large residuals related to LP03, "Library space that inspires study and learning," seem to have been found in previous studies, so this may be characteristic of the instrument rather than one that

needs local investigation. Without additional evidence, however, these ideas will remain mere hypotheses.

Kieftenbeld and Natesan's (2013) measurement invariance study provides a foundation for future research involving mean comparisons between student and faculty groups; however, replication in additional samples would increase the support for measurement invariance of the LibQUAL+ instrument. Until their findings have been replicated, researchers should thoughtfully consider whether combining students and faculty into one group (as has often been done in the past) is necessary to answer their research questions, and whether it might be more appropriate to look at scores for just one group at a time. Invariance studies may not be necessary for every conceivable division of data into groups: for example, JMU has two major library buildings and the survey asked users which library they used most often. If the demographics of the user groups are similar, and no theoretical reason is identified why these user groups might differ, perhaps an invariance study does not necessarily need to be conducted.

10. Conclusion

LibQUAL+ has become the de facto standard for providing a holistic assessment of user perceptions of library services. Sharing a common instrument has benefitted libraries in some ways. However, at least with respect to undergraduates, the Affect of Service and Information Control constructs should not be thought of as particularly distinct, and Information Control does not seem to be measured satisfactorily by this instrument, statistically and theoretically. The LibQUAL+ measure can be used to estimate perceptions of library service quality for undergraduate students, but the subscale scores for Information Control should be interpreted with some caution because they seem to measure multiple constructs. Past and future research about LibQUAL+ could be illuminated by identifying areas of local misfit and examining

residuals and factor pattern coefficients more closely. Residual analysis can not only improve the instrument but also help institutions identify additional areas for local investigation. In light of Keiftenbeld and Natesan's (2013) promising results from their measurement and structural invariance study, the richness of LibQUAL+ data offers researchers many opportunities for further exploration.

Acknowledgements

References

Accardi, M. T., Cordova, M., & Leeder, K. (2010). Reviewing the library learning commons: History, models, and perspectives. *College & Undergraduate Libraries, 17*(2), 310-329. doi: 10.1080/10691316.2010.481595

Antell, K., & Engel, D. (2006). Conduciveness to scholarship: The essence of academic library as place. *College & Research Libraries, 67*(6), 536-560.

Association for College and Research Libraries. (2013). Academic library statistics. Retrieved from http://www.ala.org/acrl/publications/trends

Association for Research Libraries. (2011). *LibQUAL+ 2011 Survey*. Retrieved from http://libqual.org/documents/LibQual/publications/ARL_Notebook_2011.pdf

Association for Research Libraries. (2012a). *General FAQs: What is LibQUAL+?* Retrieved from http://libqual.org/about/about_lq/general_faq

Association for Research Libraries. (2012b). *LibQUAL+ tools: Learn about LibQUAL+* [Presentation]. Retrieved from http://libqual.org/about/about_survey/tools

Bentler, P. M. (1998, March 10). Kurtosis, residuals, fit indices [Electronic mailing list message]. Retrieved from https://listserv.ua.edu/cgi-bin/wa?A0=semnet

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16*, 78-117.

Blummer, B. A. (2007). A literature review of academic library web page studies. *Journal of Web Librarianship, 1*(1), 45-64. doi: 10.1300/J502v01n01_04

Bontempo, D. E., & Hofer, S. M. (2007). Assessing factorial invariance in cross-sectional and longitudinal studies. In A. D. Ong & M. H. M. van Dulmen (Eds.), *Oxford handbook of methods in positive psychology* (pp. 153-175). New York: Oxford University Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.) *Testing structural equation models* (pp.136-162), Newbury Park, CA: Sage.

Bryant, F., & Satorra. A. (2012). Principles and practice of scaled difference chi-square testing. S*tructural Equation Modeling, 19*, 372-398.

Carnegie Foundation for the Advancement of Teaching. (2013). Institution lookup. Retrieved from http://classifications.carnegiefoundation.org/lookup_listings/institution.php

Cook, C. (Guest Ed.). (2002). The maturation of assessment in academic libraries: The role of LibQUAL+^{TM} [Special issue]. *Performance Measurement and Metrics, 3*(2), 34-112.

Cook, C., Heath, F., & Thompson, B. (2002). Score norms for improving library service quality: A LibQUAL+ study. *portal: Libraries & the Academy, 2*(1), 13-26.

Desai, C. M., & Graves, S. J. (2008). Cyberspace or face-to-face: The teachable moment and changing reference mediums. *Reference & User Services Quarterly, 47*(3), 242-255.

Durrance, J. C. (1989). Reference success: Does the 55 percent rule tell the whole story? *Library Journal, 114*(7), 31-36.

Durrance, J. C. (1995). Factors that influence reference success: What makes questioners willing to return? *Reference Librarian, 49*, 243-265.

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation models. In G.R. Hancock & R.O. Mueller (Eds.). *A second course in structural equation modeling* (2nd ed., pp. 439-492). Charlotte, NC: Information Age.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods 9*(4), 466-491.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research,18*, 39-50.

Freund, L., Nemmers, J. R., & Ochoa, M. N. (2007). Metasearching: An annotated bibliography. *Internet Reference Services Quarterly, 12*(3), 411-430.

Friend, L. (1985). Independence at the terminal: Training student end users to do online literature searching. *Journal of Academic Librarianship, 11*(3), 136-141.

Greenwood, J. T., Watson, A. P., & Dennis, M. (2011). Ten years of LibQUAL+: A study of qualitative and quantitative survey results at the University of Mississippi 2001-2010. *Journal of Academic Librarianship, 37*(4), 312-318. doi: 10.1016/j.acalib.2011.04.005

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

Hernon, P., & McClure, C. R. (1986). Unobtrusive reference testing: The 55 percent rule. (cover story). *Library Journal, 111*(7), 37.

Hinchliffe, L., Oakleaf, M., & Malenfant, K. (October, 2012). *What do we want to know? Articulating a research agenda for the value of academic libraries*. Presentation at Library Assessment Conference, Charlottesville, VA. Retrieved from http://libraryassessment.org/bm~doc/Hinchliffe_Lisa_2012.pdf

Horst, S. J., Finney, S. J, & Barron, K. E. (2006). Moving beyond academic achievement goal measures: A study of social achievement goals. *Contemporary Educational Psychology 32*, 667-698.

Hu, L. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Hunter, B., & Perret, R. (2011). Can money buy happiness? A statistical analysis of predictors for user satisfaction. *The Journal of Academic Librarianship 37* (5), 402-408.

Ithaka. (2010). Faculty survey 2009: Key strategic insights for libraries, publishers, and societies. Retrieved from http://sr.ithaka.org/research-publications/us-faculty-survey-2009James Madison University. (2011). James Madison University Facts & Figures, 2010-2011. Retrieved from http://www.jmu.edu/instresrch/statsum/2010_11/1_Facts%20&%20Figures_10-11.pdf

James Madison University (2013a). *Just the facts*. Retrieved from http://www.jmu.edu/about/fact-and-figures.shtml

James Madison University. (2013b). *JMU – mission*. Retrieved from http://www.jmu.edu/jmuplans/mission_statement.shtml

James Madison University. (2013c). *Defining characteristics*. Retrieved from http://www.jmu.edu/jmuplans/defchar.shtml

Jöreskog, K. G., & Sörbom, D. (2007a). LISREL (Version 8.80) [Computer software]. Lincolnwood, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (2007b). PRELIS (Version 2.80) [Computer software]. Lincolnwood, IL: Scientific Software International.

Kayongo, J., & Jones, S. (2008). Faculty perception of information control using LibQUAL+[TM] indicators. *The Journal of Academic Librairanship 34* (2), 130-138.

Kelloway, E. K. (1995). Structural equation modeling in perspective. *Journal of Organizational Behavior, 16,* 215-224.

Kieftenbeld, V., & Natesan, P. (2013). Examining the measurement and structural invariance of LibQUAL+ across user groups. *Library and Information Science Research 35*, 143-150.

Kyrillidou, M. (2009). *Item sampling in service quality assessment surveys to improve response rates and reduce respondent burden: The "LibWual+® lite" randomized control trial (RCT).* (Doctoral dissertation). Retrieved from ProQuest (UMI 3406768).

Lane, F. C., Anderson, B., Ponce, H. F., & Natesan, P. (2012). Factorial invariance of LibQUAL+ as a measure of library service quality over time. *Library and Information Science Research, 34*, 22-30. doi: 10.1016/j.lisr.2011.07.004

Letnikova, G. (2003). Usability testing of academic library web sites: A selective annotated bibliography. *Internet Reference Services Quarterly, 8*(4), 53-68. doi: 10.1300/J136v08n04_04

Lippincott, S., & Kyrillidou, M. (2004). *How ARL university communities access information: Highlights from LibQUAL+*. ARL Digital Publications.

MacCallum, R., & Austin, J. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201-226.

MacCallum, R., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111,* 490-504.

Marsh, H. W., Hau, K-T., & Wen, Z (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers to overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320–341.

Maruyama, G. M. (1998). *Basics of Structural Equation Modeling.* Thousand Oaks, CA: Sage.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

Mels, G. (2012). Structure coefficients for the LISREL model for confirmatory factor analysis. Retrieved from http://www.ssicentral.com/lisrel/techdocs/SCforCFA.pdf

Mueller, R. O. (1997). Structural equation modeling: Back to basics. *Structural Equation Modeling, 4,* 353-369.

Murfin, M. E., & Gugelchuk, G. M. (1987). Development and testing of a reference transaction assessment instrument. *College & Research Libraries, 48*(4), 314-338.

National Center for Education Statistics. (2013). *Library statistics program.* Retrieved from http://nces.ed.gov/surveys/libraries/academic.asp

Nilsen, K., & Ross, C. (2006). Evaluating virtual reference from the users' perspective. *Reference Librarian, 46*(95/96), 53-79.

OCLC, Inc.. (2010). Perceptions of libraries, 2010: Contexts and communities. Retrieved from http://www.oclc.org/reports/2010perceptions.en.html

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, & WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling, 7*, 557-595.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Refinement and reassessment of the SERVQUAL scale. *Journal of Retailing, 67*(4), 420.

Quintana, S. M., & Maxwell, S. E. (1999). Implications of recent developments in structural equation modeling for counseling psychology. *The Counseling Psychologist, 27*, 485-527.

Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling.* Mahwah, NJ: Lawrence Erlbaum.

Rimland, E. L. (2007). Do we do it well? A bibliographic essay on the evaluation of reference effectiveness. *Reference Librarian, 47*(2), 41-55.

Roszkowski, M. J., Baky, J. S., & Jones, D. B. (2005). So which score on the LibQUAL+ tells me if library users are satisfied? *Library and Information Science Research, 27*, 424-439. doi: 10.1016/j.lisr.2005.08.002

Saunders, E. S. (2007). The LibQUAL+ phenomenon. *Reference & User Services Quarterly, 47*(1), 21-24.

Schwartz, D. G., & Eakin, D. (1986). Reference service standards, performance criteria, and evaluation. *Journal of Academic Librarianship, 12*(1), 4-8.

Schwieder, D., & Hinchliffe, L. (October, 2012). *NCES Datasets and Library Value: An Exploratory Study of the 2008 Data.* Presentation at Library Assessment Conference, Charlottesville, VA.

Seeholzer, J., & Salem, J. A. (2010). The learning library: A case study of the evolution from information commons to learning commons at Kent State University Libraries. *College & Undergraduate Libraries, 17*(2), 287-296. doi: 10.1080/10691316.2010.481609

Shachaf, P., & Horowitz, S. (2006). Are virtual reference services color blind? *Library & Information Science Research, 28*(4), 501-520. doi: 10.1016/j.lisr.2006.08.009

Shill, H. B., & Tonner, S. (2004). Does the building still matter? Usage patterns in new, expanded, and renovated libraries, 1995 - 2002. *College & Research Libraries*, *65*(2), 123-150.

Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology 55*, 167-194.

Stemmer, J., & Mahan, D. M. (October, 2012). *Using library user surveys to assess the library's impact on student outcomes.* Presentation at Library Assessment Conference, Charlottesville, VA.

Thompson, B., Cook, C., & Heath, F. (2001). How many dimensions does it take to measure users' perceptions of libraries?: A LibQUAL+ study. *portal: Libraries & the Academy, 1*(2), 129-138.

Thompson, B., Cook, C., & Heath, F. (2003). Structure of perceptions of service quality in libraries: A LibQUAL+ study. *Structural Equation Modeling, 10*(3), 456-64.

Thompson, B., Cook, C., & Kyrillidou, M. (2005). Concurrent validity of LibQUAL+ scores: What do LibQUAL+ scores measure? *Journal of Academic Librarianship, 31*(6), 517-522.

Thompson, B., Cook, C., & Kyrillidou, M. (2006). Using localized survey items to augment standardized benchmarking measures: A LibQUAL+ study. *portal: Libraries and the Academy, 6*(2), 219-230.

Thompson, B., Cook, C., & Thompson, R. L. (2002). Reliability and structure of LibQUAL+ scores: Measuring perceived library service quality. *portal: Libraries & the Academy, 2*(1), 3-12.

Thompson, B., Kyrillidou, M., & Cook, C. (2007). User library service expectations in health science vs. other settings: A LibQUAL+ study. *Health Information & Libraries Journal, 24*, 38-45. doi: 10.1111/j.1471-1842.2007.00744.x

Thompson, B., Kyrillidou, M., & Cook, C. (2008). Library users' service desires: A LibQUAL+ study. *The Library Quarterly, 78*(1), 1-18. Retrieved from http://www.jstor.org/stable/10.1086/523907

Thompson, B., Kyrillidou, M., & Cook, C. (2009). Item sampling in service quality assessment surveys to improve response rates and reduce respondent burden: The "LibQUAL+ Lite" example. *Performance Measurement & Metrics 10*, 6-16.

Unigo. (2013). *James Madison University*. Retrieved from http://www.unigo.com/james_madison_university/reviews

U.S. Department of Education, The Secretary of Education's Commission on the Future of Higher Education. (2006, September). *A test of leadership: Charting the future of U.S. higher education*. Retrieved from: http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf

Vaska, M., Chan, R., & Powelson, S. (2009). Results of a user survey to determine needs for a health sciences library renovation. *New Review of Academic Librarianship, 15*(2), 219-234. doi: 10.1080/13614530903240635

Vilelle, L., & Peters, C. C. (2008). Don't shelve the questions: Defining good customer service for shelvers. *Reference & User Services Quarterly, 48*(1), 60-67.

Wei, Y., Thompson, B., & Cook, C. C. (2005). Scaling users' perceptions of library service quality using item response theory: A LibQUAL+ study. *portal: Libraries & the Academy, 5*(1), 93-104.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 57-75). Thousand Oaks: Sage.

Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist, 34*, 719-751.

Yang, Y., & Green, S. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment, 29*, 377-392.

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling 17*(3), 392-423.