**­­Validating Students’ Library Experience Survey using Rasch Model**

Jung Mi Scoulasa, \*, Beyza Aksu Dunyaa,b, Sandra L. De Grootea,

a Richard J. Daley Library, University of Illinois at Chicago, 801 S. Morgan Street, Chicago, IL 60607, USA

b Faculty of Education, Department of Educational Sciences, Bartin University, Kutlubey, Merkez/Bartin, 74100, Turkey

\* Corresponding author: Email: jscoul2@uic.edu

**Abstract**

Few studies have investigated the psychometric property of new instruments developed locally to assess students’ library experience, such as their behavior, perception and attitude towards library use. Validating a survey instrument is critical because using untested assessment instruments is more likely to lead to a lack of credibility in the results, as well as a lack of accuracy in the study’s purpose (Sullivan, 2011). This study used data from a locally developed student library experience survey distributed in Spring 2018 at a large public research university library to examine undergraduate and graduate students’ behaviors, attitudes and perceptions of library use. To test the validity of the survey as to whether six subdomains of the library experience survey contributed to a single construct, library experience, the study used Rasch model framework. This study contributes to the field as a first example using the Rasch model for library assessment purposes. Suggestions for best practices are discussed.

1. **Introduction**

With increased emphasis using evidence-based data for decision-making and the pressure often felt within academic libraries to demonstrate their value, libraries must understand user’s needs and seek feedback for improvement. One popular assessment tool to achieve this goal is user surveys. Academic libraries have used either standardized surveys (e.g., LibQUAL+, Ithaca S+R), or locally developed surveys (e.g., Gunasekera, 2010; Mirza & Mahmood, 2012), and there are pros and cons using either type.

Benefits of standardized surveys like LibQUAL+ include features to allow benchmark their own institutions with other institutions, track changes over time, and test the validity and reliability of questions (Hinchliffe, 2015). Disadvantages of standardized surveys include survey expenses, too many questions (Voorbij, 2012), difficulty in understanding questions (Thompson, Cook, & Health, 2000), and difficulty in interpreting results (Dennis, Greenwood & Watson, 2013). The benefits of locally developed surveys include customizable questions, control over the length and number of questions, and immediate use of the data for improvement. Disadvantages of local surveys include “the challenges of developing valid and reliable questions” and “lack of national benchmarking data for comparative analysis” (Hinchliffe, 2015, p.252). The decision to use either standardized surveys or locally developed surveys will depend on the institution’s needs. However, using untested assessment instruments are more likely to decrease the of credibility of the results and lack of accuracy of measurement (Sullivan, 2011).

1. **Problem statement**

Shi and Levy (2005) stated “the library has progressed from irregular statistics collection, a piecemeal approach in evaluating services, to the study of users and user satisfaction, to systematic data collection and analysis” (p.268). When using locally developed library user surveys, this issue is coupled with the “challenges of developing reliable and valid questions” (Hinchliffe, 2015, p.252). While a standardized survey like LibQUAL+ scale has been examined by various researchers assessing whether the survey instrument is reliable and valid (e.g., Natesan & Aerts, 2016), little research investigates the psychometric property of locally developed instruments designed to assess students’ library experiences involving their behaviors, perceptions, and attitudes toward library use. This study aims to apply “unusual extensions of well-known methods and tools” to college students’ attitudinal data in the domain of library use experience (Library & Information Science Research, n.d.). Using Rasch model, the current study examined the validity of a student library experience survey that was locally developed by a large public research university library. Validating a locally developed instrument is an important step to confidently rely on the findings indicating when a library is meeting user needs and identify areas for improvement. The three research questions developed for establishing construct validity evidence for such instruments are provided below:

* Research question 1: Does the set of the items in the students’ library experience survey measure one single construct: library experience?
* Research question 2: Is this instrument measuring what it is intended to measure?
* Research question 3: Are the rating scale structures for subdomains appropriate for responses of students?

1. **Literature review**

*3.1. Background*

In 2015, the assessment committee at [University Name] developed its first student library experience survey, which focused on measuring how well the University Library met students’ needs. It was distributed to students in Spring 2016. For multiple reasons, such as yes/no responses to examine the relationships between library use and students’ academic achievement, a low response rate, and an abundance of qualitative response options, the previous student survey was revised in 2017. Yes/no response scales were changed to response scales about the frequency of use (e.g. daily, weekly, monthly, never) and wording was modified for clarification (citation withheld for blinding 1). The revised survey was test piloted with 8 students based on Rubin’s (2006) guided questions, including survey perception, confusion or difficult to answer, word understanding, and terms or concepts. For more survey revision details, refer to (citation withheld for blinding 1). To review full survey questions of either version, refer to (citation withheld for blinding 3). In Spring 2018, the revised survey was distributed to undergraduate and graduate students, this time including incentives for responding. The survey response rate significantly increased from 4% (*n*=1,087, 2016) to 8% (*n*=2,277, 2018). Survey results revealed online library resource use was positively correlated with students’ Grade Point Average (GPA), whereas students’ in-person library visits were negatively associated with their GPA (citation withheld for blinding 2).

While the student surveys were pilot tested by undergraduate students, it is unknown how accurately the instrument measures the outcome of interest, and whether the survey results are credible. Since this instrument will be continuously used every other year to assess students’ needs and examine their perceptions and satisfaction with library resources and services, testing whether this instrument is reliable and valid is critical. Consequently, the next step was to validate the survey “using a modern measurement model (e.g., Item Response Theory/Rasch models) for establishing validity and reliability evidence” (citation withheld for blinding 1, p.54). The current study assesses the validity of the 2018 student library experience survey using a Rasch model.

*3.2. Library User Survey*

“Academic libraries implement user surveys in order to better understand user needs, preferences, characteristics, work practices, feelings, perceptions, behaviors, attitudes, and satisfaction with library services and resources” (Hinchliffe, 2015, p.252). While measuring the user experience is inevitable, one challenge related to user experience measurement is “how to select appropriate measures to address the particularities of an evaluation context” (Law & Van Schaik, 2010, p. 313).

Given user experience measures vary by academic libraries depending on the survey purpose, there is no one unified survey instrument measuring library experience for college students. Based on previous studies on assessment tools measuring student library experiences, it is speculated the conceptualization of the students’ library use experience consists of multi subcomponents when using the library. That is, the user experience includes usability (e.g., the ease of using the website), cognitive (e.g., perceptions or beliefs), behavioral (e.g., frequency of use of resources), and affective aspects (e.g., satisfaction) in their library interactions.

For example, University of California Berkley library developed a user survey for faculty and graduate students in collaboration with the institution’s Survey Research Center consisting of the level of use, their satisfaction with resources and services, reasons for using the library, and preference for print or electronic materials and onsite versus remote use (Maughan, 1999). Similarly, a locally developed survey conducted by Gunasekera (2010) examined undergraduate students’ needs (purpose of library visits), their satisfaction with library materials and services, the importance of library materials for learning and research, and their opinions about library quality.

While the above studies utilized user surveys to assess users’ perceptions and satisfaction on the overall library service and resources, other libraries used user surveys to focus on specific areas. For example, a study conducted by Mirza and Mahmood (2012) in the Pakistani university libraries examined how users were satisfied with and perceived library resources. The researchers in the study developed a user survey questionnaire to explore their library use (periods), library use purpose, and satisfaction level with resources and services, using a five-point Likert scale from (1) “dissatisfied” to (5) “extremely satisfied”. In this survey, the resources and services included “online public access catalog, online databases, CD-ROM databases, internet service, email, virtual/electronic reference service, selective dissemination of information, current awareness service, scanning and printing” (p.130). In a study by Anderson (2016), combined data from website transaction survey responses, library websites visit data, and enrollment trends was examined. The survey questionnaire included “satisfaction, importance, usage levels of library services and resources” (p.9). “How often do you use the e-Library (i.e., Library website)?” with a scale from never to daily, was used to measure monthly library website usage. Other examples of studies that had user surveys focusing on a specific area include: a study of user needs for music collections conducted by Lai and Chan (2009); a study of perceptions and use of online resources (e.g., databases) by Verma (2016); a study of extended library-hour services by Atuase and Koufie (2017); a study of users’ learning behaviors in library spaces by Montgomery (2014); and a study of users’ perceptions of E-books over time by Cataldo, Shelton, Carrico and Botero (2015).

With the frequent use of locally developed user surveys, one critical question arises: does the survey instrument measure what it is intended to measure? To address this issue, some studies pilot-tested or pre-tested their user surveys with small numbers of users (e.g., Maughan, 1999; Mirza & Mahmood, 2012). In defense of this, Maughan stated “the library never intended to design a statistically valid and reliable survey,” and instead the survey was pre-tested with sample of faculty and graduate students (p.355). Similarly, Mirza and Mahmood (2012) pilot tested their survey with a sample of the library users from two out of eight universities who participated in their study. However, the above studies did not clearly state whether they pilot tested or pre-tested the survey (e.g., Atuase & Koufie, 2017; Cataldo et al., 2015). As Hinchliffe (2015) comments, locally developed surveys lack reliable and valid questions, and few studies have tested whether the survey instrument assesses what it is expected to measure.

*3.3. Rasch Model*

Researchers in many disciplines including behavioral sciences, social sciences and health sciences have commonly used Likert type ordinal scales in their survey instruments to test latent traits (unobserved variables) such as abilities (e.g., Brandt, 2008), attitudes and behaviors (e.g., Hendriks, Fyfe, Styles, Skinner, Merriman & Hendriks, 2012; Waugh, 2002), feelings (e.g., Battisti, Nicolini & Salini, 2010) because the Rasch model eliminates several limitations of true score theory including sample dependency of item, test indices, and item dependency of respondents (Aksu Dunya, McKown & Smith, 2019). Additionally, the Rasch model allows researchers to evaluate the strength and weakness of the instruments by identifying item difficulty levels (varying level of item difficulty) and ability levels (varying level of endorsing a high or low degree of items) (Boone, 2016). With the adoption of Rasch models, detection of potential measurement problems such as item bias or local item dependency are facilitated, which may be missed through classical validation methods such as factor analysis (Müller, 2020). Another strength of the Rasch model allows researchers to “improve the quality of quantitative measurement at the individual and the systematic level” for items through fit statistics (Boone, 2016, p.267). Fit values capture how accurately data fit to the Rasch model. Specifically, person-fit statistics allows researchers to identify whether respondents demonstrate unusual or random answering patterns. However, few studies utilize the Rasch model to test a survey instrument in the field of library and information science. The current study uses the Rasch model to assess whether six subdomains are collectively reflected in one single construct, library experience by employing a locally developed survey instrument.

1. **Methodology**

*4.1. Sample*

The data of the current study were analyzed from the 2018 students’ library experience survey. The survey was disseminated to 28,725 undergraduate and graduate students who were enrolled in one of 15 colleges in the Spring 2018. A total of 2,277 students completed the survey (response rate=8%). More females (63.7%) participated in the survey than males (36.2%). Most participants (69.4%) were between ages 16 to 25. More than half of the participants were undergraduate students (57%). Eighty-six percent of the participants were commuters. More than 36% of the participants were White, followed by Hispanic (21.3%), Asian American (18.6%), International (13%), African American (7.2%) and other (3.5%). Overall, the survey participants were representative of the university’s population (citations withheld for blinding 3)

*4.2. Instrument*

The University Library developed the set of questionnaire items, scored on a 5 point rating scale, based on the library user experience literature measuring students’ behaviors and perceptions toward library experience across six subdomains: frequency of library visit, library space satisfaction, importance of library resources, frequency of library resource, library website ease-of-use, and likelihood of recommending library services to other students. The six subdomains commonly appeared in other libraries’ user surveys (e.g., Mirza & Mahmood, 2012; Anderson, 2016; Cataldo *et al*., 2015).

The final version included 46 items, and the items scored on an ordinal scale were selected for the current study. The open-ended questions and nominal response questions were excluded due to Rasch model requirements (ordinal level data) and specific interest of the study. In addition, responses of “I do not use this service or resource” were treated as missing since non-users cannot reflect an attitude or perception about a service or resource. As a result, 46 items were included for analysis: frequency of library use (from Q17\_1 to from Q17\_2; 2 items), library space satisfaction (from Q18\_1 to Q18\_4; 4 items), frequency of library resource use (from Q20\_1 to Q20\_11; 11 items), importance of library resources and services (from Q21\_1 to Q21\_14; 14 items), library website ease-of-use (from Q22\_1 to Q22\_11; 11 items), and likelihood of recommending library services to other students (from Q23\_1 to Q23\_4; 4 items). Below are the list of 5-Point Likert Scales that were used per domain.

* Frequency of library visit was coded from (1) never to (5) daily
* Library space satisfaction was coded from (1) very dissatisfied to (4) very satisfied and (5) I do not use this space excluded
* Frequency of library resource use was coded from (1) never to (5) daily
* Importance of library resources and services for research or coursework was coded from (1) not at all important to (4) very important, and (5) I do not use this tool/service excluded
* Library website ease-of-use was coded from (1) very difficult to (4) very easy and (5) I do not use this service excluded
* Likelihood of recommending library services to other students was coded from (1) very unlikely to (4) very likely and (5) I do not use this service excluded.

Each of the six subdomains specified above were hypothesized to contribute a single construct: library experience. The detail of item numbers, wording and scales for the student library experience survey is displayed in the Table 1.

*[Insert Table 1 here]*

* 1. *Data analysis & procedures*

While items from the student library experience survey use the same number of Likert scale (a 5 Liker scale) within the subdomains, each subdomain uses a different rating scale structure. For example, item responses were: *never, once a month, once a week, multiple times in a week, and daily* for frequency of library visits and library resource use. However, library space satisfaction responses included: *very dissatisfied, dissatisfied, satisfied, very satisfied, and I do not use this space*. Given that every subscale has unique response categories, items should be grouped to prevent mistreating frequency item responses with satisfaction item responses. In order to do that, the psychometric properties of the instrument were assessed by employing the Grouped Rating Scale Model (GRSM; Linacre, 2002), a member of family of Rasch Models (Rasch, 1960) and a specific type of Andrich Rating Scale Model (RSM; Andrich, 1978). The RSM model is used for assessing validity of the instruments with rating scale structure. It makes it possible to assess respondents’ endorsement on one rating scale category over another when the instrument’s items use Likert-type scale (Andrich, 1978; Bond & Fox, 2007). Since the survey instrument has items with different category names and numbers, the authors opted to use GRSM that groups items by rating scale structure. The mathematical formula of the GRSM is given by Linacre (2002):

/) = - -

Where:

= Probability of person *n* observed in rating scale category *j*

= Probability of person *n* observed in rating scale category *j*-1

= Severity of attitude of person *n*

= Difficulty of item *i*

= Step difficulty of category *j* considering category *j*-1

*g*= Group of items specific to item i and to the rating scale of the group

All data was analyzed using WINSTEPS version 4.5.1 (Linacre, 2020), Rasch measurement software. The criteria employed for each question is addressed below.

* + 1. *Research Question 1: Does the set of the items in the students’ library experience measure one single construct: library experience?*

Analysis of the dimensional structure of the instrument serves as confirmation that all the attitudinal items measured a single underlying construct. This construct is: what is the degree of perception of students’ library experience?. For this purpose, principal component analysis of residuals (PCA) was used to analyze if any potential pattern (another dimension) existed in the residuals. In PCA, item correlations based on residuals are analyzed to detect potential underlying dimensions that may influence observed response patterns. For this, eigenvalues are constructed as linear combinations of the items. An eigenvalue that is smaller than 2.00 implies that the contrast means a random noise in responses rather than indicating structure for a secondary underlying dimension (Linacre, 2002). An eigenvalue higher than 2.00 may imply a systematic pattern (underlying construct) in the residuals. In addition, analysis of contrast loadings is used to detect any potential secondary dimension in the data. According to a commonly-used criteria in PCA (Stevens, 2002), a contrast with an associated eigenvalue larger than 2.00 can be a separate dimension if following rules apply: (a) a minimum number of three items with absolute loading values larger than .80 were loaded on it, (b) a minimum number of four items with absolute loading values larger than .60 were loaded on it, (c) a minimum number of 10 items with absolute loadings larger than .40 were loaded on it (Stevens, 2002).

* + 1. *Research Question 2: Is this instrument measuring what we intended to measure?*

Item fit was examined as an evidence of content validity, the degree to which instrument items measure the intended construct. In Rasch context, items that did not adequately fit the model requirements are detected using mean-squared fit indices: Outfit mean-squared (MNSQ) and Infit MNSQ fit indices. Mean-squared item fit indices indicate adherence of the responses to an item to the Rasch model expected responses to that item. Reporting Outfit MNSQ statistics is recommended over Infit, unless a substantial reason exists such as response data that is heavily contaminated with irrelevant residuals (Linacre, 2008).Outfit and infit statistics have an expected value of 1.00 and values that exceed 2.00 distort interpretations made based on scores (Linacre, 2008). Therefore, the authors particularly analyzed Outfit MNSQ item fit statistic to evaluate the fit of the data to model expectations.

* + 1. *Research Questions 3: Are the rating scale structures for subdomains appropriate for responses of students?*

To examine the consistency between observed ratings and the theoretical rationale that the instrument was built on, rating scale functioning and person fit were used. Linacre’s following guidelines (1999; 2002) were adapted for analyzing rating scale functioning:

* Each category should be selected by at least 10 respondents
* Average person measures increase monotonically from one response category to another
* Outfit MNSQ for each rating scale category should be smaller than 1.4
* The category thresholds should increase in accordance with the values of rating scale categories

The authors also visually inspected rating scale functioning through rating scale category curves. Each scale category should have a unique peak along the scale if the students use the rating scale as expected.

1. **Results** 
   1. *Summary Statistics*

Before discussing validity results, it is important to outline data fit to the Grouped Rating Scale Rasch model. A high value (26.61) obtained for item separation index indicates that the respondent sample was large enough to verify the item difficulty order in the instrument (Linacre, 2002). The item reliability value that is equal to the maximum value of 1.00 for this statistic implies that the respondent sample’s measure range is large enough to reproduce an item difficulty hierarchy. Person separation index is used to group people by the level of construct being measured, where a value smaller than 2.00 implies the instrument may not be sensitive enough to distinguish people by the degree of construct being measured they have (Linacre, 2002). The person separation value of 2.59 (> 2.00) indicates the items can distinguish respondents into two groups by the construct being measured (severity of library experience). The person reliability value of 0.87 (>.80) is above the lower limit for good reliability, indicating that the instrument is able to distinguish the students with strong and weak attitudes. This finding confirms that there is no need to add items or use longer scales.

The data fit GRSM well with one poorly fitting item. The item related to Q21\_7 (importance of library resources and services-special collections and university archives) measuring the utilization of a field-specific resources offered by the library. In terms of item fit, maximum values for both infit and outfit MNSQ exceed 1.5 cutoff (1.70 and 2.03 respectively). Although the maximum values of item fit statistics exceed 1.5 cutoff, Linacre (2008) suggested that values larger than 2.00 inappropriately change the inferences made from scores.Therefore, this result can be interpreted as a noise in the data that is not substantial. However, maximum values for person infit and outfit MNSQ is too large to attribute random noise (5.57 and 9.9 respectively). This result is an indication for further review of responses by misfitting persons to specific items to pinpoint the person misfit issue.

Lastly, Figure 1 shows Wright item-person map for visually inspecting if the instrument’s items capture the respondents well. On the figure, the left-hand column locates person measures’ spread along the latent variable, which is library use in this case (represented by #) and right-hand column locates item difficulty measures (represented by X). According to figure 1, it suggests that items are well-targeted to the respondents in the sample.

[Insert Figure 1]

Figure 1. Wright Item-Person Map

*5.2 Item Fit*

Individual item fit statistics for 46 items were generated by the WINSTEP software. According to the results, the authors reviewed Q 21\_7 (Importance of library resources and services- Special collection and university archives) since the associated outfit mean-squared statistics value of 2.03 was slightly above 2.00. None of the outfit MNSQ values, except for Q21\_7 (Importance of library resources and services- special collections and archives in university), are larger than 2.00; none of the infit values, except for Q21\_7 (Importance of library resources and services- Special collection and university archives) and Q20\_11 (Frequency of library resource use- patient care tools), are larger than 1.5. Except for the question related to the use of special collection and university archives, these findings establish content validity evidence for items in the survey.

* 1. *Dimensionality*

For Dimensionality, the authors checked if the response data reflects a single underlying construct; library experience. The authors provided strong evidence for library experience construct not only through collecting students’ self-reported frequency of library use but also by asking questions related to their perceptions and satisfaction, such as how important is the library for their studies and how satisfied they were with the library. Based on the results, the Rasch measures explain 52.9% of the variance in the data (Table 2). The eigenvalue of the residuals for the first contrast was 4.00. Yet, this finding does not imply a secondary dimension itself based on Stevens’ (2002) criteria explained in the methods section. The results showed that only one item [Q20\_1 (library resource use- journal articles)] had an absolute value of loading that was slightly greater than 0.60 and 6 items [Q17\_2 (frequency of library online visits); Q20\_2 (frequency of library resource use- subject specific databases); Q21\_2 (importance of library resources and services-subject specific databases); Q22\_1 (library website ease-of-use); Q18\_2 (library space satisfaction-collaborative study spaces); Q18\_3 (library space satisfaction-group study rooms)] had absolute loading values that were greater than .40. These values suggested residual components that address further interpretation for a secondary dimension (see Table 3) is not needed, and thus the instrument is measuring only one dimension.

*[Insert Table 2]*

*[Insert Table 3]*

*5.4. Rating Scale Functioning*

The results in six separate scales are listed on Table 4. Linacre’s (2002) guidelines have been used for assessing rating scale functioning for six grouped rating scales under GRSM (see Table 4).

*[Insert Table 4]*

Visual inspection of the rating scale functioning through rating scale category curves under each group revealed that categories had distinct peaks at some point along the scale as expected (see Figure 1). These distinct peaks of each curve mean that each category on the rating scale becomes the most probable option to be endorsed at some point, which justify that respondents utilized the rating scales as expected. This result can also be interpreted as there is no redundancy in number of response categories for each group of rating scales. An example category curve image generated by WINSTEPS software for Group 1 is provided in Figure 1 below. For other groups’ category curves, please see Appendix 1. The graphical results supported the evidence that respondents can effectively distinguish and use different rating scale category by respondents.

[Insert Figure 2]

Figure 2. Category Probability Curves for Group 1 (Q17\_1 ~ Q17\_2)

Note. Each curve labeled with numbers represent a response category on the rating scale group 1. Each peak refers the point where the probability (y- axis) of observing that category is the greatest. The curves continue along the latent variable (x-axis) meaning that there is always some probability of observing any category at any point. Same interpretation applies to all figures on Appendix 1.

1. **Discussions**

Few studies describing the results of locally developed library user surveys reported whether the locally developed surveys were tested to measure the validity of the survey items. The validation of survey instruments is critical as Sullivan (2011) argued, because using untested assessment instruments are more likely to lead to lack of credibility for results and a lack of accuracy for the study’s purpose. This study established construct validity evidence for a locally developed library user survey using the Rasch model.

The first research question was: *Does the set of the items in the students’ library experience measure one single construct: library experience?* The results revealed that all attitudinal items from the student survey measured single underlying construct; what is the degree of perception of students’ library experience. In response to the question, *Is this instrument measuring what it is intended to measure?,* the results demonstrated that all of the items except Q21\_7 (Importance of library resources and services- Special collections and university archives) were within the range of both Infit and outfit MNSQ, implying that the majority of items assessed what they were intended to measure. A potential reason for misfit for Q21\_7 may be related to the survey results indicating that students tended to perceive most of the library resources and services as important. However, for a specific resource, in this case, special collections and university archives, it is possible that only a specific group of students would perceive special collections and university archives as important due to their course-related work, while most other groups of students would not, resulting in misfit. Therefore, this item needs to be revised based on the Smith, Schumacker, and Bush’s study (1998).

The results of the last research question, - *Were the rating scale structures for subdomains appropriate for responses of students?,* showed that each category had a distinct peak at some point along the scale as expected for six different rating scales, suggesting effective use of rating scale category by respondents. One of the important revisions that the authors made on the instrument was reducing the number of the rating scale categories. A frequently raised concern in survey research is whether respondents use all the rating scale categories as intended (Wolfe & Smith, 2007)? A rating scale category that is not utilized does not serve the intended purpose, which is measuring construct severity in the instrument (Bradley, Cunningham, & Gilman, 2013). According to Linacre’s (2002) guidelines, the rating scale functioning of all six rating scales in the instrument was satisfactory. The revisions on the scale structured worked as intended and the respondents used all the sub-scales effectively. One problematic issue revealed from the analysis was person fit. One of the primary reasons for revising the initial instrument was implementing frequency scales enabling correlations between student’s utilization of the library’s resources and services with GPA, not possible with only yes or no responses. Revisions also increased the clarity of some questions. In addition, to increase the 2018 survey response rate, an incentive entering respondents names in a drawing for iPads, likely attracted more students to respond and complete the survey. Consequently perhaps, person misfit pattern in the data suggested careless and random responses by some students, implying that the survey developers or researchers should promote not only participation but also mindful responding behavior among students in future administrations of the survey.

This study contributes to the Library Information and Science field as an example of using the Rasch model for library assessment purpose and demonstrates several key points for future studies. The Rasch model is an effective approach for analyzing attitude surveys that consist of subdimensions and has been extensively used in educational and psychological assessment (e.g., Aksu Dunya et al., 2019; Brandt, 2008). Given that the current study aims to measure psychological constructs such as students’ perceptions of importance and satisfaction with library collection, materials and services, the findings confirmed the application of the Rasch model to library assessment was relevant and informative. Future analyses will need to test if items function differently by sub-groups such as undergraduate and graduate students, college programs, or Pell Grant recipients versus non grant recipients, as their library needs and use differ.

This study serves as an example for other institutions who want to develop and administer tools that will accurately assess library use experiences for college students. When possible, the best practice is to test the validity of assessment instruments. However, if an assessment instrument is not already available to meet an institution’s needs and an in-house instrument is created, content experts should be asked to review the instrument in addition to piloting the instrument with a small sample of the target audience, before use. For those who are interested in learning about Rasch techniques or applying Rasch model to evaluate their own instrument, please refer to Bond and Fox (2015) for guidance.

1. **Conclusion**

The findings from the current study showed that all 46 items were considered as a single construct: library experience. In addition, the items in the studied survey measure what it purports to measure, and the different rating categories were effectively used by respondents. The findings of the current study have implications for future research. Other academic libraries who are using their own instrument can test the validation of the instrument depending on their capacity, statistical knowledge, and skills. Utilizing a tested instrument is critical as it provides credible results and ensures the study’s accuracy. If it is not possible to test within the library, librarians can partner with other skilled researchers for validating the assessment tools within their institutions. Assuming that not all academic libraries have the capacity or skills to test the validity of their own instrument, this study provides libraries with an adaptable user survey that is validated by the Rasch model that is used extensively in educational and psychological assessment.

**References**

[citation withheld for blinding 1]

Aksu Dunya, B., McKown, C., & Smith, E. (2019). Psychometric properties and differential item functioning of a web-based assessment of children’s emotion recognition skill. *Journal of Psychoeducational Assessment*. https://doi.org/10.1177/0734282919881919

Anderson, L. (2016). Library website visits and enrollment trends. *Evidence Based Library and Information Practice*, *11*(1), 4–22. https://doi.org/10.18438/B8F918

Andrich, D. A. (1978). Rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

Atuase, D., & Koufie, C. (2017). A survey of extended library-hour service in Sam Jonah Library, University of Cape Coast. *Library Philosophy and Practice (e-journal)*, 1625.

Battisti, F. De, Nicolini, G., & Salini, S. (2010). The Rasch Model in customer satisfaction survey data. *Quality Technology & Quantitative Management*, *7*(1), 15–34. https://doi.org/10.1080/16843703.2010.11673216

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly: Management Information Systems*, *25*(3), 351–370. https://doi.org/10.2307/3250921

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York; London: Routledge, Taylor and Francis Group.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-04-0148

Bond, T. G. & Fox, C. M. (2015) *Applying the Rasch model: fundamental measurement in the human sciences*. Third edition. New York;: Routledge, Taylor and Francis Group.

Brandt, S. (2008). Estimation of a Rasch model including subdimensions. *IERI Monograph Series. Issues and Methodologies in Large-Scale Assessments, 1*, 51-70.

Cataldo, T., Shelton, T., Carrico, S., & Botero, C. (Eds.). (2015). *Proceedings of the Charleston Library Conference on How Users’ Perceptions of E‐Books Have Changed—Or Not: Comparing Parallel Survey Responses,* 304–310. https://doi.org/10.5703/1288284315587

Dennis, M., Greenwood, J., & Watson, A. (2013). LibQUAL revisited: Further analysis of qualitative and quantitative survey results at the University of Mississippi. *Journal of Academic Librarianship, 39*(6), 512–516. https://doi.org/10.1016/j.acalib.2013.08.009

Gunasekera, C. (2010). Students usage of an academic library: a user survey conducted at the main library University of Peradeniya. *Journal of the University Librarians Association of Sri Lanka, 14*(1), 43. https://doi.org/10.4038/jula.v14i1.2687

Hendriks, J., Fyfe, S., Styles, I., Skinner, S. R., Merriman, G., & Hendriks, J. (2012). Scale construction utilising the Rasch unidimensional measurement model: A measurement of adolescent attitudes towards abortion. *Australasian Medical Journal, 5*(5), 251–261. https://doi.org/10.4066/AMJ.2012.952

Hinchliffe, L. J. (2015). Osservatorio: Library assessment and user surveys in academic librarianship in the United States. *AIB Studi, 55*(2), 247–257. https://doi.org/10.2426/aibstudi-11198

Lai, K., & Chan, K. (2010). Do you know your music users’ needs? a library user survey that helps enhance a user-centered music collection. *Journal of Academic Librarianship, 36*(1), 63–69. https://doi.org/10.1016/j.acalib.2009.11.007

Law, E. L. C., & Van Schaik, P. (2010). Modelling user experience - An agenda for research and practice. *Interacting with Computers, 22*(5), 313–322. https://doi.org/10.1016/j.intcom.2010.04.006

Library & Information Science Research. (n.d). Retrieved February 10, 2020. from <https://www.journals.elsevier.com/library-and-information-science-research>

Linacre, J. M. (2002). *A user’s guide to Winsteps: Rasch model computer programs*. Chicago, IL: MESA Press.

Linacre, J. M. (2008). The expected value of a point-biserial (or similar) correlation. *Rasch Measurement Transactions, 22*(1), 1114. <https://www.rasch.org/rmt/rmt221e.htm>

Linacre, J. M. (2020). WINSTEPS (Version 5.4.1) [Computer software]. Beaverton, Retrieved January 1, 2019. from http://www.winsteps.com

Maughan, P. D. (1999). Library resources and services: a cross-disciplinary survey of faculty and graduate student use and satisfaction. *Journal of Academic Librarianship, 25*(5), 354–366. https://doi.org/10.1016/S0099-1333(99)80054-8

Mirza, M. S., & Mahmood, K. (2012). Electronic resources and services in Pakistani university libraries: A survey of users’ satisfaction. *International Information and Library Review, 44*(3), 123–131. https://doi.org/10.1016/j.iilr.2012.07.005

Montgomery, S. E. (2014). Library space assessment: user learning behaviors in the library. *Journal of Academic Librarianship, 40*(1), 70–75. <https://doi.org/10.1016/j.acalib.2013.11.003>

Müller, M. (2020). Item fit statistics for Rasch analysis: can we trust them? *Journal of Statistical Distributions and Applications*, *7*(1). https://doi.org/10.1186/s40488-020-00108-7

Natesan, P., & Aerts, X. (2016). Can library users distinguish between minimum, perceived, and desired levels of service quality? Validating LibQUAL+® using multitrait multimethod analysis. *Library and Information Science Research, 38*(1), 30–38. https://doi.org/10.1016/j.lisr.2016.01.004

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.

[citation withheld for blinding 2]

[citation withheld for blinding 3]

Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*, *3*(2), 119–120. https://doi.org/10.4300/jgme-d-11-00075.1

Shi, X., & Levy, S. (2005). A theory-guided approach to library services assessment. *College and Research Libraries, 66*(3), 266–277. https://doi.org/10.5860/crl.66.3.266

Smith, R. M., Schumacker, R. M., & Bush, M. J. (1998). Using item mean squares to evaluate fit to the Rasch Model. *Journal of Outcome Measurement, 2*(1), 66-78.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Erlbaum.

Thompson, B., Cook, C., & Health, F., (2000). The LibQual+ gap measurement model: the bad, the ugly, and the good of gap measurement, *Performance Measurement and Metrics,* *1*(3), 165-178, https://doi.org/10.1108/EUM0000000007216

Verma, S. (2016). Use of online databases in central science library, University of Delhi: A survey. *Journal of Library and Information Technology, 36*(2), 104–107. https://doi.org/10.14429/djlit.36.2.9409

Voorbij, H. (2012). The use of LibQUAL+ by European research libraries. *Performance Measurement and Metrics, 13*(3), 154–168. https://doi.org/10.1108/14678041211284704

Waugh, R. F. (2002). Creating a scale to measure motivation to achieve academically. *The British Journal of Educational Psychology, 72*, 65–86. doi:[10.1348/000709902158775](https://doi-org.proxy.cc.uic.edu/10.1348/000709902158775)

Wolfe, E. W., & Smith, E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: part II--validation activities. *Journal of Applied Measurement, 8*(2), 204-34.