Applying the Rasch Model to Students’ Perceptions of Leadership Abilities

Lauren Thomas
University of Kentucky
Abstract

The following analysis examines the quality of the Emerging Leader Institute course assessment. Students complete this assessment prior to beginning class and again following the final class. Winsteps Rasch measurement software is used to compare the difficulty of assessment items with student ability and evaluate how well the assessment measures students’ perceptions of their leadership ability. Results show that students agree or strongly agree with leadership statements at a rate of 71% and 95% on the pre and post assessments, respectively. Winsteps analysis revealed that only 2 items showed serious signs of poor fit, indicating that the course meets its learning outcomes and that the assessment accurately captures the change in students’ ability to endorse statements about perceived leadership ability.
Applying the Rasch Model to Students’ Perceptions of Leadership Abilities

Problem Statement

Much like academic departments, student affairs divisions have developed learning outcomes that track the growth of desired skills, often including communication, critical thinking, comfort with diverse populations, and other vital proficiencies. Since students are not often graded on or tested over these outcomes, course assessments and program evaluations play a crucial role in the measurement of these abilities. Analysis of these assessment results may lead to improved instruments as well as identifications of gaps in course content. This paper evaluates the quality of the Emerging Leader Institute (ELI) course assessment, a survey instrument that measures students’ perceptions of their leadership abilities, using Winsteps software to compare the distribution of the assessment results to the Rasch model.

Theoretical Framework

The Rasch model is a tool used to analyze assessment data in the social sciences, often latent traits such as attitudes or perceptions. The probability that a person will indicate a certain answer is used to develop estimates of item difficulty and person ability. The Rasch model requires that instruments measure a unidimensional construct, which means that an assessment may only focus on one trait or concept. Rasch modeling can also be used to compare changes in individuals’ assessment results over time and is not compromised by missing data, making it an effective method of evaluation for the ELI assessment. The polytomous rating scale model, shown in Figure 1, is used to evaluate Likert-type data generated by the ELI assessment.

Methodology

Instrument
The ELI course assessment is a survey instrument containing 35 Likert-type items that ask students to rate their level of agreement with a statement about their leadership ability. The same survey was administered to all course participants prior to the first class period and again at the conclusion of the semester. The pencil and paper survey consisted of statements based on the ELI course objectives and allowed students to select the response they felt best described their attitudes or behaviors. The response categories for items 1-25 included (1) strongly disagree, (2) disagree, (3) somewhat agree, (4) agree, and (5) strongly agree. The response categories for items 26-34 included (1) never, (2) rarely, (3) sometimes, (4) oftentimes, and (5) always. The final item, number 35, allowed students to indicate how frequently a statement applied to them, using (1) daily, (2) weekly, (3) monthly, (4) quarterly, and (5) yearly. Rating scales were printed at the top of each page of the survey.

The statements included in the assessment reflect the learning outcomes for leadership development as defined by the University of Kentucky’s Office of Student Involvement. Although the desired skills range from handling conflict to being aware of diversity, the amalgam of these characteristics creates a portrait of the Office’s concept of a well-developed, self-aware student leader. This is the unidimensional construct that the ELI assessment is intended to measure. Assessment items, listed in Figure 5, include “I am familiar with leadership concepts and theories,” “I consider ethical implications in my decision-making processes,” and “I actively pursue positive change in my community,” among others.

Participants

All course participants were first-year and sophomore students at the University of Kentucky. The data used in this analysis were collected from the 27 students who participated in the Spring 2008 section of ELI. Fifty-nine percent of the students in this section were female,
and 41% were male. Only one section of this highly selective course is offered each semester, and all participants are chosen from a pool of applicants. Selection criteria include grade point average, co-curricular involvement, leadership experience, and essay responses.

Procedures

To evaluate the ELI assessment, Winsteps Rasch measurement software (Linacre, 2010) was used to assess the fit of the ELI pre and post assessment data with the Rasch model. Data were analyzed using the polytomous rating scale model shown in Figure 1. Winsteps output revealed information related to item difficulty, person ability, rating scale structure, and construct validity of items. Items and people not fitting the model were indicated by Winsteps-generated fit statistics. Five separate Rasch analyses were performed with this particular data set. First, results from the full pre and post assessments were analyzed separately, followed by a “stacked” data set in which responses from the pre and post assessments were combined and analyzed as a whole. These tests indicated that several items in the latter portion of the assessment did not fit the Rasch model, which might be attributable to a change in rating scales after the first 25 items. To account for this potential bias, the first 25 items from both the pre and post assessments were separated into two new data sets and a third “stacked” data set, and Rasch analyses were performed on each.

Results

As previously indicated, students chose from a five-point rating scale when responding to the ELI assessment. A frequency summation of the response categories selected on both the pre and post assessments, shown in Table 1, reveals an increase in endorsed leadership statements between the administration of the two assessments. Students were most likely to agree or strongly agree with the leadership statements on both assessments, selecting one of these two
categories at a rate of 71% and 95% on the pre and post assessments, respectively. This is also reflected in the Winsteps statistical summary; the measure of students’ ability to endorse statements increases from 1.42 logits to 2.89 logits.

Winsteps also provided information about the raw scores, difficulty measures, and mean square fit statistics for each item. The range of measured item difficulty varies based on the total number of people in a sample who correctly answered a question, but average item difficulty is always zero. The mean square fit statistic shows the amount of distortion in a measurement system; its expected value is 1.0.

Item 35, which asked students to select how often they were faced with ethical decisions, was identified as the furthest outside of the acceptable range of predictability. In each of the Winsteps analyses conducted, infit mean square values for this item exceeded the threshold of acceptability, ranging from 2.39 to 3.18. Outfit mean square values for the item ranged from 2.67 to 4.91. Items 4 and 9 on the post assessment also showed signs of serious misfit. Outfit mean square values for item 4 (I understand that I am accountable to others.) ranged from 2.09 to 2.19, and outfit mean square values for item 9 (I have identified an issue in my community that I am passionate about [am willing to take action on].) ranged from 1.80 to 2.09. Outfit mean square values for items 5, 8, 12, and 19 showed slight signs of misfit, but none as severe as items 4, 9, and 35. Infit mean square values for items 13, 19, and 34 showed slight signs that the items did not fit the Rasch model but were not degrading to the construction of a measurement system.

Figures 2 and 3 show the item difficulty and person ability hierarchies for the pre and post assessments. Items 17 (I am comfortable engaging in workteams.), 4 (I understand that I am accountable to others.), 24 (I am comfortable interacting with people different than me.), and 16 (I am aware that people are different than me.) were among the easiest (difficulty measure
less than -1.25 logits) for students to endorse before participating in ELI. These items were joined by items 25 (I understand how to plan and implement a leadership project.), 13 (I recognize the impact of nonverbal communication.), and 2 (I understand the process of collaboration.) following students’ completion of the course.

On the pre assessment, difficult to endorse items (difficulty measure greater than 1.00 logits) included items 14 (I identify with and understand my leadership style, including my strengths and areas for improvement.), 12 (I feel comfortable working down in a hierarchy.), 3 (I am familiar with leadership concepts and theories.), and 25 (I understand how to plan and implement a leadership project.). This threshold of difficulty was constructed based on the relative difficulty measures of the 35 items. Few items included in the assessment measured beyond one standard deviation of the mean difficulty level. The post assessment results indicated that items 3 and 25 dropped below one logit of difficulty while the difficulty of items 19 (I manage my time well.) and 29 (I actively pursue positive change in my community.) rose above one logit.

**Discussion**

Overall, Winsteps analysis of the ELI assessment reveals that students in the sample are comfortable endorsing their leadership ability as defined by the learning outcomes for ELI and become even more comfortable doing so after completion of the course. Students endorsed at high rates statements related to understanding, respecting, and interacting with diverse populations; recognizing nonverbal communication; communicating effectively; and making ethical decisions. Relatively low difficulty measures for these items indicate that, on average, students are very comfortable with the broader concepts of communication, diversity and
inclusion, and ethics (disregarding the misfit item number 35). These items retain their low difficulty measures on the post assessment.

The ELI pre assessment indicates that prior to completing the course students are less comfortable endorsing their knowledge of leadership theory and styles, ability to identify their own strengths and passions, ability to work down in a hierarchy, and ability to use motivational techniques to support others. Students’ lack of comfort with leadership theory and styles may be partially explained by lack of prior exposure since ELI is often students’ first introduction to the concept of leadership as an academic discipline rather than a position or an action. To this end, the curriculum is designed to include lessons on the Social Change Model of leadership, the Leadership Challenge, and the DiSC® personal assessment tool. The post assessment results reveal a decrease in item difficulty for statements related to leadership theory, strengths and passions, and motivational techniques, suggesting that ELI influences development in these areas.

Furthermore, ELI is designed for first-years and sophomores, who are often still in the process of identity development. Thus, the course incorporates reflective components intended to help students understand their personal strengths and identify social issues they feel passionately about. The coupling of students’ lower level of comfort endorsing their ability to work downward in a hierarchy and use motivational techniques indicate that perhaps students are still developing a complete understanding of how to navigate leadership as a position of influence among their peers. Winsteps output does not show a clear change in item difficulty for item 6 (I utilize motivational techniques to support others toward a common purpose.). It remains unclear why there is no change in difficulty for this item, but possible reasons include unclear language
in the phrasing of the item or inadequate coverage of motivational techniques in the ELI curriculum.

Overall, student ability measurements exceeded item difficulty on both the pre and post assessments, as indicated by the placement of the frequency markers along the central axes in Figures 2 and 3. On both the pre and post assessment item-person fit maps, only one item has a measured difficulty that exceeds the mean person ability. The lack of items at a comparable difficulty level to students’ ability levels decreases the assessment’s accuracy in differentiating between able and highly able students and also in measuring the more challenging aspects of leadership development. Since difficulty measures are low in comparison to person ability, the probability that a given student would endorse a statement is high, contributing to an elevated raw score. This phenomenon frequently occurs when Rasch analysis is applied to survey data. However, the relative easiness of the items to the students’ ability suggests that the students have met the learning outcomes for leadership development set by the student affairs staff in the Office of Student Involvement.

Probability curves like the one shown in Figure 4 indicate that each response category (1-5) appears the most likely selection at a different level of item difficulty. Figure 4 depicts the probability curves for the results of the “stacked” assessment and includes all 35 items used on the ELI assessment. Probability curves for the post assessment data show that choices 1 and 3 do not peak as the most probable selections. However, since these choices do represent the most probable selection for a specified difficulty range on the pre assessment, their inclusion in the assessment rating scale is justified.

An examination of the labels chosen for the assessment response categories also reveals a slight risk of positively skewed data due to a slanted rating scale. The scale for the first 25 items
of the ELI assessment includes two variations of “disagree” options, (1) strongly disagree and (2) disagree, and three variations of “agree” options: (3) somewhat agree, (4) agree, and (5) strongly agree. Given the middle category, students’ selections may indicate that they endorse statements at an inflated rate, depending on whether they interpret the middle category as written or as a neutral category better described as “unsure” or “indifferent.”

From Figure 4, one can also infer that “somewhat agree” is the most probable selection for items with approximate difficulties between -1.0 and 0.0 logits of difficulty. Although this range represents a relatively small portion of the difficulty distribution shown in Figure 4, items at this difficulty level comprise approximately 25% of the ELI assessment. Given the concentration of items at this difficulty level, it would be advisable to conduct further research to learn how students interpret the middle category “somewhat agree.” Possible methods of investigation could include focus groups with students and repeated iterations of the ELI assessment with different labels assigned to the middle category.

The most serious issue with the ELI assessment stems from the change of rating scale category labels throughout the assessment. As previously noted, the rating scale options 1-5 do not indicate the same responses throughout the assessment. The drastic change in rating scale labels for item 35 likely caused it to lie outside of the acceptable range of the Rasch model. Since the response category labels were consistent (strongly disagree, disagree, somewhat agree, agree, and strongly agree) for the first 25 items and labels for the next nine items (never, rarely, sometimes, oftentimes, and always) were similar to the initial rating scale, one may conjecture that students disregarded the change in response categories on item 35, which associated very different qualitative labels (daily, weekly, monthly, quarterly, and yearly) with the same numerical rating scale. In order to develop a more consistent assessment and improve item fit,
survey items should be rephrased to allow for the use of a consistent rating scale and response
category labels throughout the assessment.

Of the misfit items on the ELI Assessment, more outfit mean square values exceeded the
threshold of acceptability than did infit mean squares. The outfit statistic typically indicates
unexpected observations on an item that should be easy or difficult relative to a person’s ability
while the infit statistic indicates unexpected behavior on an item that is close to the individual’s
target ability. Thus, the excess of high outfit statistics is preferable over an excess of high infit
statistics. One may speculate that students have made careless errors or lucky guesses on the
items that exhibit high outfit statistics. On the pre assessment, these errors and guesses may be
attributed to the fact that students were still unfamiliar with some of the leadership theories and
terminology used. Prior to the beginning of class, many students were unaccustomed to the
vocabulary commonly used in leadership education and student affairs. They may assimilate
some of the terminology throughout the semester, but this is contingent upon the instructor’s
consistency in using and explaining this language.

As previously discussed, items 35 and 4 exceeded the acceptable upper bound of 2.0 for
mean square value of outfit statistics. Although the misfit of item 35 is likely attributable to the
choice of rating scale labels, item 4 *(I understand that I am accountable to others.)* uses an
appropriate rating scale and set of labels. Since the rating scale does not seem to be a source of
trouble for this item, one may look next to the phrasing used in the item stem. Item difficulty
was measured at -1.52 logits on the pre assessment and -1.78 logits on the post assessment (both
measurements fall within two standard deviations of the mean difficulty measure and well below
the mean person ability level), which means that a portion of the students probably made careless
errors on or misinterpreted this item. A potential hypothesis would be that the item stem uses
vague phrasing, making it difficult to define how and to whom students are held accountable. Appropriate forms of mitigating this problem include testing samples for other ELI classes and re-phrasing the item to more clearly reflect what is being asked.

Items 5 (I am an effective communicator.) and 9 (I have identified an issue in my community that I am passionate about [am willing to take action on].) showed slight signs of misfit, although they did not fall outside of the acceptable range for mean square statistics. Both items fell within one standard deviation of the mean item difficulty value on both the pre and post assessments so they were not among the most difficult items on the assessment. Since the difficulty measures low relative to person ability, the unexpected responses likely resulted from careless error. Although further sampling and analysis would be required to determine the cause of the misfit on these items, a reasonable hypothesis would be that item 9 may be interpreted to measure a dual-pronged construct: identifying an issue and being willing to take action on it.

**Educational Impact**

Overall, the increase in students’ ability to endorse the leadership learning outcomes between the ELI pre and post assessments indicates that the course is meeting its objective of educating students about leadership theories and issues. The difference also suggests that ELI is helping them develop an awareness of themselves as leaders and in relation to other groups of people. Several of the ELI course assignments focus on students developing a synthesis of their core values, passions, and leadership beliefs. The course content may also help students more clearly articulate some of the ideas they are already familiar with, further contributing to the increase in endorsability on the post-assessment.

Along with identifying successful components of the ELI curriculum, the issues discovered in this analysis shed light on concepts that students may have difficulty understanding.
or endorsing. In particular, the items that have higher infit statistics reveal statements that are targeted to students’ ability level but not functioning as expected. Items with high infit statistics or relatively easy items that students have difficulty endorsing reveal areas of the curriculum that may need further clarification and coverage during instructional time.

The implementation of the recommended changes to the assessment tool and reinforcement of the identified weak areas in course content will inform change that has potential to improve both students’ educational experiences and instructors’ understanding of student learning throughout the course. Analyses of results from future sections of ELI will provide information about the continuous improvement of the ELI course assessment as an evaluation instrument and may eventually lead to the development of a calibrated survey instrument with the capability of measuring students’ ability to endorse their leadership abilities. The results of ELI assessments could be bolstered by administering the same assessment to a control group of students who do not participate in the course and using the two samples to compare development over the course of the semester. In the broader spectrum of student affairs, this process of analyzing assessment results to inform course design and improve evaluation tools may be transferred to other courses and programs in order to provide better data regarding students’ development in accordance with specified learning outcomes.
References

Figure 1

*Polytomous Rating Scale Model*

\[
\pi_{njk} = \frac{\exp\left(\sum_{j=0}^{r} [\beta_n - (\delta_{ij} + \tau_k)]\right)}{\sum_{k=0}^{m} \exp\left(\sum_{j=0}^{r} [\beta_n - (\delta_{ij} + \tau_k)]\right)}
\]

Table 1

*ELI Assessment Rating Scale Categories and Frequency Distribution*

<table>
<thead>
<tr>
<th>Category</th>
<th>Label</th>
<th>Pre Assessment Observed Count</th>
<th>Pre Assessment Percent</th>
<th>Post Assessment Observed Count</th>
<th>Post Assessment Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Strongly Disagree</td>
<td>7</td>
<td>1%</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>Disagree</td>
<td>46</td>
<td>6%</td>
<td>7</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Agreed</td>
<td>151</td>
<td>21%</td>
<td>26</td>
<td>4%</td>
</tr>
<tr>
<td>4</td>
<td>Agreed</td>
<td>322</td>
<td>45%</td>
<td>243</td>
<td>36%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Strongly Agreed</td>
<td>182</td>
<td>26%</td>
<td>398</td>
<td>59%</td>
</tr>
</tbody>
</table>
Figure 2

*ELI Pre Assessment Item-Person Fit Map*
Figure 3

*ELI Post Assessment Item-Person Fit Map*
Figure 4

*ELI Stacked Assessment Probability Curve*

Figure 5

*ELI Assessment Items*

Rating Scale – Items 1-25: 1-Strongly Disagree

2-Disagree

3-Somewhat Agree

4-Agree

5-Strongly Agree

1. I have gained effective communication skills outside of the classroom.

2. I understand the process of collaboration.

3. I am familiar with leadership concepts and theories.

4. I understand that I am accountable to others.
5. I am an effective communicator.
6. I utilize motivational techniques to support others toward a common purpose.
7. I consider ethical implications in my decision-making processes.
8. I respect others’ opinions even if they are different than my own.
9. I have identified an issue in my community that I am passionate about (am willing to take action on).
10. I am confident in my ability to handle conflict.
11. I feel comfortable working up in a hierarchy.
12. I feel comfortable working down in a hierarchy.
13. I recognize the impact of nonverbal communication.
14. I identify with and understand my leadership style, including my strengths and areas for improvement.
15. I understand how to “network” with others.
16. I am aware that people are different than me.
17. I am comfortable engaging in work teams.
18. I manage my time well.
19. I actively build and participate in diverse and inclusive communities.
20. I am comfortable with “networking” concepts.
21. I work with others toward a common purpose.
22. I am comfortable making an ethical decision in the face of opposition.
23. I am comfortable interacting with people different than me.
24. I understand how to plan and implement a leadership project.

Rating Scale – Items 26-34: 1-Never
2-Rarely
3-Sometimes
4-Oftentimes
5-Always

26. I interact with others to build a “network” that benefits my organization(s).

27. I apply leadership concepts and theories to my work in groups.

28. I make ethical decisions.

29. I actively pursue positive change in my community.

30. I collaborate.

31. I practice active listening.

32. I take responsibility for my actions.

33. I voice my disagreement when confronted with prejudicial and/or stereotypical statements.

34. I set improvement goals for my personal leadership development.

Rating Scale – Item 35: 1-Daily
2-Weekly
3-Monthly
4-Quarterly
5-Yearly

35. I am faced with ethical decisions…