

Evaluating and restructuring a new faculty survey: Measuring perceptions related to research,  
service, and teaching

Kelly D. Bradley<sup>1</sup>, Linda Worley, Jessica D. Cunningham, and Jeffery P. Bieber  
University of Kentucky

---

<sup>1</sup> Use Kelly D. Bradley as author of contact. (P) 859-257-4923 or (E) [kbrad2@uky.edu](mailto:kbrad2@uky.edu)

### Abstract

The New Faculty Preparation Survey was designed to measure perceptions of recently hired faculty regarding research, service and teaching for self and institution, while considering academic preparation. A Rasch measurement model was applied to the calibration sample responses to evaluate the psychometric properties of the instrument, providing a foundation for structural modifications. Rasch results reveal sound reliability; however, item misfit exists, coupled with limited utility of the full rating scale. Studying the measurement properties of the instrument will aid in proper interpretation of the results. Accurately measuring and understanding faculty perceptions across research, service and teaching can aid in professional development training at the university level, future faculty preparation, internal review process and general faculty satisfaction.

Evaluating and restructuring a new faculty survey: Measuring perceptions related to research, service, and teaching

“Exemplary manuscripts... use measures that have established psychometric merit, and they provide evidence of the reliability and validity of those measures. Such attributes are rarely evident in the higher education research literature” (Smart, 2005, p. 470).

### Objective

The Rasch measurement model was applied to the 1997 New Faculty Preparation Survey calibration sample data to evaluate the structure and utility of the instrument, which was designed to measure perceptions of recently hired faculty regarding research, service and teaching. The study provides a foundation for accurately operationalizing, measuring, and interpreting faculty perceptions across a variable termed ‘faculty load’, which includes research, service and teaching.

### Theoretical Framework

#### *Rasch Model*

For rating scales, a straightforward approach to item analysis, particularly for development, is the Rasch model (Wright & Masters, 1982). As with all item response models, the Rasch model assumes that an additive structure underlies the observed data, and that both participants and items can be arrayed on a continuum and that the items have equal discriminative power. Thus, Rasch modeling can provide estimates of a scale’s items, as well as the response patterns of the individuals completing the survey (Andrich, 1988; Krueger & Finger, 2001; Santor & Ramsey, 1998).

One of the fundamental benefits of Rasch measurement, specific to Survey Research, is it calibrates scales to determine the psychometric distance between response options. Raw scores are not measures, and thus must be converted to become true interval measures. The Rasch model converts raw scores to their natural logarithm and places them along a 'ruler', in an effort to construct a measure, not simply allocate an assigned number. The model also overcomes the assumption of equal importance by controlling for both persons and items on the same metric. Further, with regards to missing data, the Rasch model is all encompassing, using the existing response frame to model missing responses. Bradley and Sampson (2005) summarize, stating:

Whereas the classical model produces a descriptive summary based on statistical analysis, it is limited, if not absent, in the measurement capacity. ... Applications of the Rasch model allow the researcher to identify where possible misinterpretation occurs and which items do not appear to measure the construct of interest, while producing information about the structure of the rating scale and the degree to which each item contributes to the construct. Thus, it provides a mathematically sound alternative to traditional approaches to survey data analysis (p. 13).

As an additional testament to the Rasch model's strength, Curtis & Keeves (1999), Peck (2001), Waugh (1999) and Wright and Masters (1981) concur the Rasch model is the only item response theory model that adheres to the seven principles of true measurement: (1) each item should function as intended; (2) each item can be positioned on a common scale; (3) the scale should be an interval one; (4) each person can be located along the same common scale used for items; (5) the responses should form a valid response pattern for each item; (6) estimates of precision must be available for all scale measures; and (7) each item should retain its meaning and function across individuals and groups (Curtis & Keeves, 1999; Wright & Masters, 1981).

### *Faculty Perceptions*

Why college and university faculty behave the way they do is a topic of long-standing interest to observers of higher education. Because of the interest in faculty behavior, and, by extension, what influences their behavior, there has developed a substantial literature around what motivates faculty. At the center of motivation are perceptions: Faculty will behave in certain ways because they are motivated by what they perceive to be personally satisfying and rewarding and/or because they will be sufficiently rewarded by their organizations for engaging in particular behavior.

While an extensive review of the literature is beyond the scope of this paper, our work is rooted in the larger construct of person-environment fit. As this construct applies to faculty behavior and motivation we follow in the line of others such as Blackburn and Lawrence (1995), Fairweather (2005), and Bieber and Worley (2006). Albeit with different methodologies and different types of data, each of these works seeks to understand how faculty (or future faculty in the case of Bieber and Worley's work) perceive and respond to their (anticipated) work environments.

### Methods

The New Faculty Preparation Survey was originally designed in 1997 to measure perceptions of recently hired faculty regarding research, service and teaching for self and institution, while considering academic preparation. The original instrument consisted of 32 items with a requested reflection across three domains: Self, Institution, and Graduate Training. Participants endorsed items utilizing a 6-point rating scale. A Rasch measurement model was applied to investigate the psychometric properties of the instrument, providing a foundation for structural modifications.

### *Calibration Frame*

A full study was conducted in 1997 at the University of Kentucky resulting in 144 usable responses from new faculty, with the majority being employed full time (98%) in a tenure track position (95%), with 87% having obtained a Ph.D. Various degree areas were represented including: humanities, engineering, business, education, natural sciences and social sciences. Demographic information such as gender and race, as well as details about current position and graduate training experiences were collected on each respondent.

Important to new faculty perceptions is their estimated percentage of time devoted to research, service, and teaching at their current institution as well as time devoted in these areas during their graduate training. The average percentage of time reported to teaching and research in their faculty positions were relatively similar; however, the average percentage of time tipped heavily toward research compared to teaching in their graduate training. Respondents were also asked about experiences with specific activities while in graduate school. While many respondents reported some teaching experience during graduate training, more than half of respondents did not receive training in teaching. This information provides a general picture of the calibration frame used in the study.

### *Data Analyses*

Analysis took place in two phases. Coupling a quantitative critique with a foundational review of the items resulted in a psychometric evaluation of the instrument allowing for the restructuring of the original survey to produce an instrument capable of quantifying faculty values related to research, service, and teaching. Using WINSTEPS software version 3.65, a Rating Scale model was applied beginning with persons and items to test the overall fit of the data to the model. The basic mathematical expression used for constructing measures through

responses in this model is  $\log(P_{nik} / P_{ni(k-1)}) / B_n - D_{ik}$  (Andrich, 1978) where  $P_{nik}$  represents the probability the person  $n$  when responding to item  $i$  would be observed in category  $k$ ; similarly  $P_{ni(k-1)}$ ;  $B_n$  is the attitude of person  $n$ ;  $D_{ik}$  is the difficulty of item  $i$  with the impediment to being observed in category  $k$  relative to  $k-1$ .

Survey responses were analyzed in a unidimensional approach with a hypothesis that separation by the three domains – Self, Institution and Graduate Training – was likely. In some cases, individuals responded to some, but not all, of the items. When an individual had missing data, the data were treated as missing since it was reasonable to believe respondents might not be able to answer all survey items with integrity. Further, it was decided that imputing means or other substitutes for missing data would be inappropriate since missing data are not problematic with the Rasch model, might mute variability within the data set, and could result in data and information that was not truly reflective of the answers provided by the respondents.

To evaluate the reliability and validity of the instrument, survey items and respondents were investigated. Items and persons not adequately fitting the model requirements were identified using the mean square scores, with a reasonable range determined to be within one standard deviation of the average mean square fit statistic (Wright & Stone, 2004). Inspecting the outfit mean-squares provides evidence about the fit of the data to the model. The infit mean-squares are used to determine the fit of the item within the construct. Point-biserial correlations were inspected to investigate the orientation of the latent variable to ensure that the polarity of the items were of the same sign, or direction (i.e. all point-biserial correlations were positive). The guidelines outlined by Linacre (2004) were used to evaluate the rating scale category effectiveness based on the responses. The number of observations and distribution of observations across categories were examined to describe the functioning of the rating scale

categories. Advancing average measures with each category and step calibrations ensure the rating scale measure is stable and accurate. Probability curves were used to visually inspect the rating scale category function. This information was then summarized and a theoretical review of the structure and intent of measures took place.

### Findings

The Rasch model assumes the items measure a common construct, namely faculty load, which is measured in the areas of research, service and teaching. Although such analysis may suggest invariance at the item level, it may not necessarily imply construct equivalence (Allen & Walsh, 2000). That is, item similarity across groups may reflect similarities in item interpretation but the interpretation may be different than what the survey proposes to measure. As previously discussed, results are presented across the three frameworks of representation – Self, Institution, and Graduate Training. Results are presented in two-fold. First, the fit of the data to the model are presented followed by a general discuss of the operationalization and utility of the instrument specific to measuring faculty perceptions.

#### *The Model*

First, person and item reliability estimates are important in determining fit of the data to the model. The reliability and separation are presented for the items and respondents (labeled as person). Reliability has a theoretical range of 0.00 to 1.00. With 144 participants measured, estimates are sound, with a minimum reliability estimate of 0.87, person estimate, and a maximum of 0.98, item estimate. Separation estimates ranged from 2.60 to 7.42, with the lower values being associated with the person estimates. Presented below in Table 1 are reliability and separation estimates across all frameworks.

Table 1. Reliability and separation separated by framework

	<u>Person Reliability</u>	<u>Person Separation</u>	<u>Item Reliability</u>	<u>Item Separation</u>
Self	0.87	2.60	0.97	5.34
Institution	0.94	3.83	0.98	6.95
Graduate Training	0.93	3.75	0.98	7.42

Estimates are most consistent for the frameworks of Institution and Graduate Training. The item separation estimates in these categories indicate that there are at least three distinguishable groups of thoughts for the respondents, as both person separation estimates are above 3. More so, across all frameworks, items are able to be grouped into five thresholds, as represented with item separation estimates above 5.

Item response theory, specifically here, Rasch modeling, is often used to determine whether respondents are utilizing the full rating scale. Based on the empirical evidence provided in this study, it appears the respondents are not consistently utilizing the rating scale. It is critical to consider the intended purpose of the measure and to verify that the survey is indeed capturing the variety of perceptions. Essential guidelines to meet for measure stability and accuracy, or fit, include the following (Linacre, 2004): items oriented with latent variable; each category contains at least ten observations; average measures advance monotonically within categories; and outfit mean squares less than 2.0. Certain guidelines are helpful, but not necessary for stable and accurate rating scale measures: observations display a regular distribution; ratings imply measures and measures imply ratings; and step difficulties advance by less than 5.0.

With these considerations in mind, the evaluation continues with a review of the fit statistics. Acceptable ranges were calculated using a confidence interval approach, based on the

mean estimate and standard errors across each framework. The acceptable ranges are as follows: Self – Infit (0.80, 1.34) and Outfit (0.74, 1.30); Institution – Infit (0.45, 1.69) and Outfit (0.36, 1.94); Graduate Training – Infit (0.71, 1.35) and Outfit (0.69, 1.35). Table 2 illustrates the fit statistics, and those falling outside the acceptable range are indicated by asterisks (\*).

Table 2. Item measure and fit statistics separated by framework

	<u>Framework</u>	<u>Item Measure</u>	<u>Infit MNSQ</u>	<u>Outfit MNSQ</u>
1. teach well	Self	-.91	.88	.69
	Institution	-.22	.78	.76
	Graduate Training	-.16	.84	.80
2. conduct research	Self	-.39	1.56*	1.43*
	Institution	-1.62	2.65*	3.31*
	Graduate Training	-1.66	1.17	1.09
3. do department/university committee work	Self	1.03	.84	.95
	Institution	.05	1.09	1.12
	Graduate Training	.69	1.08	1.12
4. assess student work fairly	Self	-1.32	1.10	.82
	Institution	-.12	.64	.64
	Graduate Training	-.14	.69	.68
5. experiment w/ innovative teaching strategies	Self	.14	.84	.91
	Institution	.44	.84	.84
	Graduate Training	.39	.81	.82
6. direct teaching assistants	Self	1.03	1.23	1.27
	Institution	.79	1.14	1.11
	Graduate Training	.63	.85	.81
7. be a mentor to others	Self	-.03	.81	.82
	Institution	.27	.79	.79
	Graduate Training	.01	1.01	.97
8. publish	Self	-.43	1.62*	1.39*
	Institution	-1.87	2.73*	3.66*
	Graduate Training	-1.25	1.37*	1.43*
9. do community service	Self	.67	.94	1.00
	Institution	.81	1.10	1.43
	Graduate Training	.96	1.27	1.26
10. advise students	Self	.23	.87	.80
	Institution	.15	.65	.64
	Graduate Training	.53	.85	.82

11. provide feedback to help students improve	Self	-.42	.70	.66
	Institution	.24	.45	.44
	Graduate Training	.07	.65	.65
12. lecture well	Self	-.52	1.38*	1.69*
	Institution	-.16	.64	.62
	Graduate Training	-.03	.96	.96
13. construct valid and reliable exams	Self	-.39	1.00	1.12
	Institution	.12	.49	.49
	Graduate Training	.05	.75	.71
14. interact and connect with students	Self	.51	1.15	1.25
	Institution	.93	.81	.79
	Graduate Training	.19	1.01	.97
15. apply for grants	Self	.27	1.52*	1.43*
	Institution	-1.05	2.49*	2.59*
	Graduate Training	.19	1.95*	2.04*
16. create a teaching portfolio	Self	1.01	1.16	1.18
	Institution	-.50	1.29	1.23
	Graduate Training	1.32	1.01	.78
17. provide timely feedback to students	Self	-.23	.81	.71
	Institution	.17	.64	.63
	Graduate Training	.06	.70	.70
18. use multimedia effectively	Self	.63	1.04	1.06
	Institution	.62	.81	.80
	Graduate Training	.71	1.06	1.15
19. take into consideration student differences	Self	.08	.75	.78
	Institution	.35	.52	.53
	Graduate Training	.26	.93	.89
20. conduct interdisciplinary research	Self	.34	1.27	1.25
	Institution	.14	1.64	1.90
	Graduate Training	-.35	1.56*	1.51*
21. engage classes in effective discussions	Self	-.34	1.09	.87
	Institution	.40	.64	.66
	Graduate Training	.02	.89	.95
22. initiate new research directions	Self	-.13	1.52*	1.41*
	Institution	-.57	1.83*	1.88
	Graduate Training	-.79	1.42*	1.44*
23. interact effectively with colleagues	Self	-.49	.90	.85
	Institution	-.73	1.10	1.12
	Graduate Training	-.70	.90	.98
24. construct a thoughtful syllabus	Self	-.16	1.06	.87
	Institution	.12	.60	.61
	Graduate Training	.15	.89	.91

25. present research in professional settings	Self	-.44	1.55*	1.34*
	Institution	-1.15	1.68	1.88
	Graduate Training	-1.27	1.50*	1.43*
26. understand student abilities and preparation	Self	.05	.65	.63
	Institution	.34	.56	.55
	Graduate Training	.24	.57	.58
27. direct student research (all levels)	Self	.01	1.04	1.05
	Institution	-.21	1.29	1.33
	Graduate Training	-.11	1.52*	1.48*
28. develop learning goals	Self	.42	.67	.62
	Institution	.32	.64	.65
	Graduate Training	.18	.87	.89
29. interact w/ students inside of class	Self	-.55	1.08	.82
	Institution	.23	.63	.62
	Graduate Training	-.15	.69	.68
30. organize group work in class	Self	.39	1.04	1.12
	Institution	.70	.66	.66
	Graduate Training	.34	.79	.80
31. develop your own teaching style	Self	-.62	.93	.69
	Institution	.48	.81	.81
	Graduate Training	-.07	.90	.89
32. connect with a mentor	Self	.55	1.17	1.19
	Institution	.54	1.56	1.59
	Graduate Training	-.31	1.51*	1.48*

Note: Fit statistics over the acceptable range are indicated by asterisks (\*).

Three items – 8. *publish*, 15. *apply for grants*, and 22. *initiate new research directions*, misfit the model across all frameworks – Self, Institution, and Graduate Training. Items 2. *conducting research* and 25. *present research in professional settings*, misfit for two of three frameworks. Even though these items are misfitting the model, it is believed that the content of these items is driving the misfit, the different ideologies of research and the connections to funding and dissemination. For these items, extreme responses are more prevalent, meaning ratings of 1, 2, 5, and 6 – the lowest and highest ratings available to the participants, as compared to other items on the survey. Even more, it appears the strong impressions that respondents have could be linked to

discipline, on one or many of the frameworks – self, institution, or graduate training. It is likely that differential item functioning is present; however, given the misfit, it was decided not to conduct an empirical investigation of this theory. That being said, further studies will consider this aspect by seeing that it is measured in demographics and explored in analysis.

Finally, the six rating categories were reviewed in terms of utility, in the scaffold of the empirical evidence. Ratings ‘2’ and ‘3’ do not peak in the Self framework, indicating these selections are underutilized and not contributing to the measure. All ratings, with the exception of ‘5’, peak in the Institution framework, where ‘4’ and ‘5’ seem to be collapsing into one measure. In the Graduate Training framework, all ratings peak except for ‘2’. The six-point scale was constructed to allow participants a full range of selection. Nonetheless, the calibration response frame demonstrates that this level of sensitivity is not necessary. Due to this finding, a reduction from a 6-point scale to a 4-point response frame is being considered.

#### *Utility of the Instrument*

Reorganization of the demographic items is underway. Items are separated into personal demographic items and institution related items to provide a more logical sequence. By using a web-based format, many of the demographic issues, such as redundancy and non-applicable questions, will be resolved. In this web-based platform filter questions and limited response frames will be applied as appropriate. Branching is more easily employed in the web-based format. For the body of the survey, framework will be presented in a manner so that the respondent can consider all statements in the context of a single framework.

The principal contrasts did not reveal any evident redundancy among the survey items. Obvious contrasts were present, such as research-related items versus teaching-related items. Specifically, each item was reviewed for domain fit by intention and stability of measure. For

example, the item 'advise students' was connected to the research, service, and teaching domains. More so, it was determined to be vague in the sense that faculty could have multiple perceptions of the term 'advise'. A similar content analysis took place for each Likert-type item, using a 6-point rating scale.

Given the misfit of various items, but the apparent connection to the literature, researchers revisited the survey with the intent to separate items into the original domains – Research, Service and Teaching. During this process, an additional domain was identified that is being considered for further use – Collegial Interactions. Table 3 presents the researchers' intent on how original items are distributed across domains.

Table 3. Researchers' framework for item distribution

<u>Item #</u>	<u>Item</u>	<u>Teaching</u>
1	Teach well	
4	Teach/assess student work fairly**	
5	Experiment with innovative teaching strategies	
10	Advise students	
11	Provide feedback designed to help students improve	
12	Lecture well	
13	Construct valid and reliable exams	
14	Interact and connect with students outside of class	
16	Create a teaching portfolio	
17	Provide timely feedback to students	

- 18 Use multimedia effectively
- 19 Take into consideration student differences when teaching
- 21 Engage classes in effective discussions
- 24 Construct a thoughtful syllabus
- 26 Understand student abilities and preparation
- 27 Direct student research
- 28 Develop learning goals
- 29 Interact with students inside of class
- 30 Organize group work in class
- 31 Develop your own teaching style

Research

- 2 Conduct research
- 8 Publish
- 15 Apply for grants
- 20 Conduct interdisciplinary research
- 22 Initiate new research directions
- 25 Present research in professional settings

Service

- 3 Do department/university committee work
- 6 Direct teaching assistants
- 9 Do community service

### Collegial Interactions

- 7 Be a mentor to others
- 23 Interact effectively with colleagues
- 32 Connect with a mentor

Reviewing this structure, coupled with the empirical findings, it is clear that ‘wording’ needs to be cleaned throughout. For example, item 24. *construct a thoughtful syllabus*. The word *thoughtful* is ambiguous, so it was removed from the final survey. Another example, item 27. *direct student research*, needs to have clarification. Given the range of duties faculty members have, it is important to assign a level to this, i.e. undergraduate. The relevancy of item 16. *create a teaching portfolio* was questioned, as this is a field specific requirement. Thus, it was not included in the final version. The final version is being finalized and will be redistributed and recalibrated in Fall 2009.

### Conclusion

Studying the measurement properties of this *New Faculty Survey* will aid in correct applications of the instrument and proper interpretation of the results reported. Faculty serve as a central component of the academic institution. New faculty bring to the position multiple influences, their personal ideals, academic training, and institutional framework, to name a few. Understanding faculty perceptions across these influences for the typical domains – research, service, and teaching can aid in professional development training at the university level, future faculty preparation, internal review process, and general faculty satisfaction. In addition to the utility of the content of the survey, this study has outlined the process to construct a reliable and valid measure, with credible psychometric properties.

## References

- Allen, J., & Walsh, J. R. (2000). A construct-based approach to equivalence: Methodologies for cross-cultural and multicultural personality assessment research. In R. Dana, *Handbook of multicultural/cross-cultural personality assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika*, *43*, 561-573.
- Andrich, D. (1988). *Rasch models for measurement*. Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-068. Beverly Hills: Sage Publications.
- Bieber, J. P. and Worley, L. K. (2006). Conceptualizing the academic life. *Journal of Higher Education*, *77*, 1009-1035.
- Blackburn, R. T. and Lawrence, J. H. (1995). *Faculty at work: Motivation, expectation, satisfaction*. Baltimore, MD: Johns Hopkins University Press.
- Fairweather, J. S. (2005). Beyond the rhetoric: Trends in the relative value of teaching and research in faculty salaries. *Journal of Higher Education*, *76*, 401-422.
- Krueger, R. F., & Finger, M. S. (2001). Using item response theory to understand comorbidity among anxiety and unipolar mood disorders. *Psychological Assessment*, *13*, 140-151.
- Linacre, J. M. (2004). Optimizing Rating Scale Effectiveness. In E.V. Smith & R. M. Smith, *Introduction to Rasch Measurement* (pp.258-278). Maple Grove, MN: JAM Press.
- Santor, D. A., & Ramsay, J. O. (1998). Progress in the technology of measurement: applications of item response models. *Psychological Assessment*, *10*, 345-359.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.