Calibrating Charisma: The many-facet Rasch model for leader measurement and automated coaching

Matt Barney, Ph.D.
LeaderAmp, 1091 Crystal Springs Drive, Vacaville, CA 95688
matt@leaderamp.com

Abstract. No one is a leader unless others follow. Consequently, leadership is fundamentally a social judgment construct, and may be best measured via a Many Facet Rasch Model designed for this purpose. Uniquely, the MFRM allows for objective, accurate and precise estimation of leader attributes, along with identification of rater biases and other distortions of the available information. This presentation will outline a mobile computer-adaptive measurement system that measures and develops charisma, among others. Uniquely, the approach calibrates and mass-personalizes artificially intelligent, Rasch-calibrated electronic coaching that is neither too hard nor too easy but “just right” to help each unique leader develop improved charisma.

1. Influential Instruments
It is not an overstatement to suggest that leaders architect the fate of homo sapiens. By way of persuasion and teamwork, leaders help bring the future to the present. They found important organizations like IMEKO, build civilizations, and inspire the creation of spacecraft that may land a human on Mars in 2030 (e.g. SpaceX). They are often the glue that instrumentally hold our social bonds together, and give our teams direction and purpose. And, when they are evil, they can destroy our species and others by the millions.

For too long, however, the attributes and behaviors of leadership have been managed unscientifically. Recent reviews of the vast, inter-disciplinary literature on charisma suggest that poor measurement is a key culprit accounting for the fact that in spite of decades of attempts to study leadership in a wide range of social science disciplines, leadership theory is stuck in a quagmire. Charisma, as a subset of leadership shares the same fate, in part, this is because the original scholars - Aristotle and Weber failed to be specific enough about what charisma actually is to provide a blueprint for instrument construction. Earlier attempts at defining charisma were muddied by vague and multivariate definitions of charisma that attempted to mix values, emotions, personality, behaviors, processes, attributions, and even outcomes into one unidimensional latent trait. Unlike other types of psychological measurement, the fact that followers are the ultimate judges of whether they find leader behaviors inspiring is a special challenge for traditional applications of Rasch Measurement.

This is consistent with contemporary definitions of charisma inspired by economic theory that suggest leader verbal and non-verbal behaviors are what signal follower attributions and labels of being “charismatic”. In particular, research with this cleaner conceptualization of charisma suggests that leaders behave charismatically when they communicate shared values in vivid ways, through storytelling, metaphors, and symbolism, including emotional displays. This behavioral definition also has the benefit of being concrete for skill development, and emerging evidence from randomized controlled experiments has shown that charisma, defined behaviorally, can be taught, and has improved outcomes for followers.

Fortunately, in 1989, J. Michael Linacre reformulated Fischer’s
1973 linear logistic test model into a Rasch Measurement approach engineered for just this type of situation.2,3

2. Many-Facet Rasch Model for Leaders
The Many-Facet Rasch Model (MFRM) is foundational for the objective measurement of leaders, because it extends the Rasch family of models to include facets for severity/leniency of judges and other facets that could affect a rating situation. Because it conforms to the same objective standards as the rest of the Rasch measurement methods, it allows the estimation and removal of elements that distort measurement in a myriad of other ways, not limited to severity/leniency bias. The MFRM has been shown to be robust to many forms of misfit to the Rasch model, so small perturbations distort the ultimate result in trivial, unimportant ways12. The MFRM has been used to measure leader charisma with computer-adaptive assessments that are short, precise and based on the latest science4,5,6.

2.1. A Many Facet Rasch Model for leadership
A mobile computer-adaptive measurement, coaching, and journaling platform uses a four-facet MFRM:

\[
\log \left( \frac{P_{nij}}{P_{nij-1}} \right) = B_n - D_i - R_q - F_j
\]

where

- \( P_{nij} \) = probability of observing category j when rater q response to item i for leader n
- \( P_{nij-1} \) = probability of observing category j-1 when rater q response to item i for leader n
- \( B_n \) = attribute of leader n
- \( D_i \) = difficulty of item i
- \( R_q \) = difficulty of rater q
- \( F_j \) = Rasch-Andrich threshold, the point of equal probability on the latent variable between categories j-1 and j

In earlier research, this model was applied such that an overlapping sample of raters judged the same set of leaders charismatic behaviors\(^1\), as was originally intended for the MFRM\(^2\). But this was done in a large multinational corporation where the chairman directed it to be applied this way, which is rare. More commonly, different raters judge the leadership of entirely different, non-overlapping groups of people, making the classical MFRM approach impractical for many leader measurement situations. When the judges and leaders do not overlap, their measurement estimates are in different frames of reference, and are therefore uninterpretable. This situation is common in organizations, given that raters may switch teams, resign, go on maternity leave, become busy, or otherwise be unable or unwilling to rate a given leader in a particular timeframe. To make the measurement system more robust to these likely situations, and allow for objective comparisons of quarterly developmental results, an anchored, computer-adaptive approach was adopted.

2.1.1. Computer-Adaptive Measurement. Computer-adaptive measurement uses pre-calibrated item banks to measure individuals. It starts out with a question, typically in the middle of the scale, and depending on the users prior ratings, adjusts the next questions based on the new estimate of their location on the line. The algorithm stops when one or more stopping rules are met, such as having a trivial amount of uncertainty (measurement error), or after a certain number of questions have been asked. The major benefits to computer-adaptive approaches are that they are up to 90% shorter, more precise, and are harder to misappropriate than traditional approaches where everyone is asked the same questions\(^7\). In situations where the most expensive, and time-constrained human resources are involved, computer-adaptive measurements are especially practical. In particular, the brevity of
computer-adaptive measures is especially crucial to be accepted by the senior-most leaders of an organization.  

2.1.2 Anchored Facets. The current approach uses a software platform that combines the MFRM and Computer-Adaptive methods to make a smartphone-based system convenient, and useful for the likely situations where leaders are rated by different people at different times. Just as item difficulties and Rasch-Andrich thresholds are already known apriori to closely approximate the Rasch Model, the system also first uses a computer-adaptive measure to calibrate raters. All raters to rate a short vignette about a leader, and rate him/her to also estimate their severity/leniency bias. In this way, the system is able to concatenate all raw data from raters in order to estimate the leader attribute measure.

2.1.3 Calibrated Coaching. One traditional challenge in developing leaders is helping them develop on their jobs. Usually, people get a multi-page report that explains the result and recommendations for improvement. But as with any skill, it is easy to forget to practice communicating with metaphors, stories, and moral values so that a leaders charisma improves in the eyes of followers. Developmental Psychological literature suggests that it is especially important that individuals practice charismatic tasks that are neither too easy, nor too hard, but just right for their proficiency level.
To help leaders develop their charisma, based on their baseline, calibrated feedback content can be used to coach people through their smartphones to practice on the days and times that they desire. Each leader uses a smartphone graphical user interface, to schedule mornings they want personalized automated suggestions for practicing their charismatic behaviors. They also schedule a time later that evening when they reflect upon that deliberate practice, and journal about it in a mobile journal.

To make sure each person’s artificially intelligent coaching is “just right” for their current level of proficiency, the MFRM is used to calibrate two and three-sentence “eCoaching” statements for each level of the latent trait, using a construct map\(^9\), the latest domain-specific content, and science around persuasion\(^10\) and delivering feedback\(^6\). For example, signifying the importance of a team’s work is quite easy for leaders to do more frequently, if they’re already seen as low in charisma, this is a good place for them to start to practice, more frequently. They’re more likely to develop by mastering these easy tasks before they attempt more complex charismatic behaviors such as modulating the intonation of his/her voice to emphasize key points\(^6,11\). Only very highly charismatic leaders should practice behaviors that are exceptionally difficult, such as using alliteration and metaphors appropriately, with fervor. These are typically only done extremely well by famous leaders such as Martin Luther King Jr., JFK, and Steve Jobs who were considered especially charismatic by their followers because of their exceptional intuitive mastery of charismatic behaviors.

Table 1 shows a sample of eCoaching calibrations. Uniquely, this approach can be used with very senior and famous leaders, such as Elon Musk, who can give his verbatim eCoaching statements that can be calibrated with the MFRM, and mass-personalized to all SpaceX, Solar City and Tesla employees who work for him. In this way, the dual MFRM calibrations allow personalization of advice in each unique person’s zone of proximal development (“Goldilocks Zone”) from highly credible expert leaders whose time is scarce. While we have not yet tested the motivational utility of calibrating verbatim scripts from famous or highly credible leaders, it seems likely to be welcomed by prospective leaders, and worthy of future empirical tests.

**Table 1: Example MFRM Calibrated Coaching Statements for Mass-Personalized eCoaching**

| logits | eCoaching Push Notifications                                                                 | Mean-Square Infit | Mean-Square Outfit | Expected Point-Measure Correlation | Actual Point-Measure Correlation |
|--------|-----------------------------------------------------------------------------------------------|-------------------|--------------------|-----------------------------------|---------------------------------|
| -2.18  | Today, underscore the importance of your team’s work. What can you do today to put the value you all create in the top of your team’s mind? | 1.26              | 1.26               | 0.61                              | 0.74                            |
| -0.66  | What can you do to express your true emotions about how you feel about the future? Make an extra effort to talk about your emotions in an impassioned way today, as you work with your team. | 0.76              | 0.74               | 0.73                              | 0.74                            |
Summary

The leader assessment and development industry is worth over USD20B in North America alone, but rarely uses engineering-worthy measurements on charisma or any other attribute. Rasch-measurement approaches to developing a leader’s charisma to inspire his/her team is one subset of examples that can make a big difference to transform the workplace. This is acutely true when objective measurement is combined with technology that can scale to millions or even billions of people. The key is to ensure that the measurements and development are tailored to each unique persons’ proficiency level, to mass-personalize development. Because mobile devices are ubiquitous, even in the developing world, mobile measurement and development systems present an improved path toward scaling evidence-based management development.

References

1. Barney M 2010 Leadership @ Infosys (Penguin India)
2. Linacre J M 1989 Many-facet Rasch Measurement (Chicago: MESA Press)
3. Fischer G H 1973 Acta Psychologica 37 359-374
4. Day D V & Barney M 2012 ed W H Mobley, Y Wang, & M Li, Advances in Global Leadership (Vol. 7) (London, UK: Emerald)
5. Barney M 2013. Leading Value Creation: Organizational Science, Bioinspiration and the Cue See Model (New York: Palgrave)
6. Barney M 2015 Amplifying Leaders: Using Quantified Self Technologies to Grow Leaders like Olympians (CA: LeaderAmp Press)
7. Barney M 2012 Ed. J McHenry & M McCall. Leaders Building Leaders: How Senior Leaders Develop Their Protégés. (Soc. for I/O Psych, San Diego, CA)
8. Barney M, & Fisher W 2016 Ann Rev Org Psychol Org Behav 3
9. Wilson M 2005 Constructing Measures: An Item Response Modeling Approach (New York: Routledge)
10. Cialdini R 2008 Influence: Science and Practice (New York: Prentice Hall)
11. Antonakis J, Fenley M, Liechti S 2011 Academy Mgmt L & E 10 374-396
12. Linacre, JM 2016. Brief Explanation of the Theory behind Many-Facets Rasch Measurement. Facet manual, http://www.winsteps.com/facetman/theory.htm
13. Antonakis J, Bastardoz N, Jacquart P, Shamir B 2016. Charisma: an ill defined and ill-measured gift. An. Rev. of OP & OB, 3, 293-319.
14. Bass BM 1985. Leadership and Performance Beyond Expectations. NY:Free Press
15. Bass BM, Avolio BJ 1995. MLQ Multifactor Leadership Questionnaire. Redwood City, CA: Mindgarden.
16. Antonakis J, Fenley M, Liechti S. 2011. Can charisma be taught? Tests of two interventions. Acad. Manag. Learn. Educ. 10:374–96
17. 