American Journal of Economics and Business Administration

Original Research Paper

A Primer on the One-Parameter Rasch Model

Randy McCamey

College of Business Administration, Tarleton State University, PO Box T-0330, Tarleton Station, TX 76402, USA

Article history
Received: 16-12-2014
Revised: 20-12-2014
Accepted: 31-12-2014

Abstract: The Rasch measurement model improves on traditional test construction by creating tests in which the person's ability is independent of the sample of items used and the norm group used to calibrate the test. This article is an introductory review of the Rasch Model and describes properties of the Item Characteristic Curve (ICC) and discusses the utility of having person ability and item difficulty on a common scale.

Keywords: Item Response Theory, Rasch, Test Construction, Item Difficulty

Introduction

Consider the following scenario: An employment test is created and is then standardized using the brightest, most intelligent workers in the company. In subsequent hiring situations, job applicants are tested and almost none pass the test, thus they are not hired. As a consequence, position's go unfilled, production falls behind, existing employees work overtime regularly, morale and productivity suffer and the company loses money. This scenario highlights one of the criticisms of traditional test construction-tests must be standardized, or "norm'ed", correctly for the population being tested. If not and the test is used outside the norm group parameters, the test results may be invalid and thus any decisions made using those test results become questionable. In specific hiring and selection processes, this scenario could also create, however unintended, adverse impact to one or more protected populations. Another criticism of traditional test construction dependent on norm groups is that even if the test were standardized correctly, many populations change over time and thus old norms can become invalid for current applications.

Beyond the requirements of a norm group to standardize tests using traditional test construction methods, tests often require a large number of items to measure a person's ability. If a method could be devised to provide a better assessment of test items such that items measuring the same ability level could be eliminated, then the tests themselves could be shortened and test takers would be less likely to suffer from "test fatigue" as a result. A recent study by Shu'aibu *et al*. (2013) uses Rasch to identify the likelihood of redundant items is their questionnaire which could lead to fewer item responses required by respondents.

Rasch Measurement

One modern model of measurement used in the social sciences is the 1-parameter Item Response Theory (IRT) model. Georg Rasch, a Danish mathematician, had an interest in teaching statistics and in measurement models, in particular the IRT models. During the 1960's, Rasch developed his now-famous 1-parameter logistic model (the Rasch Model) to estimate a person's trait level from their responses to test items (Embretson and Reise, 2013). Although the Rasch Model and the 1-parameter IRT model use different algorithms for calculations, the results are virtually identical.

The Rasch Model, as with IRT models in general, promised to overcome weaknesses in Classical Test Theory (CTT). Specifically, IRT promises to overcome circular dependency of CTT which is the situation, as described by Fan (1998), where the person statistic is item dependent and the item statistic is examinee (person) dependent. The Rasch Model improves on traditional test construction in the sense that Rasch creates item-free and person-free tests. That is, the Rasch Model allows tests to be constructed where the measure of a person's ability is independent of the sample of items used and is independent of the norm group used to "calibrate" the test (Hashway, 1978). In the simplest form, a person's response to an item is the dependent variable in the Rasch Model and the independent variables are the person's trait score (theta or θ) and the item difficulty (b).

The Rasch Model can be used for measurement (i.e., locating a person on the latent continuum) or exploratory data analysis (i.e., understanding the structure of items or selecting a useful subset of items). The Rasch Model permits identification of items or behaviors that are ordered (e.g., what are the sequence of skills one needs to become a computer programmer) and thus the variable unit measure has the same meaning across the scale (Andrich, 1988). IRT modeling also allows statistical adjustments in scores and thus the development of more meaningful comparisons (Hambleton, 2000).

DOI: 10.3844/ajebasp.2014.159.163

Using item response theory allows two distinct advantages over simple classical test theory. First, it allows researchers to more accurately rank respondents in terms of their patterns of responses (Crocker and Algina, 1986; Hambleton, 1983). Although some researchers have argued that IRT does not produce scores necessarily different from classical test theory, IRT is maximized at the tails of the distribution (Fan, 1998). This is an important consideration when working with individuals who tend to score at either extreme of a distribution. Second, using IRT estimates allows for the generalization of scores to both the population of interest and to future users, whereas classical test theory results will not generalize to future users.

One of the practical applications of IRT modeling is to diagnose test instruments (i.e., item or test analysis). Table 1 lists the partial output of a Rasch analysis of a graduate level mid-term exam (n = 39) using the RASCAL for Windows (1995) software by Assessment Systems Corporation. It should be noted that IRT methods require much larger data sets, however this data set is introduced for heuristic purposes.

Item difficulty (*b*) is the main parameter of interest in the Rasch Model and is defined as the position on the latent trait variable where it is expected the person has a 50% probability of answering the item correctly. Note that item numbers 16, 7 and 13 all have the same item difficulty. Also note that there is a substantial difference between item difficulty for questions 16, 7 and 13 (-2.740) and items of the next higher item difficulty, items 15 and 19 (-2.028). Having this knowledge allows the researcher to modify one or more of items 16, 7 and 13 to fill in the item difficulty gap between –2.740 and –2.028 if so desired. If the researcher is confident that the range of person abilities is being adequately measured by the test, there is evidence through this analysis to remove 2 of these 3 items (16, 7 and 3) with the same item difficulty. This allows the instructor or researcher to measure the same range of person abilities using a single item at the difficulty level of -2.740. Item analysis can be continued in this example as items 15 and 19 also have the same item difficulty (-2.028), as do items 2, 10 and 24 (-1.269). As this example shows, IRT modeling software can provide a convenient method for researchers to optimize both the number and difficulty of items on a test or assessment.

Assumptions of the Rasch Measurement

The first assumption of the Rasch Model is that there is only one latent dimension underlying the items. This assumption is called *unidimensionality*; the item pool should be unidimensional and measure a single latent trait. This factor is not a severe limitation of the method since one can easily eliminate items that appear to violate the assumption. Harvey and Hammer (1999) also report that unidimensionality can be overcome by dividing the instruments into subscales or factors for those instruments with available subscales such as the Myers-Briggs Type Indicator.

A second assumption of the Rasch Model is the local independence of items. That is, items should not give information that could be used to answer any subsequent item. Statistically, local independence means that the items do not correlate with each other (i.e., the items are uncorrelated or have a Pearson *r* at or near zero). Embretson and Reise (2013) describe this concept statistically as being the probability of solving any item *i* where the outcome of that item is independent of any other item.

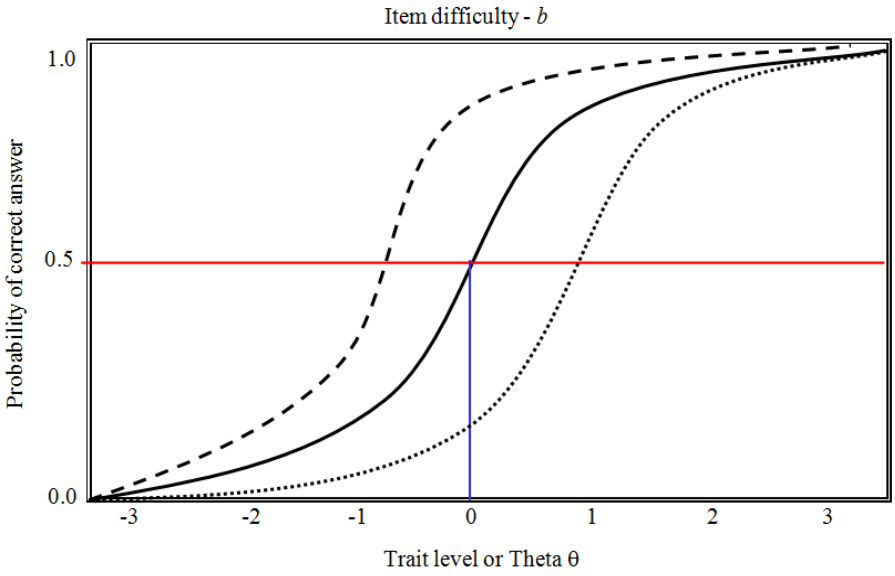

Fig. 1. Item characteristic curves for 3 items of varying difficulty

DOI: 10.3844/ajebasp.2014.159.163

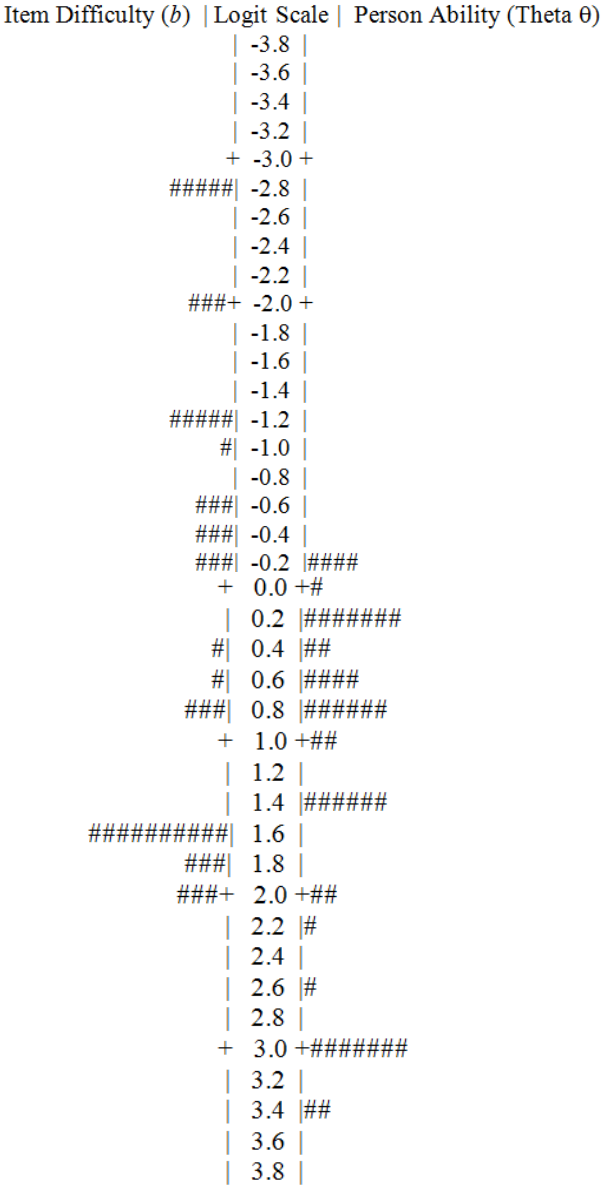

Fig. 2. Item by Person distribution map, Note. Data for this table was compiled from graduate level mid-term exams and is provided for illustrative purposes only

Table 1. Item parameter estimates sorted by item difficulty

Item no.	Item difficulty
16	-2.740
7	-2.740
13	-2.740
15	-2.028
19	-2.028
2	-1.269
10	-1.269
24	-1.269
14	-1.006

The Item Characteristic Curve

The Rasch Model can be used to measure magnitudes of variables using a single continuum (Andrich, 1988). IRT measurements are often used with, but not limited to, inventories or tests that utilize dichotomous responses (e.g., binary data, 0 and1). It is important to note that, in the Rasch Model, items do not have to be dichotomous, merely the scoring of the items is required to be dichotomous in nature. Items may be any type that allows a yes/no or right/wrong scoring regardless of the number of possible choices (distracters) given. Several polytomous IRT models are available to handle multiple-ordered Likert-type responses such as the graded-response model, the partial credit model and the rating scale model (Harvey and Hammer, 1999; Embretson and Reise, 2013). To simplify the discussion, the remainder of this paper will assume dichotomous responses are used.

The Item Characteristic Curve (ICC) is a plot of the latent trait (θ) on the x-axis by the probability of a correct response on the y-axis. The item characteristic function can be described as a mathematical representation of the relationship between a person's position on the latent trait dimension and the probability the person will correctly answer an item of a given difficulty (Hashway, 1978). The scale for the latent trait is typically described as a logarithmic measure (natural log or base e) thus forming a trait scale that is interval or near-interval in nature. According to David and Chih-Hung (2000), by creating an interval scale, parametric statistics are less subject to violation of assumptions and logit measures may temper bias at extremes in the scale. The Rasch Model makes an assumption analogous to equal measurement error for each item and thus are said to be equally discriminating. The visual representation of this item characteristic function is the Item Characteristic Curve (ICC) as seen in Fig. 1. Other similar examples of Rasch ICC's may be found in Azeem and Gondal (2011) and Kersten *et al*. (2014). Figure 1 represents a plot of three separate ICC's on the same scale.

Again, item difficulty is defined as the point at which a person has a 50% probability of answering the item correctly. In Fig. 1 for example, using three dichotomously scored items, the location where the item characteristic curves cross the 0.5 probability line is the item difficulty. Thus, item difficulty for the 3 items illustrated in Fig. 1 is –1.0, 0.0 and 1.0 respectively. A function of the IRT is that it allows item difficulty and person ability to be plotted on the same scale. If a person's ability level exceeds an item's difficulty level, the person will generally pass the item (i.e., the probability of a correct answer increases) (Safrit *et al*., 1989).

Item Response Theory Coefficients

The full IRT model produces 3 parameters for a given data set: Parameters *a*, *b* and *c*. Parameter *a* refers to the slope of the ICC. The slope tells the researcher something about the discriminating power of the item. As the slope, *a*, becomes larger (i.e., more of a vertical orientation), the greater the ability of the item to discriminate between small changes in theta (person ability). As the slope, *a*, becomes smaller or lesser (i.e., more of a horizontal orientation), the less the ability of the item to discriminate between small changes in theta. In the 1-parameter model, parameter *a* is most often assumed to be 1.0 but may be fixed at some other predefined constant (Henson, 1999).

Parameter *c* is the "guessing parameter" and can help researchers take into account the respondent's ability to guess the answer to the item. For example, the probability of guessing a correct answer to a multiple-choice item with 4 options is 25%. In this case, at lower theta values, the ICC would become asymptotic to 0.25 rather than 0. In the Rasch Model, parameter *c* is most often assumed to be 0 but may also be fixed at some other predefined constant (Henson, 1999).

The Rasch Model

Common to the full IRT model and the Rasch Model, parameter *b* is the item difficulty and is defined as the position on the latent trait variable where it is expected the person has a 50% probability of answering the item correctly. The further to the right on the plot the ICC stands, the greater the item difficulty as only those individuals with a higher theta would have a 0.5 or greater probability of having a correct answer. Figure 1 represents an ICC for 3 items of varying difficulty. An important concept to note is that theta (θ), the latent trait or characteristic of the individual being measured, uses the same scale as the *b* parameter (item difficulty). According to Harvey and Hammer (1999), the location of the person and item parameters on a common scale represents an important, if not the critical, characteristic of IRT models. Hashway (1978) reinforces this concept as he discusses how the Rasch procedure assumes that both items and subjects occupy positions on the same latent trait dimension (i.e., the same scale).

Again, referring to Fig. 1 for the Rasch Model, the only characteristic distinguishing one item's difficulty from another is the location of the ICC on the horizontal axis (theta). The further left the ICC is on the graph, the lower the item difficulty and the further right the ICC is on the graph, the larger the item difficulty (i.e., the more difficult the item). The Rasch Model also assumes that all Item Characteristic Curves are the same shape, which in the practical world is probably not completely true. As noted previously, the Rasch Model holds constant both the item discrimination parameter, *a* and the guessing parameter, *c*.

If the person's theta (latent trait) exceeds the item difficulty, the person is more likely to answer the item correctly. Conversely, if the person's theta is less than the item difficulty, the person will likely not answer the item correctly. This is an important point to the test developer. If the test item difficulty far exceeds the student's ability (theta), students will do poorly and the test will not yield significant information regarding the true ability of the students. Conversely, if the test item difficulty is significantly below that of the student's ability (theta), similar results occur: No significant information regarding student ability will be generated. IRT methods will often help discriminate between students with abilities at extremes of the distribution of scores by assisting the test developer in the development of items with many different item difficulties to assess different person (ability) levels.

For example, using three dichotomously scored items (Fig. 1), the location of subjects on the trait level continuum (x-axis) corresponds to their ability or trait level. The location of the items corresponds then to each item's difficulty levels. If a person's ability level exceeds an item's difficulty level, the person will generally pass the item (i.e., the probability of a correct answer increases) (Safrit *et al*., 1989). In a slight variation, Petersen *et al*. (2012) discuss using the Rasch methodology to evaluate raters making medical diagnosis (i.e., rater bias) based on the level of difficulty in rating patients.

Per the previous discussion, Hashway (1978) describes how the Rasch procedure places or calibrates both items and subjects (persons) to occupy positions on the same latent trait dimension (i.e., the same scale). Figure 2, an item by person distribution map generated from the same data set as that used for Table 1, is provided as a visual example to help relate the concept of trait level and item difficulty being on the same scale. Figure 2 shows the logit scale occupying the central potion of the map. Item difficulty (*b*) is displayed graphically to the left of the logit scale. Each marker (#) represents the percent of items at a particular item difficulty. Notice that several items, as a percentage, have an item difficulty of –2.8. From the previous discussion of item difficulty, using the values listed in Table 1, these markers correspond to items 16, 7 and 13. The map obviously rounds the item difficulty values. In this example, the item difficulty of –2.8 on the map corresponds to the calculated value of –2.740 for items 16, 7 and 13.

Person ability (theta) is displayed graphically on the right side of the logit scale on item by person distribution map (Fig. 2). Again, each marker (#) represents the percent of examinees at a particular person ability or theta level. In this example, theta levels of the examinees reside at the upper level of the item difficulties. In some cases, theta levels exceed item difficulties.

In simple terms, Fig. 2 shows the results of a Rasch assessment which creates a common scale for both item difficulty and person ability. In this example, it is easy to see that the item difficulties span a wide range (-2.8 to 2.0) and are generally below the person abilities. The person abilities are generally higher than the item difficulties and also in a narrower range (-0.2 to 3.4). What does this mean to the instructor or test developer? In general terms this assessment will allow the instructor to see the abilities of the students in relation to the difficulty of the items on a test. The logical outcome in this example is that the instructor or test developer could refine the test by removing items that have a low item difficulty, reducing the number of items that have the same item difficulty and adding items at a higher difficulty level.

Conclusion

The Rasch Model is gaining in use due to the widespread growth of computer applications and the increasing sophistication of computer programs to run demanding mathematical operations (Harvey and Hammer, 1999). This model is also being used in an array of different subject areas from Human Resources (Wang and Stahl, 2012) to the field of medicine (Petersen *et al*., 2012) as more researchers are seeing the benefits of the Rasch approach to item analysis. The Rasch Model assists test developers by providing a platform to calibrate instruments to be independent of the norm reference group. The Rasch Model is also helpful in diagnosing instruments by calibrating item difficulty and person ability to a common scale. This function of the Rasch Model allows test developers and instructors to create better instruments in terms of optimizing the number of items, eliminating items of the same difficulty and more closely matching the level of difficulty of the items to the abilities of the examinees.

Acknowledgement

This work was supported by the University of North Texas department of Applied Technology and Performance Improvement and the author appreciates the faculty for their support.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Andrich, D., 1988. Rasch Models for Measurement. 1st Edn., Newbury Park, CA: Sage Publications, Newbury Park. ISBN-10: 080392741X, pp: 95.

Azeem, M. and M.B. Gondal, 2011. Math proficiency assessment based upon item response theory. Int. J. Interdisciplinary Soc. Sci., 6: 105-122.

Crocker, L.M. and J. Algina, 1986. Introduction to Classical and Modern Test Theory. 1st Edn., Holt, Rinehart and Winston, New York, ISBN-10: 0495395919. pp: 527.

David, C. and C. Chih-Hung, 2000. A discussion of item response theory and its applications in health status assessment. Medical Care, 38:66-72. DOI: 10.1097/00005650-200009002-00010

Embretson, S.E. and S.P. Reise, 2013. Item Response Theory for Psychologists, 1st Edn., Psychology Press, ISBN-10: 1135681465, pp: 384.

Fan, X., 1998. Item response theory and classical test theory: An empirical comparison of their item/person statistics. Educ. Psychol. Measurement, 58: 357-381. DOI: 10.1177/0013164498058003001

Hambleton, R.K., 1983. Applications of Item Response Theory. 1st Edn., Vancouver, Canada: Institute of British Columbia.

Hambleton, R.K., 2000. Emergence of item response modeling in instrument development and data analysis. Med. Care, 38: 60-65. DOI: 10.1097/00005650-200009002-00009

Harvey, R.J. and A.L. Hammer, 1999. Item response theory. Counseling Psychologist, 27: 353-383. DOI: 10.1177/0011000099273004

Hashway, R.M., 1978. Objective Mental Measurement: Individual and Program Evaluation using the Rasch Model. 1st Edn., Praeger Publishers, New York. ISBN-10: 0275902978, pp: 105.

Henson, R.K., 1999. Understanding the one-parameter Rasch Model of item response theory. Educational Research Association, San Antonio, USA.

Kersten, P., P.J. White and A. Tennant, 2014. Is the pain visual analogue scale linear and responsive to change? An exploration using rasch analysis. PLoS ONE. DOI: 10.1371/journal.pone.0099485

Petersen, J.H., K. Larsen and S. Kreiner, 2012. Assessing and quantifying inter-rater variation for dichotomous ratings using a Rasch Model. Statistical Methods Med. Res., 21: 635-652. DOI: 10.1177/0962280210394168

Safrit, M.J., A.S. Cohen and M.G. Costa, 1989. Item response theory and the measurement of motor behavior. Res. Exercise Sport, 60: 325-335. DOI:10.1080/02701367.1989.10607459

Shu'aibu, B., A.S. Bappah and M.S.B. Saud, 2013. Modelling ICT integration in teaching and learning of technical education. J. Soc. Sci., 9: 81-88. DOI: 10.3844/jsssp.2013.81.88

Wang, N. and J. Stahl, 2012. Obtaining content weights for test specifications from job analysis surveys: An application of the many-facts Rasch Model. Int. J. Testing, 12, 299-320.