PEL Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana
2015, 52(2), 1-5

Pontificia Universidad Católica de Chile
Facultad de Educación

Special Section Introduction

Advances in Rasch Modeling: New Applications and Directions for Objective Measurement Science

[1]Brent Duckor, [2]María Verónica Santelices, and [3]Steffen Brandt

[1] College of Education, San Jose State University, California, USA
[2]Faculty of Education, Pontificia Universidad Católica de Chile
[3]Art of Education, Statistical Analyses

Over 50 years ago, Georg Rasch helped found the field of Item Response Theory with the model that bears his name, distinguished by the use of a single parameter to model the relationship between item difficulty and person ability (Rasch, 1960/1980, 1977). Long considered the gold standard in «objective» measurement, various extensions of the Rasch model have been proposed, applied, and investigated for use in educational testing and assessment settings. According to the Program Committee of the Objective Measurement Institute (2000), «objective» measurement is «the repetition of a unit amount that maintains its size…no matter which instrument…is used and no matter who or what relevant person or thing is measured.» The Rating Scale Model, the Partial Credit Model, and the Randomized Coefficient Multinomial Logit Model are all part of the Rasch family. Initially focused on reading achievement, Rasch models are now employed in such diverse fields as health care, physical therapy, nursing, pharmaceuticals, and competitive sports.

© 2014 PEL, http://www.pensamientoeducativo.org - http://www.pel.cl

ISSN: 0719-0409 DDI: 203.262, Santiago, Chile
doi: 10.7764/PEL.52.2.2015.1

The Rasch model puts our theory of the structure of any latent variable to a rigorous scientific test by asking: «Does the data fit the model?» (Wright & Master, 1980). From a constructing measures perspective (Wilson, 2014), the Rasch model framework allows researchers to explore the validity of the scales, starting from construct theory and contrasting that theory with the evidence of the empirical orderings of persons and items on a Wright map. Much in line with the current notion of the validity argument (Kane, 2015), the Rasch framework helps explore the internal structure of the construct and the validity of its content, both at the same time.

In addition, Wilson (2005) reminds us that the Rasch framework puts forth a set of first principles for measuring persons and calibrating items on a single scale; far from being just another scaling technique, this approach is fundamentally important to the interpretations made in measurement science (Duckor, Draney, & Wilson, 2009). When choosing and evaluating a measurement model, researchers should think spatially, in terms of a geographic map. The idea of the «location» of an item response with respect to the location of another item response only makes sense if that relative meaning is independent of the location of the respondent involved. That is to say, the interpretation of relative locations needs to be uniform no matter where the actual respondent is. This invariance requirement corresponds to the idea that an «inch represents a mile» or a «meter represents a kilometer», wherever you are on a geographical map.

Over the years, the one-parameter (Rasch) model has faced challenges from item response theorists (IRT) who favor more flexible models and better data fit. Some IRT researchers (Birnbaum, 1968) have pushed to include additional parameters in order, for example, to model variation in item discriminations (2PL) or variation in guessing probabilities (3PL). While the Rasch model requires that all items be equally discriminating in order to define the ability that is to be measured, the 2PL and 3PL models allow discrimination to vary across items and calculate it recursively as part of the estimation process. This also impacts the estimation of the item difficulty parameter, creating a sharp difference between how Rasch and 2PL/3PL item difficulties can be interpreted. Because item discrimination is at least in part a property of how a particular sample of examinees interacts with an item and is not exclusively a property of the item, and because examinees vary across tests, the inescapable consequence is that the scores calculated using the 2PL and 3PL models are not guaranteed to be as generalizable across tests as scores calculated from data that are constrained to fit the requirements of the Rasch model.

Thus, although the 2PL and the 3PL may fit the data better, the gain often comes at a cost. First, the interpretation of the test scores is not as clear, since the different discrimination parameters of the items result in different weightings of the items in the test score; whether this is due to an item or a sample characteristic is unknown. Second, the scientific principle of checking on one's theoretical expectations about a construct (in a validity framework) is lost if researchers cannot falsify hypotheses about data structure. Moreover, recent studies (San Martín, González, & Tuerlinkx, 2015) have shown the unidentifiability of the 3PL model, examining how, even after fixing the difficulty, the discrimination, and the guessing parameters of an item, the remaining items' parameters are still unidentified by the observations, do not have an empirical interpretation, and cannot be unbiasedly and consistently estimated.

These reasons explain why the Rasch model is still used today and remains an important tool in the psychometrician's and educational measurement specialist's toolkit.

In accordance with the traditions established by the predecessor series, *Objective Measurement: Theory into Practice* (Vols. 1-5) and *Advances in Rasch Measurement* (Vols. 1-2), this journal is pleased to offer both the theoretical and practical applications of Rasch measurement models in this issue. All papers were originally presented at the International Objective Measurement Workshop (IOMW) 2014 in Philadelphia, PA, United States. The IOMW is a biennial conference, which takes place before the conference of the American Educational Research Association (AERA), and gathers experts from around the world to share their work in the areas of Rasch modeling, psychometrics, and philosophy of measurement. Manuscripts were solicited for thematic coherence and fit, and each of them was blind-reviewed by at least two experts. The five papers gathered for the special topic, «Advances in Rasch Modeling: New Applications and Directions», consider the Rasch model from different angles and applications.

The first paper from Andrich (2015), «Components of Variance of Scales with a Subscale Structure using Two Calculations of Coefficient α», addresses the issue of analyzing dimensionality, the coefficient alpha, and subscales variance. It notes that scales constructed to measure a single variable are nevertheless composed of subscales of items, which measure different aspects of the variable. Using a simulation study, the paper proposes a simple method that can be used to provide a more comprehensive summary of the properties of a scale with subscales than is possible with an estimate of the reliability coefficient. It shows that, with some common simplifying assumptions and using a bifactor structure, the ratio of two calculations of the coefficient alpha, one at the level of the items and the other at the level of the subscales, can be used to obtain (a) the proportion of true common variance, (b) the proportion of the true unique variance, (c) the proportion of the true common variance relative to the sum of the true common and unique variances, and (d) the summary correlation among subscales immediately corrected for attenuation due to error. The paper shows how Rasch scholars can shed new light on traditional problems and approaches about reliability for those using the coefficient alpha.

The second paper, by Behizadeh and Engelhard (2015), «What Is a Valid Writing Assessment from the Perspectives of the Writing and Measurement Communities?» addresses issues related to validity, consequential validity, writing assessment, and communities of practice, with a particular focus on Rasch measurement theory. The study examines the concept of validity in two distinct communities of practice: the writing research and educational measurement communities. It highlights the contributions that *Rasch measurement theory* (Rasch, 1960/1980) brings to understanding and evaluating validity. By connecting technical and non-technical perspectives on validity, the authors explore points of consensus and convergence regarding validity. This research has implications for improving research, theory, and practice in writing assessment for scholars and practitioners.

The third paper, by Fisher and Wilson (2015), «Building a Productive Trading Zone in Educational Assessment Research and Practice» explores the challenges of measurement across different institutional contexts. It argues that diverse viewpoints regarding the act of measuring —from the science laboratory to the classroom to the marketplace— can be reconciled with the use of boundary objects that allow for shared meanings. The authors describe how psychometrically modeled exemplars known as *construct maps* and *Wright maps*, developed based on the Rasch model, function as boundary objects and can serve as a basis for productive analogies in educational assessment by (a) preserving relational structures, (b) making isomorphic mappings between systems, and (c) facilitating systematicity, understood as mapping systems of higher order relational structures. Using the case of the BEAR Assessment System and its accompanying software, the paper explores how such technologies support practical alliances of teaching, policy-making, assessment and curriculum development, psychometrics, and information technology.

In the fourth paper, Korpershoek (2015) presents «An Investigation of the Reliability and Validity of the Utrecht-Management of Identity Commitments Scale Adapted to Measure Students' Identity Formation Processes at their University», which addresses issues of commitment and identity development processes in psychological measurement. The paper examines evidence for the construct validity and predictive validity of the measurement framework presented by Crocetti, Rubini, and Meeus (2008), which was originally developed to measure the identity formation processes (such as achieving commitments) of individuals in various domains. Drawing from a pool of university students and using a multidimensional Rasch model, the paper then explores an adapted version of the Utrecht-Management of Identity Commitments Scale (U-MICS) to measure university students' identity formation processes, which are a part of students' personal identities at their university. The results showed some signs that internal structure evidence can represent students' identity formation processes at the university, although some items need further improvement to fit the multidimensional model better.

The fifth paper, by Williamson (2015), «Measuring Academic Growth Contextualizes Text Complexity», addresses the challenges and opportunities for measuring growth in reading. Using Rasch first principles for specific objective measurement, the paper argues that optimal measurement of students' academic growth requires a scale that is unidimensional, continuous, equal-interval, developmental, and invariant with respect to location and unit size. It provides an empirically validated example of Rasch measurement with an operationalized reading construct theory. Using the Lexile scale, the paper illustrates a text-complexity continuum, where persons (readers) and items (texts) are brought onto a common «academic growth» scale. The paper further explores recent educational policy developments (College and Career Readiness Standards) in the U.S. that recommend increasing students' exposure to complex texts. The author notes that parametric modeling of alternative growth curves can better frame conversations about how exactly students might attain these particular college and career readiness goals.

These five articles are fine examples of the combination of the theoretical and empirical work that characterizes the Rasch analyses, and we hope that they will motivate readers to learn more about the Rasch models and their potential uses for evidence-based validity arguments. They show how Rasch model concepts and statistical modeling are applicable to a broad range of issues, including instrument development, innovative research programs, and public policy. We look forward to readers joining the next IOMW conference to be held in Washington, D.C. in spring 2016.

References

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395-479). Reading, MA: Addison-Wesley.

Crocetti, E., Rubini, M., & Meeus, W. (2008). Capturing the dynamics of identity formation in various ethnic groups. Development and validation of a three-dimensional model. *Journal of Adolescence, 31,* 207-222. doi: 10.1016/j.adolescence.2007.09.002

Duckor, B., Draney, K., & Wilson, M. (2009). Measuring measuring: Toward a theory of proficiency with the Constructing Measures framework. *Journal of Applied Measurement*, *10*(3), 296-319.

Kane, M. T. (2015). Explicating validity. *Assessment in Education: Principles, Policy & Practice*. Retrieved from http://dx.doi.org/10.1080/0969594X.2015.1060192

Program Committee of the Objective Measurement Institute (2000). *Definition of objective measurement*. Retrieved from http://www.rasch.org/define.htm

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago: The University of Chicago Press.

Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *The Danish Yearbook of Philosophy, 14*, 58-93.

San Martín, E., González, J., & Tuerlinkx, F. (2015). On the unidentifiability of the fixed-effects 3PL model. *Psychometrika*, *80*(2), 450-467. doi: 10.1007/s11336-014-9404-2

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wilson, M. (2014). Considerations for measuring learning progressions where target learning is represented as a cycle. *Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana, 51*(1), 156-174. doi: 10.7764/PEL.51.1.2014.12

Wright, B. D., & Masters, G. N. (1981). *The measurement of knowledge and attitude*. Chicago: Statistical Laboratory, Department of Education, University of Chicago.