Invariance and the Three Traditions of Measurement in the Human Sciences

George Engelhard, Jr.[1a]

[1]The University of Georgia, Athens, GA, USA

[a]gengelh@uga.edu

Keywords: invariant measurement, Rasch measurement theory, test-score tradition, scaling tradition, structural tradition, explanatory Rasch models.

Abstract: This chapter presents key aspects of invariant measurement. Invariant measurement can be broadly viewed as the application of scientific principles to the development of scales. Three traditions of measurement are identified and described (test-score, scaling, and structural traditions) for categorizing widely used measurement theories. Each of these traditions has a distinctive perspective on invariant measurement. Rasch measurement theory offers the opportunity to integrate these three perspectives based on extensions of the Rasch model that include the development of explanatory measurement models with both item and person covariates.

1. Personal reflections on Rasch measurement theory

This chapter is based on my presidential address at the annual meeting of Pacific Rim Objective Measurement Society in Macau in 2023. In preparing for my presidential address, I thought about my long-term commitment to studying and teaching about Rasch measurement theory. I first met Professor Ben Wright at the University of Chicago in the summer of 1977 (45 years ago). He shared with me his ETS lecture [1], and an article by Bruce Choppin [2]. Needless to say, I very quickly became hooked on these new ideas regarding item-invariant person measurement and person-invariant item calibration. In addition to the power of these ideas, Ben Wright made a strong case for *why* measurement should meet the requirements of invariant or objective measurement. Recently, the work of Sinek [3] on successful leaders introduced a golden circle model that argues for the importance of stressing why before dealing with the what and how in business and other endeavors. Ben Wright knew *why* we needed invariant or what he called objective measurement, he made it clear *how* to use Rasch measurement theory to achieve this goal, and he stressed the *what* in terms of the creation of invariant measures. I view the *why, how,* and *what* as follows:

- Why: Our goal is to create measures to represent key constructs for improving human sciences (theory, research, and practice).
- How: These measures are clearly defined based on scientific principles related to invariance using Rasch measurement theory.
- What: These invariant measures can be used to discover invariant structural relationships.

This chapter focuses on an introduction to selected aspects of the how and what related to invariant measurement.

Q. Zhang (ed.), *Proceedings of the Pacific-Rim Objective Measurement Symposium (PROMS 2023)*,
Atlantis Highlights in Social Sciences, Education and Humanities 23,
https://doi.org/10.2991/978-94-6463-494-5_2

2. Purpose

Invariant measurement is the use of scientific principles to guide measurement and create scales. Various measurement theories offer us different perspectives on invariance, and I have found it useful to identify three traditions of measurement: test score, scaling, and structural traditions. These traditions have the characteristics of paradigms, and they also provide different lens for seeing and not seeing a variety of measurement issues. The purpose of this chapter is to consider the following research questions:

- What is invariant measurement?
- What are the three major research traditions in measurement?
- How does invariance appear within the three traditions of measurement?

In addressing these questions, Rasch measurement theory provides the underlying frame of reference.

3. What is invariant measurement?

The mathematical rendition of the requirements of invariance in a probabilistic framework ... is the major contribution to Rasch's work having the characteristics of a theory [4].

The philosopher Robert Nozick identified invariance as an essential feature of science related to the concept of objectivity. Nozick [5][6] described four specific and related aspects of objectivity. First, evidence for objectivity is provided by accessibility from different perspectives. It is accessible by different observers, and objectivity can be replicated under different conditions. The second aspect of objectivity is intersubjectivity. This implies independent agreement among observers. The third aspect relates to independence. Objectivity is evident when it is independent of the particular beliefs, desires, and hopes of observers. Finally, the fourth aspect relates to invariance, which underlies all of the other aspects of objectivity. According to Nozick [6],

> what is objective about something, I have claimed, is what is invariant from different angles, across different perspectives, and under different transformations. Yet often what is variant is what is especially interesting. We can take different perspectives on a thing (the more angles the better), and notice which of its features are objective and invariant, and also notice which of its features are subjective and variant (p. 102)

This tension between variance and invariance appears within the three research traditions of measurement. As a preview, the test-score tradition focuses on reducing error variance, while the scale tradition focuses on creating invariant scales. Scientific measurement includes a consideration of both variance and invariance.At its core, invariant measurement involves the application of basic science principles to our activities in creating, developing, and using scales.

Measurement plays a foundational role in all human sciences. An early statement of this perspective is that "the history of science is the history of measurement"[7]. A similar view of the importance of measurement is echoed in the 21st century:

> It would be difficult to overstate the value and importance of measurement in nearly every aspect of society. Every time we purchase or eat food, take prescribed medicine, travel in a vehicle, use a phone or computer, or step inside a building, we place our trust int the results of measurement—and for the most part, that trust seems well earned, and as such measurement is commonly associated with precision, accuracy, and overall trustworthiness [8].

Measurement theories are conceptual models that provide a systematic way to understand, evaluate, and interpret assessment processes. One view of measurement is based on Stevens who stated that "measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules" [9]. Of course, this definition does not define the specific rules that are embedded within different measurement theories. The three research traditions (test score, scaling, and structural traditions) discussed later in this chapter highlight the categorization of measurement models with different characteristics based on different guiding principles or rules. The concept of invariant measurement is the underlying principle underlying numerous measurement theories. Essentially, the "rules" mentioned by Stevens [9] come from our measurement theories, and the quest for invariance in our measures is based on the application of scientific principles to guide our efforts.Different measurement theories provide various perspectives and approaches for considering variance and invariance. This chapter stresses the utility of Rasch measurement theory as an underlying base to develop, apply, and evaluate the measures or scales that have been developed to represent a latent variable of construct. Rasch measurement theory provides a strong foundation for considering a variety of measurement issues related to invariance.

Measurement theories play an important role in many aspect of the human science because they

- define the aspects of quantification that are defined as problematic,
- determine the statistical models and appropriate methods used to solve these problems,
- determine the impact of our research in the social, behavioral and health sciences,
- frame the substantive conclusions and inferences that we draw, and ultimately
- delineate and limit the policies and practices derived from our research work in the social, behavioral and health sciences

Given the ubiquitous aspects of measurement, it is important to consider the basic principles related to invariance that guide our evaluations of how well our measures are functioning. It is also important to broaden these principles beyond traditional indices of psychometric quality to include a consideration of invariant structural relationships among scales.

The connections between invariance and measurement has been discussed by several measurement theorists who have stressed the importance of invariance. Examples of key measurement theorists are Thorndike [11], Thurstone [12], Guttman [13][14], Lazarsfeld [15], Mokken [16], and Rasch [17][[18][19]. These measurement theorists have important perspectives on to conceptualize measurement with regard to invariance, among other issues. In a series of publications, Engelhard [20][21][22] discussed and compared these measurement theorists based on their perspectives related to key measurement concepts, including invariance.

One major commonality between these selected theorists is that their measurement theories emphasize the central role of a single invariant and unidimensional latent variable. A second common feature that at cuts across these measurement theories is the idea that person measurement should be independent or invariant over different sets of items. The thirdcommonality is that item calibrations and procedures for locating items on the latent variable should be independent of particular persons. These three issues can be briefly described as follows:

- Invariant unidimensional continuum: Items and persons must be simultaneously located on an underlying latent variable.

- Item-invariant measurement of persons: The measurement of persons must be independent of the particular items used. A moreable person must always have a better chance of success on an item than a less able person.
- Person-invariant calibration of test items: The calibration of the items must be independent of the particular persons used. Any person must have a better chance of success on an easy item than on a more difficult item.

In answer to the question posed in this section, invariant measurement is the use of scientific principles to guide measurement. Invariant measurement includes the creation of an invariant unidimensional continuum to represent an underlying latent variable (construct). Invariant measurement also provides the opportunity to achieve item-invariant measurement of persons, and person-invariant calibration of items based on this invariant unidimensional continuum.

4. What are the three major research traditions in measurement?

There are a plethora of measurement theories, and Laudan's concept of research traditions is as a useful frameworkto broadly categorize these measurement theories [23]. Research traditions are similar to paradigms[24], and research programs[25]. Laudan identifies three key characteristics defining research traditions: they delineate problematic aspects of quantification, prescribe methods to address these issues, and significantly influence the social science research methodologies. The three major research traditions are test-score, scaling, and structural traditions. Each tradition is characterized by distinct historical roots and methodologies. These research traditions not only shape the conceptualization of measurement, but also impact the very fabric of social science research, offering varied assumptions, requirements, and perspectives that guide researchers in their pursuit of psychometrically sound measurement.

As its label implies, the test-score tradition focuses on test scores with a primary concern with reducing measurement error, and the decomposition of an observed test score into several components including a true score with various error components. Linear models define the form of measurement models in the test-score tradition. The primary goal is the estimation and reduction of error variance. This tradition has its roots in the psychometric work of Spearman [26] on classical test theory and generalizability [27].

Scaling theory is the second dominant research tradition in measurement theory. Scaling theory has its roots in 19th century psychophysics, and it has continued to the present through various forms of item response theory [28] [17]. The focus of measurement theories in the scaling tradition is on item-person responses. Non-linear models are used as the form of models in the scaling tradition. A major goal of the scaling theory tradition is the creation of an invariant scale represents the location of both items and persons on a latent variable scale that represents a construct.

The last research tradition is the structural tradition. This tradition includes path analysis[29][30], factor analysis [31], structural equation models [32], and explanatory item response models [33]. The characteristics of each of these three traditions in terms of four issues (measurement models, focus, form of models, and illustrative theorists)are summarized in Table 1.

Table 1 Three Research Traditions for Classifying Measurement Theories.

Issue	Test-Score Tradition	Scaling Tradition	Structural Tradition
Measurement models	Classical Test Theory Generalizability theory	Psychophysical Models Absolute Scaling Item Response Theory Parametric/non-parametric	Path analysis Factor Analysis Structural Equation Modeling Explanatory Item Response Models
Focus	Test Scores	Item-Person Responses	Correlationsbetween items and latent variables
Form of models	Linear models	Non-linear models	Structural models
Primary Goal	Estimate variance components to reduce error	Estimate invariant scales to use for measurement	Estimate invariant relationships between variables (observed/latent)
Illustrative theorists	Spearman, C. Cronbach, L.J.	Guttman, L. Lazarsfeld, P. Mokken, R. Rasch, G.	Wright, S. Thurstone, L.L. Joreskog, K. De Boeck, P. Wilson, M.

5. How does invariance appear within the three traditions of measurement?

A major weakness common to both classical test theory and generalizability theory is the sample-dependent nature of the estimation procedures ... one of the critical properties of the item response theories… is that item parameters are invariant across groups of examinees while at the same time estimates of examine ability or trait level are invariant across sets of items measuring the same ability or trait. [34]

Invariance is a key aspect of science, and the application of scientific principles to measurement can yield invariant measurement. The three measurement traditions discussed in this section are test-score, scaling, and structural traditions. Table 1 summarizes the issues distinctive to each tradition. Figure 1 illustrates the connections between these three research traditions.

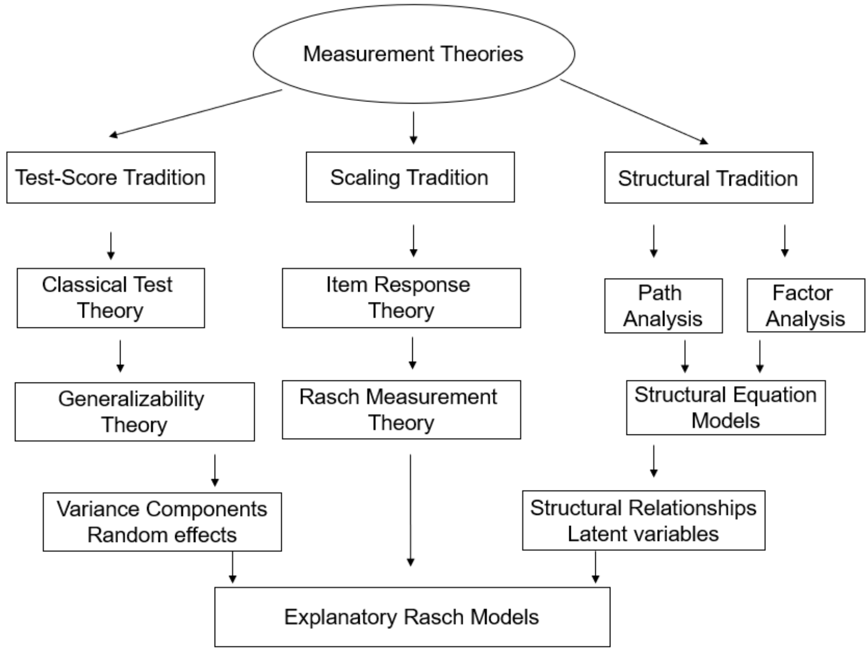

Figure 1 The three traditions of measurement theories.

The test-score tradition includes classical test theory and generalizability theory. As shown in Figure 1, these two theories focus on estimation of variance components and random effects. These essential define the focus on invariance for measurement theories within this tradition. The scaling tradition includes item response theory and Rasch measurement theory. In this case, the goal is to estimate invariant measures to create a continuum to represent a latent variable or construct. Finally, the structural tradition includes path analysis and factor analysis models that are frequently combined to form structural equation models. The structural tradition is concerned with the identification of invariant relationships between variables. Lastly, Figure 1 suggests that issues from all three traditions can be integrated into explanatory Rasch models that are discussed in more detail later in this chapter.

5.1 Test-Score Tradition

Classical test theory (CTT) was developed in the early 20th century, and it represents the founding of the test-score tradition. CTT is based on three key ideas: measurements have errors, these errors can be modeled as a random variable, and correlation coefficients can be corrected for these measurement errors to produce so called disattenuated correlations. The key figure in the development of CTT is Spearman [26] who proposed a method for correcting correlations coefficients to account for measurement errors – his correction for attenuation.

It is important to note that the test-score tradition tends to focus on the "variant" aspects of measurement with the goal of reducing of sources of error variance. This stands in contrast to theories within the scaling tradition that address many of the issues related to invariant measurement. However, issues related to variance and invariance are in a sense two sides of the same coin.

5.2 Scaling Tradition

The connections between the scaling tradition and measurement in the social, behavioral, and health sciences have their roots in 19th century psychophysics. Psychophysics focuses on the scaling or calibration of stimuli (items and tasks), while measurement focuses on the scaling of responses from individuals. Psychophysics for example focused on how well people can distinguish variation in lifted weights and other aspects of sensory perceptions related to sight, hearing, smell, and taste. A detailed treatment of this duality between psychophysics and psychometrics appears in the work of Mosier [35][36]).

The scaling tradition has a number of distinctive features. The single and most distinctive feature of the scaling tradition is development of a visual representation of the construct as a variable map. Embretson [37] has highlighted progress that has been made in the movement from the test-score tradition to the scaling tradition as represented by item response theory that she has called *the new rules of measurement.*

5.3 Structural Tradition

The structural tradition is an important and distinct tradition that is related to both the test score and scaling traditions, but warrants consideration as a separate tradition. Structural models play an important role in current measurement practice as part of the validity argument for the proposed interpretation of test scores. From the perspective of invariant measurement, the structural tradition seeks to identify invariant relationships among both observed and latent variables. Examples of structural models include factor analysis, path analysis, structural equation modeling, and item response theory with covariates.The structural tradition takes invariance to the next level by focusing on the discovery and exploration invariant relationships between latent and observed variables.

In summary, the three research traditions (test-score, scaling, and structural) offer a useful way of viewing a variety of measurement theories created during the 20th century that continue to echo into the 21st century. Scientific perspectives of measurement can focus on either variant or invariant aspects.

5.4 Combining aspects of the three traditions

An exciting development in measurement is the recognition that Rasch models can be estimated using generalized linear mixed models (GLLMs). De Boeck and Wilson [33] made an important distinction between descriptive and explanatory item response models. Descriptive models do not include any covariates in the model, while explanatory models can include both item and person covariates. GLMMs can be used to develop a variety of explanatory Rasch models, and to construct a bridge between the scaling and structural traditions from a Rasch measurement perspective. Since GLMMs also include random effects, there are aspects of these models with implications related to the test-score tradition with its emphasis on variance components.

The dichotomous Rasch model is labeled as doubly descriptive with no covariates. The linear logistic Rasch model with item covariates is labeled as an item explanatory model, and the latent regression Rasch model with person covariates is labeled as a person explanatory model. Finally, models with both item and person covariates are labeled as a double explanatory model. This yields four general models: the Dichotomous Model, Linear Logistic Rasch Model, Latent Regression Rasch Model, and Combined Covariates Rasch Model. These are summarized in Table 2. It is beyond the scope of this chapter to describe and illustrate these models.

Table 2 Rasch Models with Item and Person Covariates for Dichotomous Data.

	Covariates		**De Boeck & Wilson (2004)**
Models	**Item**	**Person**	
1. Dichotomous Model	No	No	Doubly descriptive
2. Linear Logistic Rasch Model	Yes	No	Item Explanatory
3. Latent Regression Rasch Model	No	Yes	Person Explanatory
4. Combined Covariates Rasch Model	Yes	Yes	Doubly Explanatory

It is important to note that explanatory Rasch models combine aspects of each measurement tradition, and they can be estimated with generalized linear mixed models to yield a broader framework for exploring invariant measurement and invariant relationships. A map of the descriptive and explanatory Rasch models is shown in Figure 2. Descriptive Rasch models include dichotomous and polytomous models (Rating Scale and Partial Credit Models). Multifaceted models can also be considered descriptive Rasch models. These models can be extended by adding item and person covariates to yield a number of exciting explanatory models. In the next few years, I predict that this will be an active and exciting area of progress related to Rasch measurement theory.

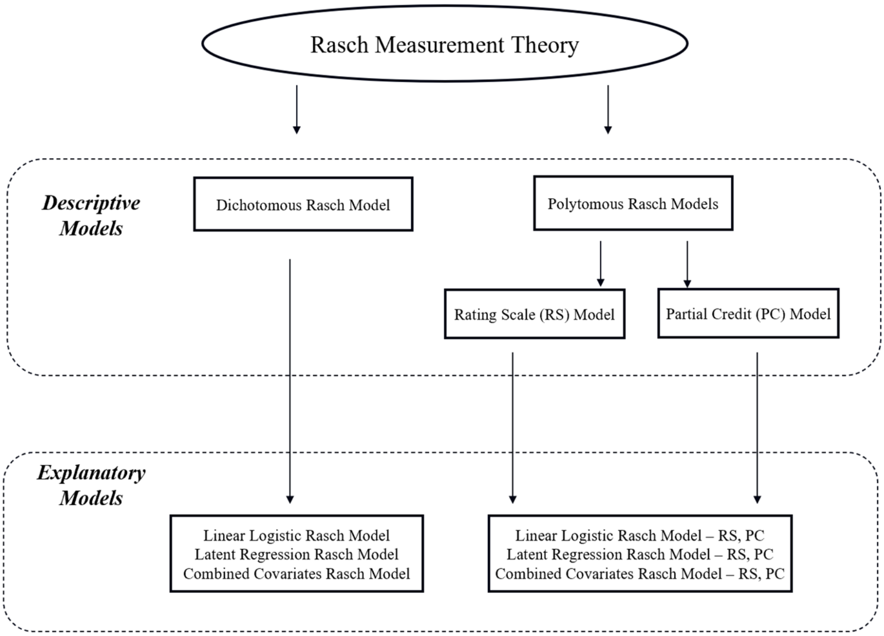

Figure 2 Descriptive and Explanatory Rasch Models.

6. Summary

[T]he fundamental goal of science is to find invariants [38].

I started this chapter with personal reflections on my commitment to teaching about Rasch measurement theory. The chapter focuses on selected issues related to the *how* and *what* of invariant measurement and invariant structural relationships. However, it is important to deepen our understanding of *why*: the goal of our efforts is to work towards the identification, measurement, and appropriate use of key constructs for improving human sciences based on the principles of invariance. My hope is that researchers around the world will join the quest for the development and use of invariance broadly conceived to guide our research, theory, and practice in the human sciences. This can be briefly stated as:

- Why: Goal is to create measures to represent key constructs for improving human sciences (theory, research, and practice)
- How: Measures are clearly defined based on scientific principles related to invariance using Rasch measurement theory.

- What: Creation of invariant measures that can be used to discover invariant structural relationships.

In summary, invariance is a key goal in science and measurement. The creation of invariant measures to represent key constructs, and the discovery of invariant relationships between these constructs is major goal of the human sciences.

References

1. Wright, B. D. (1968). Sample-free test calibration and person measurement. In *Proceedings of the 1967 invitational conference on testing problems* (pp. 85-101). Princeton, NJ: Educational Testing Service.

2. Choppin, B. (1968). Item bank using sample-free calibration. *Nature, 21* (August 24), 870–872.

3. Sinek, S. (2011). *Start with why: How great leaders inspire everyone to take action.* Penguin.

4. Andrich, D. (2018). A Rasch measurement theory. In F. Guillemin et al. (Eds), *Perceived Health and Adaptation in Chronic Disease* (pp. 66-91). Routledge.

5. Nozick, R. (1998). Invariance and Objectivity. *Proceedings and Addresses of the American Philosophical Association, 72*(2), 21-48.

6. Nozick, R. (2001). *Invariances: The structure of the objective world.* Harvard University Press.

7. Cattell, J. M. (1893). Mental measurement. *The Philosophical Review, 2*(3), 316-332.

8. Mari, L., Wilson, M., & Maul, A. (2021). *Measurement across the Sciences.* Springer.

9. Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677-680.

10. Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York: Routledge.

11. Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements.* Teacher's College, Columbia University.

12. Thurstone, L. L. (1959). *The measurement of values.* Chicago: The University of Chicago Press.

13. Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*(2), 139–150.

14. Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction* (Volume IV, pp. 60–90). Princeton, NJ: Princeton University Press.

15. Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star & J.A. Clausen. (Eds), *Measurement and prediction* (pp. 362–412). Princeton, NJ: Princeton University of Press.

16. Mokken, R. J. (1971). *A theory and procedure of scale analysis*. The Hague: Mouton/Berlin: De Gruyter.

17. Rasch (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research. (Expanded edition, Chicago: University of Chicago Press, 1980).

18. Rasch, G. (1961). On general laws and meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the fourth Berkeley Symposium on mathematical statistics and probability* (pp 321-333). . Berkeley: University of California Press.

19. Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy, 14,* 58–94.

20. Engelhard, G. (1991). Thorndike, Thurstone and Rasch: A comparison of their approaches to item-invariant measurement. *Journal of Research and Development in Education, 24*(2), 45–60.

21. Engelhard, G. (1992). Historical views of invariance: Evidence from the measurement theories of Thorndike, Thurstone and Rasch. *Educational and Psychological Measurement, 52*(2) 275–292.

22. Engelhard, G. (1994). Historical views of the concept of invariance in measurement theory. In M. Wilson (Ed.), *Objective measurement: Theory into practice, Volume 2* (pp. 73–99). Norwood, NJ: Ablex.

23. Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific change.* Berkeley: University of California Press.

24. Kuhn, T. S. (1970). *The structure of scientific revolutions (2nd ed.).* Princeton, NJ: Princeton University Press.

25. Lakatos, I. (1978). *The methodology of scientific research programs.* Cambridge: Cambridge University Press.

26. Spearman, C. (1904). General intelligence objectively determined and measured, *American Journal of Psychology, 15*, 201-293.

27. Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

28. Baker, F. B., & Kim, S. (2004). *Item response theory: Parameter estimation techniques. Second edition, Revised and expanded.* New York: Marcel Dekker.

29. Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*, 557–585.

30. Wright, S. (1934). The Method of Path Coefficients, *The Annals of Mathematical Statistics, 5*(3), 161-215.

31. Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

32. Joreskog, K.G. (2007). Factor analysis and its extensions. In R. Cudeck & R.C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 47-77). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

33. De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer.

34. Messick, S. (1983). Assessment of children. In P. H. Mussen (Ed.), *Handbook of child psychology, volume 1: History, theory, and methods* (pp. 477–526). New York: Wiley.

35. Mosier, C. I. (1941). Psychophysics and mental test theory II: The constant process *Psychological Review, 48*, 235–249.

36. Mosier, C. I. (1940). Psychophysics and mental test theory: Fundamental postulates and elementary theorems. *Psychological Review, 47*, 355–366.

37. Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8*(4), 341–349.

38. Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*(1), 1-20.