Measurement, Explanation, and Invariance: Next Generation Invariant Measurement

George Engelhard, Jr.[1]

[1] The University of Georgia, Athens, GA 30602
gengelh@uga.edu

Abstract. Science is built upon the relationships between measurement, explanation, and invariance. This chapter emphasizes that the principles of invariance are essential to measurement and scientific objectivity more generally. Rasch measurement theory highlights the role of objectivity in creating and using invariant scales. In addition to this connection between measurement and invariance, science relies on discovering explanations that reflect stable or invariant relationships between well-defined variables. The scientific process seeks to discover invariant relationships across various conditions. Explanatory Item Response Models (EIRMs) are described in this chapter because these models can provide an approach for linking measurement and explanations based on principles of invariance. These measurement models can be enhanced in ways that provide a strong foundation to guiding the next generation of invariant measurement.

Keywords: Invariant measurement, explanatory Rasch models, next generation measurement.

1 Introduction

Measurement remains one of the essential aspects of life in the 21st century. It affects almost every aspect of our lives ranging from checking the weather to see if it is going to be a rainy day to personal measures of our health and well-being in the physician's office. Measurement permeates our interactions with others through commerce and other aspects of modern life.

My expertise focuses on educational and psychological measurement. It is safe to say that measurement is a major issue that includes the appropriate uses of tests in education for accountability to inform policy in our schools, for admission to colleges, and for information to parents and teachers about individual student learning. Unfortunately, measurement problems exist related to the use of most tests, and as emphatically pointed out be Ben Wright, "truth of the matter is that the so-called measurements are no damn good!" (Wright, 1968, p. 85). It is clear that measurement problems still remain, and that fundamental misunderstandings of the appropriate meaning and use of scores obtained from both cognitive (school achievement) and

© The Author(s) 2026
C. Y. Lye et al. (eds.), *Proceedings of the Pacific Rim Objective Measurement Symposium (PROMS 2025)*, Atlantis Highlights in Social Sciences, Education and Humanities 56,
https://doi.org/10.2991/978-94-6239-834-4_3

affective assessments (socioemotional characteristics of children) is a continuing problem.

Fortunately, the outlook for improvements in educational measurement are not necessarily as bleak as perceived by Wright (1968). There are some very basic steps that have the potential to lead to progress in our assessment systems that have been known for well over a century. A general goal of this chapter is to introduce the reader to ideas about next generation assessment systems.

Section Two of this chapter explores the concept of invariance and its implications for measurement theory. To explore the connections between measurement and invariance, I consider the measurement theory of Rasch (1960/1980,1961,1977). Next, I discuss the relationship between explanation and invariance with implications for examining structural relationships that underlies science broadly conceived (Godfrey-Smith, 2003; Nozick, 2001; Woodward, 2000, 2021). In the next section, I consider how measurement, invariance and explanation can be combined to create a set of basic guiding principles for explanatory Rasch models. Explanatory Rasch models include item and person covariates that can guide our perspectives on invariant structural relationships and Rasch measurement theory.

2 Measurement and Invariance

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 (Andrich, 2018, p. 79).

Rasch's goal was to develop models for measurement that support "individual-centered statistical techniques in which each individual is characterized separately and from which, given adequate data, the individual parameters can be estimated" (Rasch 1960/1980, p. xx). He also sought to develop the requirements for invariance related to specific objectivity. To accomplish these goals, Rasch started with a perspective on science that included the concept of objective comparisons. In his words (Rasch, 1964),

> Looking then for concepts [of measurement] that could possibly be taken as primary it seems worth-while to concentrate upon two essential characteristics of "scientific statements": 1. they are concerned with "comparisons" [and] 2. the statements are claimed to be "objective" [quotes in the original] (p.2)

As pointed out by Nozick (2001), objectivity in scientific statements depends on several features including accessibility, intersubjectivity, independence, and most importantly, invariance that implies the first three characteristics. 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. These four features of objectivity in science can be applied to measurement processes. There are close connections between Nozick's (2001) views on invariance in science, and Rasch's (1977) perspective on specific objectivity and invariant measurement.

Rasch identified objective comparisons as a key concept in general science, as well as specifically in measurement. Rasch defined objective comparisons as follows:

> The comparison between two stimuli [items] should be independent of which particular individuals were instrumental for the comparison; and it should also be independent of which other stimuli [items] within the considered class were or might also have been compared.
>
> Symmetrically, a comparison between two individuals should be independent of which particular stimuli [items] within the class considered were instrumental for the comparison; and it should also be independent of which other individuals were also compared, on the same or on some other occasion (Rasch 1977, pp. 331–332).

Rasch recognized the importance of identifying the conditions under which objective comparisons are invariant. He also introduced the idea of specific objectivity that stresses that invariance may not occur under all conditions. He argued for the importance of specifying a frame of reference to delineate the conditions for invariant measurement. Specifying frames of reference is related to Nozick's (2001) observation that it is important to notice both objective and invariant features, as well as to "notice which of its features are subjective and variant" (p. 102).

Invariant and variant features are two sides of the same coin in the sense that establishing invariance includes an exploration of sources of variance that may place limits on the consistency and generalizability of the measuring process. For example, in the measurement of temperature, researchers do not expect all instruments to function across the full range of temperatures (Chang, 2004) – we would use a different thermometer to measure a child's fever as compared to the temperature of molten steel. It is also important to note that the search for variant aspects implies anomalies that play a key role in the history of science (Andrich, 2002; Kuhn, 1970). It should also be noted that differential item functioning can be viewed from the perspective of invariance in the sense that the key question underlying DIF is whether or not items are functioning in a comparable way across subgroups of persons. DIF analyses can also

be viewed from this perspective as explanatory Rasch models that explore why certain items function differentially.

Rasch developed a philosophy of measurement that included the application of basic scientific concepts including objectivity and invariance. In his words, "the concept of *objectivity* raises fundamental problems in all sciences. For a statement to be scientific, "objectivity" is required" (Rasch 1964, p. 1). Rasch measurement theory represents scientific principles based on the concept of specific objectivity (invariant comparisons) applied to models for measurement. Rasch measurement theory aims to create invariant scales and provide objective measures.

In the next section, we turn our attention to explanation and invariance in science that has the potential to unite the concepts of measurement, explanation, and invariance.

3 Explanation and Invariance

> *Science aims at the discovery … of relationships that are invariant under some large or relevant range of circumstances (Woodward, 2021, p. 267).*

Explanation plays a key role in science (Godfrey-Smith, 2003). A major aim of science is to describe the structure of the world. Science also aims to explore why things happen. Although philosophers of science may disagree, it is our perspective that science has a strong explanatory role in addition to its descriptive role. Measurement is a key aspect of science that does not merely describe latent variables – it also plays a role in explaining relationships among variables.

Woodward (2000, 2021), who is a philosopher of science, has made a strong case for the connections between explanation and invariance. His philosophical perspective provides an opportunity for extending the principles of invariance to explore relationships between variables. According to Woodward (2000), "explanatory relationships are relationships that in principle can be used for manipulation and control in the sense that they tell us how certain (explanandum) variables would change if other (explanans) variables were to be changed or manipulated" (p. 198). He also stresses that the utility of an explanation depends on the degree and domain of invariance. In his words,

> A generalization is invariant if it is … change-relating and it is stable or robust in the sense that it would continue to hold under a special sort of change called an intervention (Woodward, 2000, p. 198).

Woodward (2000) argued that relationships between two or more variables are invariant if they continue to hold as other conditions vary. The changes over which the relationships are invariant define a domain of invariance. Invariance is a relative matter and includes the identification of the frames of reference under which relationships are

stable and consistent. It is also important to note that degrees of invariance may hold. Measuring instruments may not function in the same way under all conditions. In other words, invariance is not an all or nothing matter because generalizations may be invariant under some interventions and not invariant under others. The concept of domain invariance suggested by Woodward (2000) is similar to Rasch's concept of specific objectivity and frames of reference (Rasch, 1977).

Measurement theories including the Rasch model focus on the creation of an invariant continuum to represent a latent variable or construct. Structural models on the other hand consider connections between latent variables. Explanatory item response models seek to identify invariant relationships between variables. Structural ideas are not typically related to measurement issues. I believe that there are essential connections between measurement and structural analyses. In the next section, I focus on measurement, explanation and invariance with explanatory item response models as a promising methodology for next generation invariant measurement.

4 Measurement, Explanation, and Invariance: The Next Generation Invariant Measurement

It is evident that the two approaches, the descriptive measurement approach and the explanatory approach, can be combined. Thus, explanation is seen as complementary to measurement (De Boeck & Wilson, 2004, p. 3).

De Boeck and Wilson (2004) edited an influential volume that summarized much of the previous work on explanatory item response models. A key feature of EIRM is the addition of item and person covariates to item response models for measurement. The introduction of covariates for both items and persons provides the opportunity to examine structural relationships that can support explanatory analyses. Another key feature of EIRMs is that parameter estimation can be approached using generalized linear mixed models (McCulloch, 2003). The use of GLMMs opens up a host of options for linking measurement and explanation in the human sciences. EIRMs have been developed for a variety of item response models including the Rasch model (De Boeck & Wilson, 2004).

Rasch measurement theory can be used as an overarching framework for both descriptive and explanatory measurement models. Figure 1 provides an overview of common Rasch models for both dichotomous and polytomous data. The commonly used descriptive Rasch models include the dichotomous model (Wright & Stone, 1979), the two polytomous models: Rating Scale Model (Andrich, 1978) and the Partial Credit Model (1982). Each of these descriptive models can be extended with item and person covariates to form explanatory Rasch models that may not be as familiar.

Table 1. Descriptive and explanatory Rasch measurement models.

Models	Logit (η_{ni})	Person Covariate	Item Covariate
Descriptive Model			
Dichotomous Model	$\eta_{ni} = \theta_n - \delta_i$	No	No
Explanatory Models			
Linear Logistic Rasch Models	$\eta_{ni} = \theta_n + \sum_{k=1}^{K} \beta_k X_{ik}$	No	Yes
Latent Regression Rasch Models	$\eta_{ni} = (\sum_{j=1}^{J} \vartheta_j Z_{nj} + \theta_n^*) - \delta_i$	Yes	No
Combined Covariates Rasch Models	$\eta_{ni} = (\sum_{j=1}^{J} \vartheta_j Z_{nj} + \theta_n^*)) + (\sum_{k=1}^{K} \beta_k X_{ik})$	Yes	Yes

Note. η_{ni} represents logits, θ_n is the person location on the latent variable, δ_i is the item location, β_k is the item subset location, and ϑ_j is the person subgroup location. X is the matrix of item covariates, while Z is the matrix of person covariates.

Table 1 shows four Rasch models that can be classified based on whether item and person predictors are included. Descriptive models do not include any covariates (e.g., Dichotomous Model), Linear Logistic Rasch Models include item covariates, and the Latent Regression Rasch Models include person covariates. Finally, the Combined Covariates Rasch Models that include both item and person covariates.

The explanatory models can be used to explore relationships between variables. They can provide an approach for exploring causal relationships including ideas about explanation and invariance as described by Woodward (2000). They can be used to provide explanations of the influences of various covariates (items, persons) on the probability of positive responses.

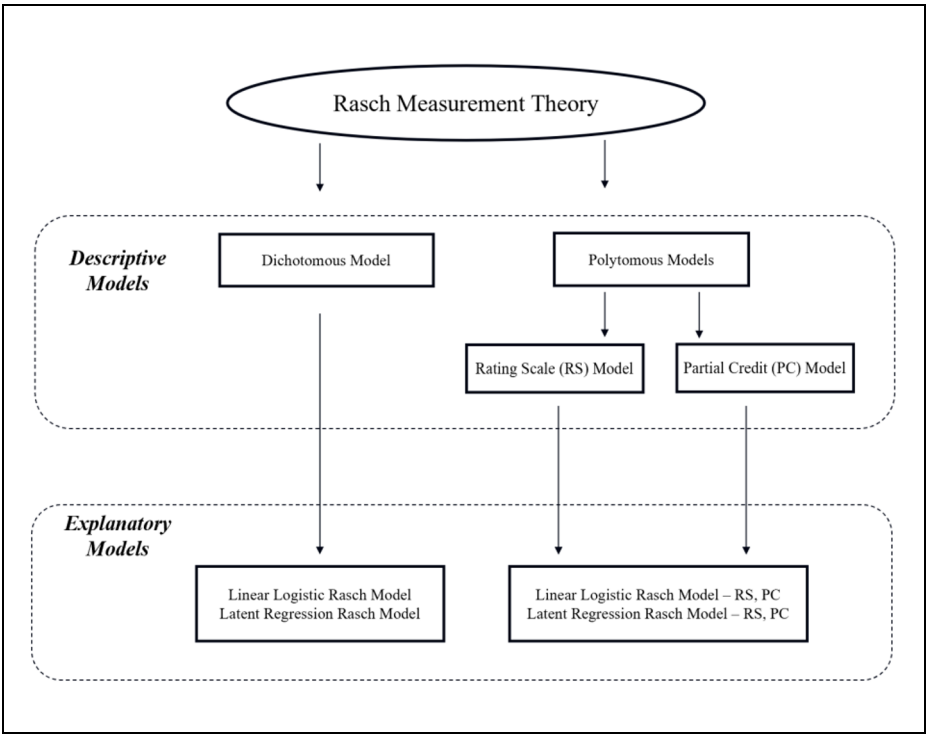


Fig 1. Overview of the descriptive and explanatory models.

The principles of invariant measurement are summarized in Figure 2. The main idea is that invariance is required for psychometrically sound measurement, and that Rasch measurement theory in this case can be used to reveal invariant scales. The four components of measurement constructs, items, persons and structural aspects all guide the essential principles for the next generation of invariant measurement defined by as follows:

- Constructs should be unidimensional and defined by latent variables,
- the measurement of persons should be invariant across items,
- item calibrations should be independent of specific persons, and finally,
- structural analyses with latent variables should be invariant across relationships among variables.

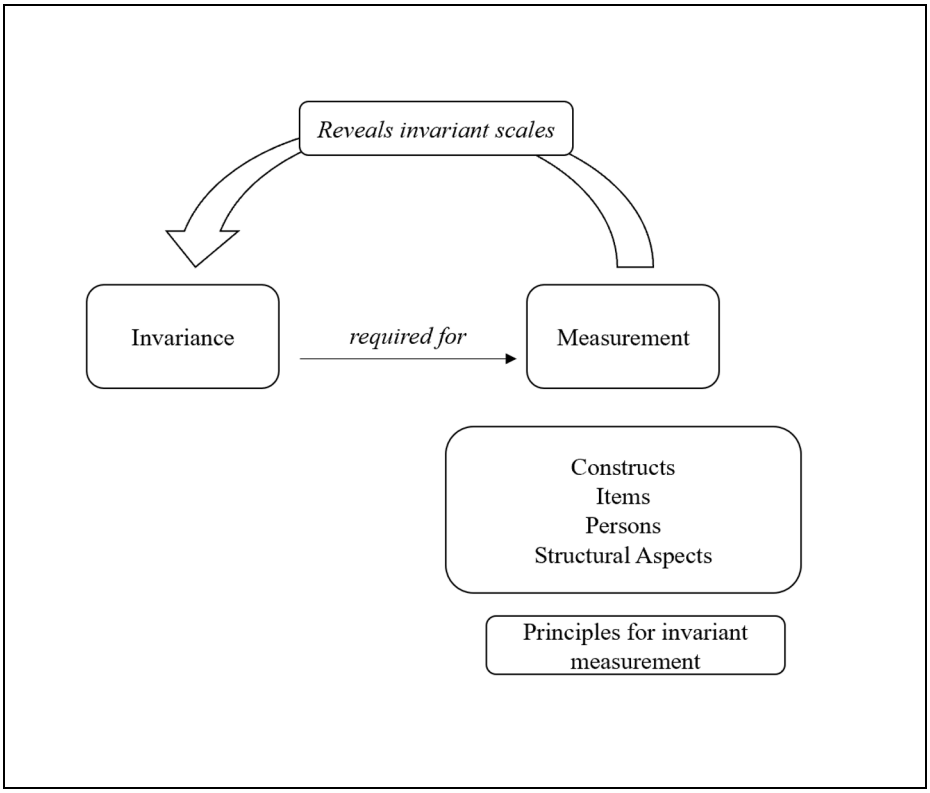


Fig 2. Principles of invariant measurement.

Figure 3 extends these principles to invariant measurement and relationships. Principles of invariant measurement are embedded under the broader idea that measurement facilitates explanation regarding structural relationships. The principles for invariant relationships are simply the expansion of the idea that structural analyses with latent variables should be invariant across relationships among variables. Measurement is an important aspect of science that does not merely describe latent with invariant measurement, but also plays a role in explaining the relationships among variables.

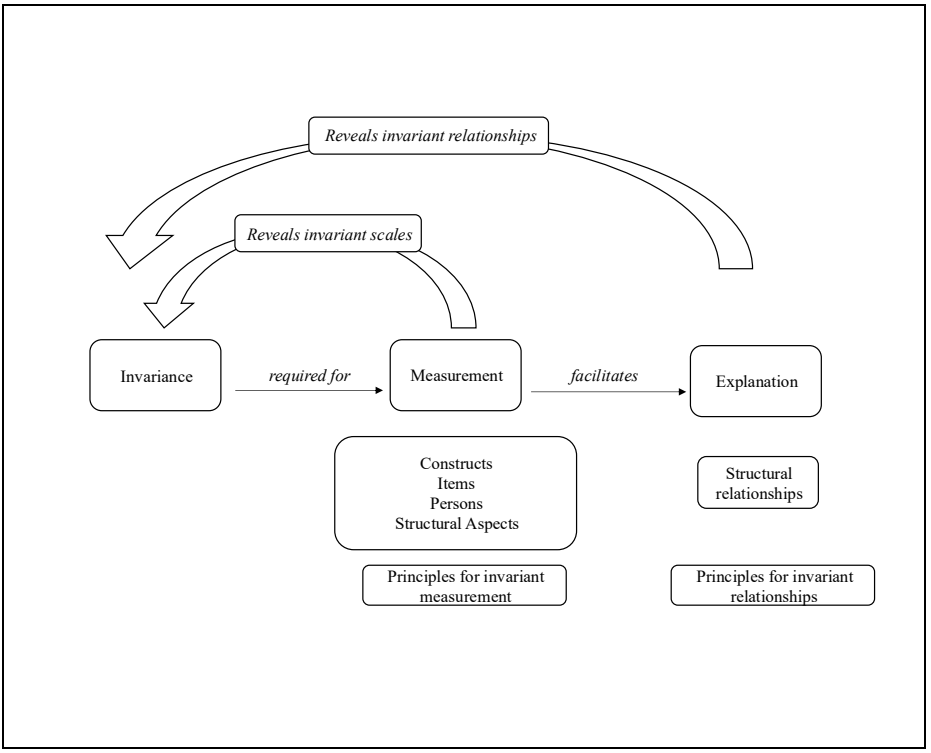


Fig 3. Principles of invariant measurement and relationships.

5 Discussion

Have you ever wondered why progress is so slow? Something must be wrong. I believe progress will continue to be slow until we find a way to work with measurements which are objective, measurements which remain a property of the person measured regardless of the test items he answers or the company he keeps (Wright, 1968, p. 101).

This chapter has explored the foundational and emerging intersections between measurement, explanation, and invariance, proposing a conceptual and methodological framework for the next generation of invariant measurement. At its core, this framework emphasizes the scientific imperative of invariance — not only as a technical requirement of measurement models but as a philosophical foundation for objectivity and explanatory power in the human sciences.

Drawing from Rasch's notion of specific objectivity and Woodward's account of explanatory invariance, meaningful measurement and explanation both rest on identifying relationships that are stable under a range of conditions. Rasch measurement

theory, particularly through its probabilistic formulation of invariant comparisons, offers a rigorous way to represent and test for objectivity in measurement. At the same time, Woodward's interventionist theory of explanation introduces a complementary lens for understanding how relationships between variables must demonstrate invariance to be considered causal or explanatory.

The integration of explanatory item response models (EIRMs) represents a pivotal advancement in this regard. These models bridge the traditional gap between descriptive measurement and structural explanation by incorporating person and item covariates directly into the measurement model. This development aligns with broader scientific aims: not merely to describe performance or traits, but to understand the mechanisms behind them. As shown in the classification of Rasch models in Table 1, this integration allows researchers to systematically explore how invariant relationships emerge across different conditions and subgroups, enhancing both the validity and the interpretability of measurement outcomes.

A central argument advanced in this chapter is that the future of measurement lies in the deliberate fusion of descriptive and explanatory frameworks. By embracing explanatory Rasch models, researchers are better equipped to investigate not just whether measures are invariant, but *why* and *for whom* they are invariant or variant. This perspective transforms differential item functioning (DIF) analysis from a diagnostic tool into an explanatory endeavor, revealing substantive insights about population heterogeneity and the functioning of constructs across contexts.

However, this next generation of invariant measurement also requires a conceptual shift. It compels us to think critically about frames of reference, domains of generalization, and the contingent nature of objectivity itself. Invariance is not an all-or-nothing condition, but a property that holds within specific boundaries. Identifying and articulating these boundaries is as much a philosophical as it is a statistical task, reinforcing the need for theory-driven measurement practices.

By situating measurement within the broader goals of explanation and invariance, we gain a more integrated and powerful approach to understanding human behavior and cognition. The Rasch model, extended through explanatory modeling, offers a robust foundation for this endeavor. As measurement continues to influence high-stakes decisions in education, psychology, and beyond, embracing these principles is not only scientifically sound — it is ethically imperative. In summary, the hope is that progress in the next generation invariant measurement will provide the integration of measurement theory with explanatory models to deepen our understanding of the world, and to discover invariant relationships in the human sciences.

Acknowledgments. This chapter is based on my presidential address presented at the Pacific Rim Objective Measurement Society meeting in Singapore (July 2025).

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573.
2. Andrich, D. (2002). Understanding Resistance to the Data-model Relationship in Rasch's Paradigm: A Reflection for the Next Generation. *Journal of Applied Measurement, 3*(3), 325-359.
3. Andrich, D. (2018). A Rasch measurement theory. In F. Guillemin et al. (Eds), *Perceived Health and Adaptation in Chronic Disease* (pp. 66-91). Routledge.
4. De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach.* Springer.
5. Godfrey-Smith, P. (2003). *Theory and reality: An introduction to the philosophy of science.* Chicago: The University of Chicago Press.
6. Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Princeton, NJ: Princeton University Press.
7. Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149–174.
8. McCulloch CE. (2003). Generalized linear mixed models (GLMMs). In *Generalized linear mixed models*. Beechwood OH and Alexandria VA: Institute of Mathematical Statistics and American Statistical Association, 2003. pp. 28–33.
9. Nozick, R. (2001). *Invariances: The structure of the objective world.* Harvard University Press.
10. 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).
11. 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.
12. Rasch, G. (1964). Objective comparisons. In *Lectures given at the UNESCO Seminar, Voksenasen, Oslo.*
13. 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.
14. Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science, 51*(2), 197-254.
15. Woodward, J. (2021). *Causation with a human face: Normative theory and descriptive psychology.* Oxford University Press.
16. 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.
17. Wright, B. D., & Stone, M. (1979). *Best test design: Rasch measurement*. Chicago: MESA Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

