2

CONSTRUCTING A RASCH SCALE

Rasch measurement theory can be used as the basis for solving a variety of measurement problems. In the previous chapter, we proposed using four components as the basis for developing a Rasch scale: constructing, evaluating, using, and maintaining a scale. Each of these components maps to a common measurement problem encountered in the social sciences. These problems are the definition of a latent variable, measurement invariance (e.g., differential item functioning), interchangeability of items (e.g., test equating), and standard setting (e.g., setting a cut score based on performance standards). This chapter discusses the first component in relation to the measurement problem of constructing a scale that can be used to define a latent variable based on the principles of Rasch measurement theory.

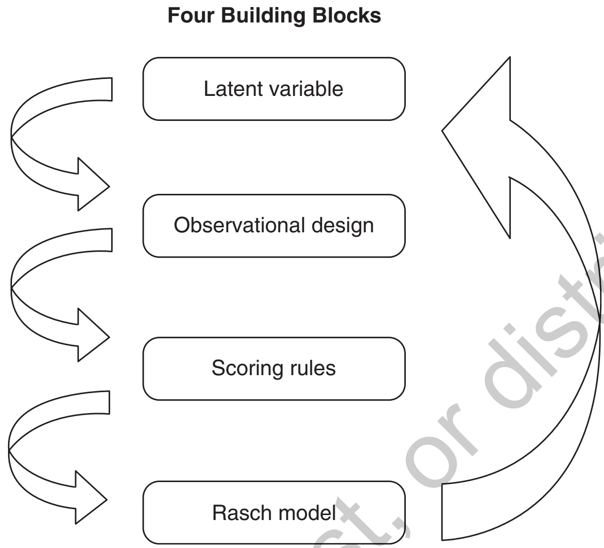
Once a researcher has decided to use a scale to represent an important latent variable (construct), the first step is to begin the construction of the scale. The construction of a scale involves several steps. The approach used here is based on modifications to the constructing measures approach suggested by Wilson (2005). The essential building blocks for constructing a latent variable scale include specification of the latent variable (construct) to be measured, creation of an observational design (e.g., items or questions), development of a set of scoring rules, and application of the Rasch model to observed data to create an empirical Wright map.

This chapter begins with a description of the building blocks that can be used to create a Rasch scale. Next, we use an international scale constructed to measure individual experiences of food insecurity as an illustrative example (Cafiero, Viviani, & Nord, 2018). This includes the use of the Rasch model to evaluate a small data set. Finally, we provide a summary of the chapter and highlight key points.

2.1 Building Blocks for a Rasch Scale

The creation of a meaningful and useful scale based on Rasch measurement theory is described using the four building blocks as shown in Figure 2.1 (latent variable, observational design, scoring rules, and

Figure 2.1 Building Blocks for Constructing a Rasch Scale



Rasch model). The specific question addressed is: What are the essential steps for constructing a Wright map based on Rasch measurement theory? Each building block is described in detail in this section.

Latent Variable

The first building block in creating a scale starts with the initial imagery of a latent variable (Lazarsfeld, 1958). It is important to note that we are creating a unidimensional scale. Unidimensional scales play key roles because "they coincide with the use of unidimensional language in social science theories—language that is intended to clarify the meaning of those theories" (McIver & Carmines, 1981, p. 86). The concept of unidimensionality is relative in essence, e.g., if the items measure both mathematical and reading components to the same degree, the items may be scalable on a unidimensional scale (Andrich, 1985; Lumsden, 1957).

The measurement of food insecurity is used for our illustrations in this chapter. The purpose of the Food Insecurity Experience (FIE) scale is to obtain evidence regarding food insecurity in a global context

(Cafiero et al., 2018). Overall, food insecurity is defined very generally as follows:

> *Food security is said to exist when all people, at all times, have physical, social and economic access to sufficient safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life.*
>
> Declaration of the World Summit on Food Security, World Summit on Food Security, Rome, November 16–18, 2009

Cafiero et al. (2018) recognize that:

> *although food security is inherently multi-dimensional, one critical dimension is continued access to adequate food. The United Nations Food and Agriculture Organization (FAO) has undertaken a project called Voices of the Hungry (VoH) to develop and support a survey-based experiential measure of access to food, called the Food Insecurity Experience Scale.*
>
> (p. 146)

A similar approach is used in the United States for measuring food insecurity at the household level (Coleman-Jensen, Rabbitt, Gregory, & Singh, 2015). The FIE scale measures food insecurity conceived as the "condition of not being able to freely access the food one needs to conduct a healthy, active and dignified life … resulting from the inability to access food due to lack of money or other resources" (p. 147). Narrowing down the broad theoretical definition of food insecurity to focus on a single dimension, access to adequate food guides the creation of a hypothesized Wright map for measuring food insecurity.

The hypothesized Wright map for food insecurity is shown in Figure 2.2. There are several features that should be noted in Figure 2.2. First, the scale for measuring food insecurity is represented by a line. This line reflects a theoretical continuum that ranges from low food insecurity to high food insecurity. This continuum includes qualitative descriptions of ordered levels of food insecurity for persons that range from mild food insecurity through moderate food insecurity to severe food insecurity. These levels represent substantively important distinctions that are used by policy makers who address issues of food insecurity around the world. Second, the line representing the latent variable of food insecurity is defined by an ordered set of items that reflect the experiences of food insecurity. These ordered items reflect a

Figure 2.2 Hypothesized Wright Map for Measuring Food Insecurity

Food Insecurity Levels		**Food Insecurity Experiences (Items)**
High food insecurity		*Hard to affirm*
		Experiencing hunger
Severe food insecurity		
		Reducing quantities, skipping meals
Moderate food insecurity		
		Compromising quality and variety of food
Mild food insecurity		
		Worrying about ability to obtain food
Low food insecurity		*Easy to affirm*

Source: Based on Coleman-Jensen et al. (2015).

hypothesized expectation regarding the order of items from easy to affirm with a Yes response (i.e., worrying about ability to obtain food) to hard to affirm (i.e., experiencing hunger).

Observational Design

After a researcher has created a hypothesized Wright map (e.g., Figure 2.2), the next step is the creation of a set of observable indicators or items to represent food insecurity. Observational designs frequently include various item classifications and domains that guide the creation of specific items. A classic example is the creation of educational achievement tests using item classifications based on Bloom's Taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Lane, Raymond, and Haladyna (2016) provide a detailed consideration of various guidelines for test and item development for assessments used in a variety of assessment contexts.

Table 2.1 shows the eight items that are included in the FIE scale (Cafiero et al., 2018) used in this book. These items reflect the observational design used to represent food insecurity. This scale is based on a careful consideration of previous scales that have been used to measure household and individual food insecurity around the world, such as the US Household Food Security Survey Module (HFSSM), the Escala Brasileira de Insegurança Alimentar (EBIA), the Escala

Table 2.1 English Version of the Food Insecurity Experience Scale

Item	*Questions*	*Label*
1	During the last 12 months, was there a time when you were worried you would not have enough food to eat because of a lack of money or other resources?	Worried
2	Still thinking about the last 12 months, was there a time when you were unable to eat healthy and nutritious food because of a lack of money or other resources?	Healthy
3	Was there a time when you ate only a few kinds of foods because of a lack of money or other resources?	Few Foods
4	Was there a time when you had to skip a meal because there was not enough money or other resources to get food?	Skipped
5	Still thinking about the last 12 months, was there a time when you ate less than you thought you should because of a lack of money or other resources?	Ate Less
6	Was there a time when your household ran out of food because of a lack of money or other resources?	Ran Out
7	Was there a time when you were hungry but did not eat because there was not enough money or other resources for food?	Hungry
8	During the last 12 months, was there a time when you went without eating for a whole day because of a lack of money or other resources?	Whole Day

Note. Respondents answer yes or no to these questions (http://www.fao.org/in-action/voices-of-the-hungry/fies/en/).

Latinoamericana y Caribeña de Seguridad Alimentaria (ELCSA), the Escala Mexicana de Seguridad Alimentaria (EMSA), and the Household Food Insecurity Access Scale (HFIAS) (Coleman-Jensen et al., 2015). The selection of items also included a consideration of the interpretability of these items and conditions across different cultures and contexts by the creators of the scale.

Scoring Rules

Scoring rules specify how the person responses are coded. For our example, the responses to the eight items are simply scored dichotomously (Yes = 1 and No = 0). A response of Yes indicates an affirmative response to the item, and it leads to a higher level of food insecurity. The items are combined into sum scores with higher sum scores indicating more severe food insecurity of persons.

There are other examples of scoring rules that include different types of rating scales, such as the rating scale model (Andrich, 2016) and partial credit model (Masters, 2016) that are part of the Rasch family of models (Wright & Masters, 1984). It is also possible to combine categories in ways that reflect different scoring rules. Engelhard and Wind (2018) provide guidance on different models for polytomous or rating data using different Rasch models.

Rasch Model

The final step is the use of a measurement model to link the observed responses to items and persons based on their locations on a latent variable scale. Rasch (1960/1980) started with a simple idea that a person's response to an item depends on the difficulty of the item and the ability of the person. He selected a probabilistic model based on the logistic response function because of its desirable properties related to specific objectivity (i.e., invariant measurement). The dichotomous Rasch model in its modern form can be written as:

$$\phi_{ni1} = \frac{\exp(\theta_n - \delta_{i1})}{1 + \exp(\theta_n - \delta_{i1})} \tag{2.1}$$

The Rasch model in Equation 2.1 can be viewed as an operating characteristic function that relates the differences between locations of persons (θ) and items (δ) on a latent variable to the probability of success or affirmation on a dichotomous item. This distance reflects a comparison between each person and item that predicts a probability of a positive response for each person on an item.

We find it useful to conceptualize the Rasch model as a probabilistic version of Guttman scaling (Andrich, 1985). In his words:

> *… technical parallels between the SLM [Rasch model] and the Guttman scale are not a coincidence. The connections arise from the same essential conditions required in both, including the requirement of invariance of scale and location values with respect to each other.*
> (Andrich, 1988, p. 40)

Engelhard (2005) describes Guttman scales in detail as deterministic- and ideal-type models. Table 2.2 provides a simple example to show the

Table 2.2 Illustration of Guttman (Perfect) and Rasch (Probabilistic) Item Response Patterns

	Panel A				*Panel B*			
	Perfect Pattern (Guttman)				*Probabilistic Pattern (Rasch)*			
Person Scores	*A Easy*	*B*	*C*	*D Hard*	*A Easy*	*B*	*C*	*D Hard*
4	1	1	1	1	0.98	0.95	0.75	0.65
3	1	1	1	0	0.95	0.75	0.65	0.45
2	1	1	0	0	0.75	0.65	0.45	0.34
1	1	0	0	0	0.65	0.45	0.34	0.25
0	0	0	0	0	0.45	0.34	0.25	0.15

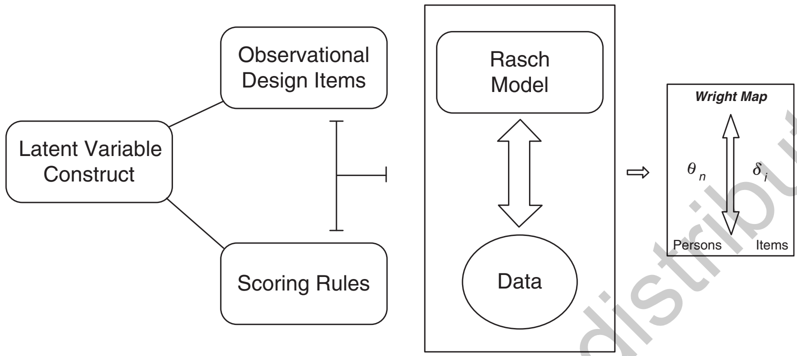
Note. These values are used for illustration.

triangular pattern of Guttman scales. Panel A in Table 2.2 shows this pattern when items are ordered from easy to hard and the persons are ordered based on their sum scores. A similar triangular pattern appears with the Rasch probabilities when items are calibrated, and persons measured on the latent variable scale. This is illustrated in Panel B of Table 2.2. Probabilities that are greater than 0.50 indicate that a person is more likely to affirm an item, although the stochastic nature of the process also recognizes that a person may not affirm the item. This is in contrast to Guttman scaling that defines a perfect scale with a deterministic model.

Figure 2.3 provides a flowchart of the progress from the idea of a latent variable (construct) to a Wright map. Once we have a general conception of the latent variable, the next steps are to create an observational design and scoring rules that guide the creation of items and indicators that we plan to use to define the latent variable. Next, the items are administered to a sample of persons to collect responses, and the observed data are analyzed with the Rasch model. The Rasch model is the measurement model that connects the observed data to the visual representation of our latent variable on the Wright map. Finally, the Wright map displays the location of both items and persons empirically on the latent variable scale.

There are numerous excellent introductions to the technical details for estimating item and person locations on the latent variable based on the Rasch model. We highly recommend Baker and Kim (2004) as an advanced text addressing methods for estimating the parameters of item

Figure 2.3 Progression From Latent Variable (Construct) to Wright Map



response theory (IRT) models. Baker and Kim (2017) have also published R syntax for a variety of IRT models.

In this book, we use the Facets computer program (Linacre, 2018a) to estimate the parameters of the Rasch model. The syntax for the Facets computer program and also an R program (Everyone's Rasch Measurement Analyzer—ERMA) is available online (https://study.sagepub.com/researchmethods/qass/engelhard-rasch-models). Sample data sets are also available online.

2.2 Illustrative Analyses

In this section, we use a dichotomous Rasch model to analyze the FIE data. The dichotomous responses of 40 persons to 8 items (FIE scale) are shown in Table 2.3. These data reflect food insecurity experiences for the United States. Table 2.4 presents the summary statistics for the eight items from a Rasch analysis of these data responses using the Facets computer program (Linacre, 2018a). The first column is the item number, while the second column provides a short label describing the content of the item.

Column 3 shows the proportion of Yes responses to each item. The items range from easy to affirm (Item 3—Few Foods) to hard to affirm (Item 8—Whole Day). The next two columns indicate the calibration of the items in logits (and standard errors) that represent the locations of the items on the Wright map. The Wright map is shown in Chapter 1 (Figure 1.6). The next four columns report how well the observed data

Table 2.3 Food Insecurity Experience Data

	Items							
Person	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1	1	0	1	0	1	0	0	0
2	0	0	1	1	1	0	1	0
3	0	0	1	0	1	0	0	0
4	0	0	0	0	0	1	1	0
5	0	0	1	0	0	0	0	0
6	0	0	1	0	0	0	0	0
7	1	1	1	0	1	0	0	0
8	0	0	1	0	0	0	0	0
9	1	1	1	0	0	1	0	0
10	0	0	1	0	0	0	0	0
11	0	1	1	0	0	1	0	0
12	0	1	0	0	0	1	0	0
13	1	1	1	1	1	1	0	0
14	0	1	0	0	0	0	0	0
15	0	1	0	0	0	0	0	0
16	0	0	1	0	0	0	0	0
17	0	1	1	1	1	1	1	0
18	0	1	1	0	0	0	0	0
19	0	0	1	0	0	0	0	0
20	0	1	1	0	0	0	0	0
21	0	1	1	1	0	1	0	0
22	1	1	1	1	1	1	1	0
23	1	1	0	1	1	1	1	1
24	0	1	1	0	1	0	0	0
25	1	1	1	1	0	0	1	1
26	1	1	1	1	0	1	0	0
27	1	0	1	0	1	0	0	0
28	0	0	1	1	0	0	1	0
29	0	1	1	1	1	1	1	1
30	0	0	0	0	1	0	0	0

Table 2.3 *(Continued)*

Person	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
	Items							
31	0	1	0	0	0	0	0	0
32	0	1	0	0	0	0	0	0
33	1	1	1	0	1	1	1	0
34	1	0	0	0	0	0	0	0
35	1	1	1	1	1	0	0	0
36	0	0	0	1	1	0	1	0
37	1	1	1	1	1	1	0	0
38	0	1	0	0	0	0	0	0
39	0	1	1	0	0	0	0	0
40	0	0	1	0	0	0	0	0

Note. Items are scored as follows: 0 = No, 1 = Yes.
Source: Based on Coleman-Jensen et al. (2015).

Table 2.4 Summary Statistics for Items (Ordered by Proportion of Yes Responses)

Item	*Label*	*Proportion of Yes Responses*	*Measure*	*S.E.*	*Mean Squares Infit*	*Mean Squares Outfit*	*Fit Category Infit*	*Fit Category Outfit*
8	Whole Day	0.08	2.85	0.67	0.74	0.23	A	B
7	Hungry	0.25	0.81	0.46	1.04	0.81	A	A
1	Worried	0.33	0.21	0.43	1.02	1.16	A	A
4	Skipped	0.33	0.21	0.43	0.69	0.47	A	B
6	Ran Out	0.33	0.21	0.43	0.92	0.71	A	A
5	Ate Less	0.40	−0.32	0.41	1.03	0.93	A	A
2	Healthy	0.60	−1.59	0.39	1.20	0.98	A	A
3	Few Foods	0.73	−2.38	0.41	1.12	4.42	A	D

Note. Fit categories: A ($0.50 \leq MSE < 1.50$), B ($MSE < 0.50$), C ($1.50 \leq MSE < 2.00$), D ($MSE \geq 2.00$).
MSE, mean square error.

Table 2.5 Fit Category Based on Mean Square Error (*MSE*)

MSE	*Interpretation*	*Fit Category*
$0.50 \leq MSE < 1.50$	Productive for measurement	A
$MSE < 0.50$	Less productive for measurement, but not distorting of measures	B
$1.50 \leq MSE < 2.00$	Unproductive for measurement, but not distorting of measures	C
$2.00 \leq MSE$	Unproductive for measurement, and distorting of measures	D

fit the Rasch model. The Infit statistics are sensitive to unexpected responses to items that are close to person locations, while the Outfit statistics are sensitive to responses to items that are located farther from person locations. Table 2.5 shows a framework suggested by Engelhard and Wind (2018) to further categorize the items.

Using this framework, none of the items misfit based on the Infit statistics, while Item 3 (Healthy) fits based on the Infit statistic and misfits based on the Outfit statistics. Further details on the Infit and Outfit statistics will be discussed in more detail in Chapter 3.

Table 2.6 provides similar information for persons. These analyses provide a validation of the responses of each person. Food insecurity ranges from high for Person 22 who responds "yes" to 88% of the items to low for Person 34 who responds "yes" to 13% of the items. As with the items, the measures indicate the location of the persons in logits on the Wright map (Figure 1.6).

Based on the Infit statistics, a summary of the fit categories for the persons are as follows: A (75.0%), B (10.0%), C (7.5%), and D (7.5%). The fit categories based on the Outfit statistics are as follows: A (37.5%), B (47.5%), C (5.0%), and D (10.0%). Misfitting persons will be discussed in more detail in Chapter 3.

Rasch measurement theory provides the connection between the observed data and the creation of a scale including location parameters for items and persons. A Wright map provides the visual outcome of this process. There are two representations for a Wright map with the hypothesized map for food insecurity (Figure 2.2) and the empirical map shown in Chapter 1 (Figure 1.6).

Table 2.6 Summary Statistics for Persons (Ordered by Proportion of Yes Responses)

				Mean Squares		*Fit Category*	
Person	*Prop Yes*	*Measure*	*S.E.*	*Infit*	*Outfit*	*Infit*	*Outfit*
22	0.88	2.63	1.23	0.36	0.16	B	B
29	0.88	2.63	1.23	1.80	1.61	C	C
23	0.88	2.63	1.23	2.02	9.00	D	D
13	0.75	1.47	0.96	0.59	0.41	A	B
37	0.75	1.47	0.96	0.59	0.41	A	B
17	0.75	1.47	0.96	0.81	0.64	A	A
33	0.75	1.47	0.96	0.81	0.64	A	A
25	0.75	1.47	0.96	2.02	1.83	D	C
35	0.63	0.67	0.85	0.72	0.54	A	A
26	0.63	0.67	0.85	0.88	0.72	A	A
7	0.50	−0.02	0.83	0.69	0.55	A	A
9	0.50	−0.02	0.83	0.87	0.68	A	A
21	0.50	−0.02	0.83	0.87	0.68	A	A
2	0.50	−0.02	0.83	1.40	1.35	A	A
24	0.38	−0.72	0.86	0.54	0.44	A	B
11	0.38	−0.72	0.86	0.72	0.61	A	A
1	0.38	−0.72	0.86	1.16	0.95	A	A
27	0.38	−0.72	0.86	1.16	0.95	A	A
28	0.38	−0.72	0.86	1.49	1.40	A	A
36	0.38	−0.72	0.86	2.13	2.13	D	D
39	0.25	−1.51	0.95	0.40	0.29	B	B
18	0.25	−1.51	0.95	0.40	0.29	B	B
20	0.25	−1.51	0.95	0.40	0.29	B	B
3	0.25	−1.51	0.95	0.92	0.68	A	A
12	0.25	−1.51	0.95	1.39	1.21	A	A
4	0.25	−1.51	0.95	2.16	2.49	D	D
6	0.13	−2.60	1.18	0.55	0.24	A	B
8	0.13	−2.60	1.18	0.55	0.24	A	B
10	0.13	−2.60	1.18	0.55	0.24	A	B
5	0.13	−2.60	1.18	0.55	0.24	A	B

Table 2.6 *(Continued)*

				Mean Squares		*Fit Category*	
Person	*Prop Yes*	*Measure*	*S.E.*	*Infit*	*Outfit*	*Infit*	*Outfit*
16	0.13	−2.60	1.18	0.55	0.24	A	B
19	0.13	−2.60	1.18	0.55	0.24	A	B
40	0.13	−2.60	1.18	0.55	0.24	A	B
14	0.13	−2.60	1.18	1.04	0.48	A	B
15	0.13	−2.60	1.18	1.04	0.48	A	B
31	0.13	−2.60	1.18	1.04	0.48	A	B
32	0.13	−2.60	1.18	1.04	0.48	A	B
38	0.13	−2.60	1.18	1.04	0.48	A	B
30	0.13	−2.60	1.18	1.52	1.39	C	A
34	0.13	−2.60	1.18	1.62	2.25	C	D

Note. Fit categories: A ($0.50 \leq MSE < 1.50$), B ($MSE < 0.50$), C ($1.50 \leq MSE < 2.00$), D ($MSE \geq 2.00$).
MSE, mean square error.

2.3 Summary

This chapter introduces the essential steps of scale construction based on Rasch measurement theory. The construction of a Rasch scale can be reflected by a flowchart that is shown in Figure 2.3. The first step includes the conceptual formation of a latent variable that is a focus of theory and practice within a broader substantive theory. We use the example of food insecurity as our focal latent variable. The next step guides our selection of items and observations for the design of the scale that is used to operationally define the latent variable. As an illustrative example, eight items (Table 2.1) are used to define the FIE scale. Meanwhile, a set of scoring rules is developed to code the observations onto an ordinal scale. The responses to the FIE scale are scored dichotomously (0 = no, 1 = yes). The last step links the Rasch model to observed data that are collected based on the responses of persons to the items. The Rasch model connects the observed data to the calibration of the items (location of items on the scale) and the measurement of persons (location of persons on the scale). The outcome of this step includes the creation of a Wright map that shows the simultaneous location of persons and items on the Rasch scale.

The empirical Wright map for the illustrative data is shown in Figure 1.6. It is important to remember that our goal is to ultimately create a scale that is validated for its intended purposes and uses. It is also important that the research community accepts the scale as a consensus view of the construct being measured by the scale. We think of this process as being organized around a Wright map with two aspects: a hypothesized Wright map and an empirical Wright map. The FIE scale provides a good example of a scale that has been widely recognized, and it is used throughout the world to measure food insecurity (Cafiero et al., 2018).

A Rasch scale meets the requirements of invariant measurement when good model-data fit is obtained. One point of confusion in the literature on psychometrics is the failure to adequately distinguish between the unobservable latent variable that is hypothesized a priori (hypothesized Wright map) and the empirical analyses of model-data fit to determine whether or not our intentions are realized in our specific data set. We view the examination of model-data fit as part of the evaluation of whether or not a successful scale for our latent variable (empirical Wright map) has been constructed. Chapter 3 describes in more detail the steps for evaluating model-data fit for a Rasch scale including item and person fit analyses.