POLYTOMOUS ITEM RESPONSE THEORY MODELS

Chapter 1 – Introduction

Measurement Theory

Mathematical models have been found to be very useful tools in the process of human enquiry. It can be argued that their use is what has driven the engine of science for the past century and has made it such a successful enterprise (Wright, 1997). Bartholomew (1987) suggests that a mathematical model elucidates the conceptual grounding and explicates the framework for a project and in doing so provides a context from which to conduct analyses and interpret results. It should be noted that here and in the remainder of the text we use the term mathematical model in its broadest sense and not, as it is sometimes used, as a proxy for deterministic models.

At a basic level, mathematical models can provide a means to quantify phenomena of interest. The process of counting objects involves the use of a relatively simple and direct mathematical model. Simple counts play a large role in the social sciences, but the usefulness of this approach is less certain here than in areas such as commerce and engineering. For example, Thorndike (1904) noted the difficulty of measuring as simple a skill as spelling, since spelling proficiency depends not only on how many words a person can spell (simple count) but also on how difficult those words are to spell.

Problems with simple mathematical models in the social sciences have led to the development of more appropriate models. In psychology, one of the earliest formalizations of a measurement theory for mental tests (Cattell, 1890), or psychological phenomena generally, is a set of mathematical models that have come to be called classical test theory (CTT). CTT derives from the pioneering work of Spearman (1904) and is built on measurement concepts borrowed from the physical sciences (Mislevy, 1996; see e.g., Nichols, 1998). One of the central concepts that CTT has borrowed from measurement in the physical sciences is the idea of errors in measurement. Indeed, it is partly because there obviously are, potentially large, errors in psychological measurement that test theories are required (Lord & Novick, 1968). Measurement theory is also needed because the phenomena that psychologists study (e.g., traits) are not themselves directly measurable and must be studied indirectly, through the measurement of other observable phenomena (Lord & Novick, 1968).

Ultimately, difficulties with testing the assumptions of CTT and with applying the resulting model in practice have led to the development of alternative measurement models. These models are essentially extensions and liberalizations of the classical theory (Brennan, 1998).

Item Response Theory

Item Response Theory (IRT) is an extension of CTT with mathematical roots that run deep in psychology, in the work of Fechner in the 1860s (Baker, 1992) and in Thurstone's (1925) early work (Bock, 1997a; Weiss, 1983). In IRT these mathematical roots form an item-based test theory which itself has roots in the psychological measurement work of Binet, Simon and Terman as far back as 1916 (Baker, 1992; Bock, 1997a). The formal basis of IRT as an item based test theory is generally attributed to the work of Lawley (1943; see Baker, 1992; Weiss, 1983). His pioneering work was in turn significantly expanded by Lord (1952) who also formalized IRT's role as an extension of the classical theory (Baker, 1992). Subsequent work by Lord (1980; Lord & Novick, 1968) and Birnbaum (1968) has been instrumental in establishing an understanding and acceptance of IRT among psychological measurement practitioners. The Danish mathematician Rasch (1960; 1966; 1977) played a similarly influential role by separately developing a specific class of IRT models and showing that it had a number of highly desirable features.

General treatments of IRT can be found in Hambleton and Swaminathan (1985) and Andrich (1988a). The mathematical foundation of IRT is a function that relates the probability of a person responding to an item in a specific manner to the standing of that person on the trait that the item is measuring. In other words, the function describes, in probabilistic terms, how a person with a higher standing on a trait (i.e., more of the trait) is likely to provide a response in a different response category to a person with a low standing on the trait. This mathematical function has a pre-specified form (usually a logistic ogive) and is now generally referred to as an item response function (IRF).

The main advantage of IRT is the fact that the item location (b) and the person trait level (θ) are indexed on the same metric. Therefore, when a person's trait level is higher than the item location on the trait continuum, that person is more likely than not to provide a trait-indicating (positive, or true) response. The converse is true when a person's trait level is below the item location.

In typical discussion of dichotomous IRT models the item category that represents a positive response, and is subsequently coded 1, is described as indicating a "correct" response to an item (the alternative category, coded 0, indicates incorrect responses). Furthermore, the item location parameter b, is commonly referred to as the item difficulty parameter.

References to correct responses and item difficulties lose their common meaning in the context of tests that attempt to measure a respondent's typical performance rather than their maximum performance. Measures of typical performance (MTPs) include, for example, tests used in personality, attitude, and interest measurement – what might be called predilection measurement in contrast to aptitude measurement. While the language that has built up around IRT has a common meaning that is out of place in typical performance measurement, the mathematical models that comprise IRT are quite neutral about their application to either maximum or typical performance measures.

Applying the IRT Model

The mechanics of IRT can most easily be presented in terms of a dichotomous model, that is, a model for items with only two response alternatives. Typically, such items require responses that are either correct or incorrect. They can also be personality-type measurement items with, for example, true or false response options. It was just noted earlier the central feature of IRT is that it specifies a way to model the probability of a particular response to a test item with respect to a continuous, underlying trait. This is typically accomplished by means of a monotonic function: the IRF.

The IRF shown in Figure 1.1 is a function that reflects the probability of selecting a positive (correct or keyed) response to an item. It can more generally be thought of as reflecting the probability associated with moving from one response category to the next, along the entire trait continuum. In other words, the function depicts the probability of making the transition from responding in one category to responding in the next, across the boundary between categories which the ogival IRF represents.

INSERT FIGURE 1.1 ABOUT HERE

When modeled by a logistic ogive, the IRF usually requires the estimation of two parameters. One is the location parameter which describes where along the trait continuum the function is centered. The center of the function is defined as midway between its lower and upper asymptotes. More generally, the center of the function is at the point of inflection of the curve. The letter "b" typically signifies the item's location parameter.

The item location parameter is modeled in the same metric as the parameter that describes a person's trait level. As a result it also represents the amount of the trait being measured that is required for a person to be as likely to respond in the particular response category being modeled as to not respond in this category.

The second parameter that must be estimated to provide a description of an IRF is the parameter that indexes the slope of the function. This parameter, signified by the letter "a", is usually estimated at the point of inflection of the function which is also where the steepness of the slope is greatest. This parameter gives an indication of how well an item discriminates among people along the trait continuum. In other words, it shows how well an item can tell people apart with respect to the amount of a trait that they have. Another way of saying this is that highly discriminating items can tell with greater accuracy than poorly discriminating items whether people who have trait levels that are close together, and whose ability is close to the item's difficulty, are likely to provide different responses.

With these two parameters to describe the function it is convenient to write the equation for the logistic form of an IRF as

$$P(\theta) = \frac{e^{a(\theta - b)}}{1 + e^{a(\theta - b)}} \quad , \qquad (1.1)$$

where $P(\theta)$ is shorthand for $P(x = i|\theta)$ and this in turn denotes the probability that the response to item *x* is with option *i* at a given trait level (θ). In this case, response option *i* refers to the positive item response option (i.e., more of the trait; a correct answer when there are only two options). With the dichotomous model, the equation for the function that describes the probability of responding in the alternative category (less trait; wrong answer) is simply 1 minus the right hand side of Equation 1.1.

Typically, only the response function for one response category is explicitly modeled since the functions for each category are the complement of one another. Usually only the positive category is modeled, with a monotonically increasing function (Figure 1.1). The complementary nature of the category functions means that knowing the characteristics of one function tells you all you need to know about the other function.

However, measurement items with multiple response options also exist, and their use is becoming more prevalent. Such items include rating scale items, such as the ubiquitous Likert-type items, as well as ability test items that provide partial credit for partially correct answers, portfolio assessment test formats, and even multiple-choice items when each response option is scored separately.

Polytomous IRT models for such items operate quite differently from dichotomous models. In these cases, knowledge of the characteristics of one of the response category functions does not determine the characteristics of the other category functions and each category function must therefore be modeled explicitly. A corollary of the non-determinate nature of the category response functions is that they are no longer exclusively monotonic functions. In the case of items with ordered categories, only the functions for the extreme negative and positive categories are monotonically decreasing and increasing respectively. Figure 1.2 shows that the function for the second category rises as the probability of responding in the most negative category decreases, but only up to a point at which time it decreases as the probability of responding in the next category increases. Ultimately the "next category" is the extreme positive category, which has a monotonically increasing function.

INSERT FIGURE 1.2 ABOUT HERE

In as much as the application of IRT is an exercise in curve fitting, the presence of non-monotonic functions presents special problems. Such functions can no longer be simply described in terms of a location and a slope parameter. Actually selecting the appropriate mathematical form and subsequently estimating parameters for such unimodal functions is a significant challenge. Fortunately, in the case of ordered polytomous items, a solution to this problem has been found by treating polytomous items essentially as concatenated dichotomous items. Multiple dichotomizations of item response data are combined in various ways to arrive at appropriate response functions for each item category. The different ways in which the initial dichotomizations can be made and differencet approaches to combining dichotomizations result in a variety of possible polytomous IRT models. In addition, different types of polytomous items require, or allow, different features to be incorporated into an applicable IRT model. The result is a range of possible polytomous models that far outstrips the number of available dichotomous models. Before addressing these issues in detail, a brief discussion will be provided to suggest why the extra complexity involved in modeling polytomous items might be justified.

Reasons for Using Polytomous IRT Models

Perhaps the simplest and most obvious reason for the development of polytomous IRT models is the fact that polytomous items exist and are commonly used in applied psychological measurement. To be a comprehensive measurement approach, IRT must provide appropriate methods for modeling these data. The need for polytomous response formats may be most acute in the measurement of personality and social variables. Kamakura and Balasubramanian (1989) suggest that dichotomous distinctions are often less clear in this context than in ability measurement settings and that “more subtle nuances of agreement/disagreement” (p. 514) are needed than dichotomous items permit. Similarly, Cox (1980) notes that items with two or three response alternatives are inadequate in this context, because they cannot transmit much information and they frustrate respondents.

The existence of polytomous data also has consequences for statistical data analysis. Wainer (1982) points out that responses to polytomous items can be thought of as data distributions with short tails. This may affect statistical procedures, such as obtaining least squares estimators, which rely on assumptions of a Gaussian distribution. Rather than simply proceeding as though the data met the assumptions, a better approach, according to Wainer, is to use procedures specifically designed for this type of data, especially relevant IRT models.

Prior to IRT, the two most common methods for dealing with polytomous data were Thurstone and Likert scaling. Thurstone scaling is similar to IRT in that it scales items on an underlying trait using a standardized scale (Thurstone 1927a; 1927b; 1929; 1931). However, to achieve Thurstone scaling one must assume either that the trait is normally distributed in the population of interest or one must select items so that this is the case (Thurstone 1925; 1927b). Likert (1932) showed that a simple summated rating procedure produced results equal to or better than Thurstone's method. The simplicity of the approach meant that Likert scaling was soon widely adopted as the method of choice for rating data such as those used in attitude measures. However, Likert scaling assumes a linear relationship between the response probability and the underlying trait (Hulin, Drasgow & Parsons, 1983). Neither the normality assumption of Thurstone scaling nor the linearity assumption of Likert scaling are particularly plausible.

There are also psychometric issues that make polytomous items attractive in comparison to dichotomous items. At a general level, one such issue is that polytomous items measure across a wider range of the trait continuum than do dichotomous items. This occurs simply by virtue of the fact that polytomous items contain more response categories than dichotomous items.

Strictly speaking, of course, all items, dichotomous and polytomous, measure across the entire range of the trait continuum, from negative to positive infinity. However, the amount of measurement information provided by an item is peaked above the trait scale location of that item and then drops, often precipitously, at higher and lower trait levels. Paradoxically, the more information that an item provides at its peak the narrower the range of the trait continuum that the item provides useful information about. The advantage of polytomous items is that, by virtue of their greater number of response categories, they are able to provide more information over a wider range of the trait continuum than dichotomous items.

Masters (1988a) and Bejar (1977) note that the entire purpose of using more than two categories per item is to try to obtain more information about the trait level of the people being measured so that more precise trait level estimates can be obtained. Masters (1988b) also points out that more detailed diagnostic information about respondents and items can be obtained from polytomous test items.

Samejima (1975; 1977a) demonstrates the increase in statistical information that is available from a polytomous IRT model in comparison to a dichotomous model. Conversely, though in a different context, Cohen (1983) demonstrates that reducing continuous or multiple category data to the dichotomous level leads to a systematic loss of measurement information.

Finally, Kamakura and Balasubramanian (1989) note two practical benefits that follow from using the more informative polytomous items in the context of marketing research. They are: i) testing time and cost are decreased; and ii) there will be a positive effect on respondent motivation to participate in the measurement if they are required to respond to fewer test items. However, these benefits are more likely to be obtained in any context using simple fixed measurement scales, such as rating scales. They are unlikely to apply in ability measurement settings, where the creation and scoring of polytomous items incurs greater time and financial costs.

Polytomous IRT Models

As was mentioned in the previous section, the key to making ordered polytomous IRT models work in practical terms is dichotomization. At this point, for the sake of simplicity we will concentrate the discussion on the case of polytomous models for items with ordered categories. Ordered polytomous items are simply those where the response categories have an explicit rank ordering with respect to the trait of interest. Likert-type attitude items and partial credit cognitive ability items are examples of ordered polytomous items. Responses to such items are also referred to as graded responses in the literature. Items on the Raven Progressive Matrices test and most multiple-choice test items are examples of items where the response options are not designed with an explicit ordering.

Polytomous items are categorical items in the same way as dichotomous items; they simply have more than two possible response categories. Categorical data can effectively be described in terms of the number of categories into which the data can be placed. Ordered categories are defined by boundaries or thresholds that separate the categories. Logically, there is always one less boundary than there are categories. Thus, for example, a dichotomous item requires only one category boundary to separate the two possible response categories. In the same manner, a five-point Likert-type item requires four boundaries to separate the five possible response categories.

Two Types of Probabilities

Compared to dichotomous models, the primary additional complication with polytomous IRT models, is that the distinction between response categories and the boundaries that separate them reflects two different types of conditional probabilities. These are: (1) the probability of responding in a given category; and (2) the probability of responding positively rather than negatively at a given boundary between two categories.

In the dichotomous case the two probabilities amount to the same thing. That is, the probability of responding positively rather than negatively at the category boundary (modeled by the IRF) also represents the probability of responding in the positive category. When there are more than two categories this is no longer the case because there is always at least one category that is defined by two boundaries. In that case, the probability of responding *in* such a category is determined by the characteristics of, at least, the two adjacent boundaries.

Measurement practitioners are typically most interested in the probability of responding in a given category, as this is the basis for determining respondents' levels of ability, trait, or attitude. However, we have already noted that the unimodal response functions for middle categories (Figure 1.2) are difficult to model mathematically. Therefore, most polytomous IRT models proceed by modeling each category boundary separately with a dichotomous model and then combining all the boundary information for an item in a formally specified manner. Applying a dichotomous model at each category boundary gives the probability of responding positively rather than negatively at the specific boundary. Hence the need for explicitly ordered categories. Combining the dichotomous information from each boundary gives the probability of responding *in* each individual category.

Consider a specific item with five categories separated by four category boundaries. Determining the probability of responding positively rather than negatively at the first boundary will, at a minimum, involve the responses in Category 1 and Category 2. However, the responses in Category 2 will also be required for determining the probability of responding positively rather than negatively at the second boundary (the boundary between Category 2 and Category 3). Thus the probability of actually responding *in* Category 2 is a combination of the probability of responding positively at the first category boundary and the probability of responding negatively at the second category boundary.

Therefore, while the IRF of a dichotomous IRT model can represent both the probability of responding on the positive side of the category boundary and also responding *in* the positive category, there is no single polytomous equivalent of the dichotomous IRF. The two kinds of probability simply cannot be represented in one type of polytomous response function. Instead, polytomous models require a specific function simply to represent the probability of responding at a category boundary. Because this particular function dichotomizes the response data at a category boundary it has the typical ogive shape of the dichotomous IRF. Since it only represents the category boundary half of the response probability description represented by dichotomous item IRFs, it will be referred to as the category boundary response function (CBRF). Examples of CBRFs are shown in Figure 1.3.

INSERT FIGURE 1.3 ABOUT HERE

The other type of function that polytomous models require to represent the probability of responding *in* a given category was shown in Figure 1.2. In keeping with Chang and Mazzeo (1994) and Weiss and Yoes (1991), this type of function will be referred to as an item category response function (ICRF) to reflect its specific item category role. Unfortunately, researchers have used many different terms to refer to CBRFs and ICRFs. A selection of these is listed in the glossary.

Two Types of Polytomous Models

There is also a distinction within both dichotomous and polytomous IRT models that is based primarily on the underlying measurement philosophy. The distinction is between Rasch-type models that attempt to conform to fundamental measurement theory and more pragmatically based models that do not. Conforming to fundamental measurement theory ultimately requires that IRT models adhere to the principle of specific objectivity. At the level of model parameters, specific objectivity requires that comparisons among item parameter values be independent of person parameter values and any other item parameters that are not being compared (and vice versa for person parameter comparisons). Rasch (1977) determined that for this to occur, the two types of parameters (item and person) in the measurement model must be additive and separable. This means that they must be in the form (θ-b) and cannot have a multiplicative form as in $a(\theta$-$b)$. Thus, while it was mentioned earlier that estimation of a dichotomous IRF usually requires the estimation of two parameters, this is not true if adherence to the principle of specific objectivity is required. In that case the discrimination (a) parameter must not be estimated. The dichotomous Rasch model therefore only requires (allows) the estimation of an item location parameter. At the parameter estimation level, specific objectivity is closely linked to the mathematical principle of sufficiency (Rasch, 1977) which provides the link between the measurement theory and the practical testing model in that it allows person and item parameters to be estimated independently (Wright, 1997).

In the polytomous arena models that do not conform to fundamental measurement theory are direct successors to the Thurstone (1928a; 1928b; 1931; 1937; Edwards & Thurstone, 1952) scaling approach. They are also historical successors to equivalent models in the field of bioassay (Aitchison & Bennet, 1970; Aitchison & Silvey, 1957). Their IRT form however, is largely due to a general framework developed by Samejima (1969; 1972; 1988).

Category boundaries

Two practical differences also present themselves when comparing Rasch and Thurstone/Samejima polytomous models. The first, perhaps obvious, difference is the mathematical form of the CBRF used to dichotomize the polytomous item data. Rasch models typically use a dichotomous Rasch model IRF to define category boundaries. Thurstone/Samejima models, on the other hand, typically use the 2-parameter logistic (2PL) model IRF to define category boundaries. As would be expected, the Rasch model choice of function is constrained by the additivity requirement of fundamental measurement theory in so far as this theory forms the basis for the concept of specific objectivity (Wright, 1997). The second, less obvious, consequence of applying fundamental measurement theory to polytomous data is that it constrains the set of item category responses to which the dichotomization process is applied and it does so in a particular manner. That is, it determines in a distinctive way which response data are used to define a Rasch CBRF.

As described earlier, a CBRF represents the probability of responding positively at a category boundary rather than responding negatively. However, in a polytomous item there are at least two ways to respond positively rather than negatively. This is presented graphically in Figure 1.4. Here it can be seen that "positively rather than negatively" can refer to just the two categories immediately adjacent to the category boundary (bottom half of Figure 1.4). Alternatively, the phrase can refer to all of the possible response categories for an item above and below the category boundary respectively (top half of Figure 1.4). There is also a third possibility. The phrase can also refer to a combination of the two alternatives just described. This leads to something of a hybrid polytomous model which has thus far had little impact and will therefore not be discussed further here. Interested readers are referred to the work of Tutz (1990; 1997).

INSERT FIGURE 1.4 ABOUT HERE

In order to maintain parameter separability and sufficient statistics for parameter estimation, that is, in order to maintain the theoretical and practical advantages of specific objectivity, Rasch-type polytomous models dichotomize polytomous data in the first manner just described (Wright & Masters, 1982; Molenaar, 1983). That is to say that the requirement for additive parameters means that Rasch model CBRFs only dichotomously model responses in categories immediately separated by a given boundary (Wright, 1997). As a result, the dichotomizations involve local comparisons and ignore the context of the responses in all other item categories. In contrast, Thurstone/Samejima models dichotomously model all possible response category responses above and below each category boundary respectively. These dichotomizations can be described as a set of global comparisons and directly incorporate the entire item category response context at each boundary. Thus, any given Rasch-model CBRF models responses in only two specific categories while each Thurstone/Samejima-model CBRF models the responses in every item category.

This is perhaps most easily seen by taking a specific response as an example. In Figure 1.4, a capital X is used to represent a response in Category 4 of the demonstration item. The vertical, dashed line shows where this response would be counted in the category boundary dichotomization process, for each category boundary, in both the Rasch and the Thurstone/Samejima approaches. The example shows that in the Thurstone/Samejima approach this response would be used in estimating the category boundaries for all four of the boundaries separating the five response categories of this item. The response registers on the positive side of the first three category boundaries and as a response on the negative side of the boundary between Categories 4 and 5.

This approach is in marked contrast to the Rasch case. Since Rasch dichotomization involves only responses in categories adjacent to a specific boundary, the example response in Category 4 is only directly relevant to modeling two category boundaries. It registers as a positive response in the third category boundary dichotomization and as a negative response in modeling the fourth category boundary.

Item Category Response Functions

These two ways to model polytomous item category boundaries lead to two different ways of calculating ICRFs. Recall that the ICRFs provide the probability of responding *in* a particular category as a function of trait level (θ). While details about the form and location of category boundaries is helpful, from a measurement, item analysis and test construction perspective it is the probability of responding in a category that is typically required. Since it is difficult to directly model this probability, it is obtained by combining information about category boundaries. How this information is combined naturally depends on how it is initially obtained.

Thurstone/Samejima. If P_{i_g} is the probability of responding in a particular category (g) to item i and $P^*_{i_g}$ represents a CBRF in the Thurstone/Samejima case (where both are conditional on θ), then

$$P_{i_g} = P^*_{i_g} - P^*_{i_{g+1}} \quad . \tag{1.2}$$

That is, the probability of responding in a particular category is equal to the probability of responding above (on the positive side of) the lower boundary for the category (i_g) minus the probability of responding above the category's upper boundary (i_{g+1})[1]. It can be seen that this gives the probability of responding between the category's two boundaries which, intuitively and in fact, equals the probability of responding in the given category. Thus, for example, a response such as that in Figure 1.4, in the fourth category (i.e., g=3) of a 5-category item would be represented algebraically as $P_{i_3} = P^*_{i_3} - P^*_{i_4}$.

This approach requires an additional definition for the probability of responding to an item at all (i.e., $P^*_{i_0}$) which is set to equal 1. Without this definition it would not be possible to calculate the probability of responding in the first category. Similarly, the probability of responding above the highest category ($P^*_{i_5}$ in our example) is set to zero. This effectively states that the probability of responding in Category 5 of a 5-category item is equal to the probability of responding positively at the fourth CBRF. More generally, the probability of responding in the highest category of an item with ordered categories is equal to the probability of responding positively at the highest CBRF. This type of polytomous model is referred to as a difference model by Thissen and Steinberg (1986).

All Thurstone/Samejima models take the form described in Equation 1.2. The various models within the family differ only in terms of how restrictions are applied to the parameters of the CBRFs, both within and across items. Specific restrictions and the different models that result from their implementation will be described in Chapter 4.

Rasch. The situation is somewhat different for Rasch models, largely because the dichotomization process involves the limited context of a local comparison. One result of this situation is that polytomous Rasch models rarely involve explicitly modeling category boundaries. Since these boundaries provide localized information, they are combined algebraically into a general expression for each ICRF. The expression which describes the probability of responding in a given item category takes the form

$$P_{i_g} = \frac{e^{z_{i_g}}}{\sum e^{z_{i_h}}} \quad . \qquad (1.3)$$

Here, z_{i_g} represents a sum of the differences between a given trait level and the location of each category boundary $\left(\text{i.e., } \sum\left(\theta - b_{i_g}\right)\right)$ up to the particular category (g) in question. Seeing how this works in practice requires explicitly expanding out the equation for a specific example. If we again take the example from Figure 1.4, of a response in Category 4 (g=3) of a 5-category item, then Equation 1.3 expands to show that the probability of this occurring is determined as

$$P_{i_3} = \frac{e^{(0)+(\theta-b_{i_1})+(\theta-b_{i_2})+(\theta-b_{i_3})}}{e^{(0)} + e^{(0)+(\theta-b_{i_1})} + e^{(0)+(\theta-b_{i_1})+(\theta-b_{i_2})} + e^{(0)+(\theta-b_{i_1})+(\theta-b_{i_2})+(\theta-b_{i_3})} + e^{(0)+(\theta-b_{i_1})+(\theta-b_{i_2})+(\theta-b_{i_3})+(\theta-b_{i_4})}} . \qquad (1.4)$$

In a sense, the numerator describes the likelihood of someone at a given trait level responding in the positive category of each dichotomization up to the category in question. The denominator is the sum of the numerator values for every category in the item. It also ensures that the probability of responding in any single, given category does not exceed 1, and that the accumulated probabilities of responding in a category, across all the categories for an item, sum to 1. The zero elements in the exponentials are (in this case) redundant in all of the exponential terms except the first term in the denominator. They are included here for completeness of exposition. Their function is to represent the possibility of a negative response at the first dichotomization, that is, failing to pass beyond the first CBRF location. In the denominator the first exponent describes the requirement that a response must be made in one of the available categories even if it is only a negative response at the first category boundary. This probability is of course 1, showing that this element of the equation corresponds logically, to the constraint in the Thurstone/Samejima case that $P_{i_0}^{*}$ equal 1. Given the form of the general expression for Rasch polytomous models shown in Equation 1.3, this type of model is described as a divide-by-total model by Thissen and Steinberg (1986). The major types of polytomous Rasch models, their relationships to each other and variants of the major models are described in Chapter 3.

However, we will begin our description of specific models in Chapter 2 by looking at the most general of the models outlined in this book. It is a model that does not require items with ordered categories.