PSYCHOMETRIC LATENT RESPONSE MODELS

Eric Maris

UNIVERSITY OF NIJMEGEN

In this paper, some psychometric models will be presented that belong to the larger class of *latent response models* (LRMs). First, LRMs are introduced by means of an application in the field of *componential item response theory* (Embretson, 1980, 1984). Second, a general definition of LRMs (not specific for the psychometric subclass) is given. Third, some more psychometric LRMs, and examples of how they can be applied, are presented. Fourth, a method for obtaining maximum likelihood (ML) and some maximum a posteriori (MAP) estimates of the parameters of LRMs is presented. This method is then applied to the *conjunctive Rasch model*. Fifth and last, an application of the conjunctive Rasch model is presented. This model was applied to responses to typical verbal ability items (open synonym items).

Key words: latent response models, componential IRT, conjunctive Rasch model, incomplete data, EM algorithm, verbal ability, open synonym items.

Cognitive Componential Item Response Theory

One of the best known item response theory (IRT) models is the one-parameter logistic, or *Rasch model* (Fischer, 1974; Rasch, 1980). This is a model for a dichotomous random variable Y_{pi}, usually denoting correct/incorrect, characterized by the following probability of a correct response:

$$P(Y_{pi} = 1; \beta_p, \delta_i) = \frac{\exp (\beta_p - \delta_i)}{1 + \exp (\beta_p - \delta_i)}. \tag{1}$$

In (1), the indices p ($p = 1, \ldots, P$) and i ($i = 1, \ldots, I$) denote, respectively, the persons and the items. The parameters β_p and δ_i are usually denoted as ability and item difficulty.

As an example, we consider a person's response to a typical verbal intelligence test item: a question asking for one or more synonyms of a given word (e.g., woods, ship, job), which will be called an *open synonym item* (the somewhat unusual formulation in which more than one synonym may be given as a response, will be justified later). The response to such an item is scored correct if the correct synonym is among the synonyms the subject has given as a response.

We now consider two different conceptualizations of the solution process. First, from the point of view of the Rasch model, the solution process is considered as being governed by a single person characteristic and a single item characteristic, of which the first has a monotone *increasing* relation with the probability of a correct response, and the second a monotone *decreasing* one. This is a purely formal conceptualization of the solution process.

GENERATE		EVALUATE
trees		No
lakes		No
domain		No
forrest		Yes
nature		No

FIGURE 1.
Covert responses in two subprocesses (generating and evaluating) of the solution process of an open synonym item.

Second, in cognitive psychology, typically a verbal description is given of the solution process (i.e., the cognitive theory). For example, the process involved in solving an open synonym item can be considered as consisting of two subprocesses: *generating* and *evaluating* (De Boeck, 1988). In the generating process, words that are candidates for being synonyms are generated, and in the evaluating process, each is evaluated as to whether they are a synonym. This is illustrated in Figure 1 for the stimulus word *woods*.

The strength of the cognitive psychologist's approach in comparison with the psychometrician's, is that complex behavior is considered as the result of more elementary behavior. On the other hand, the strength of the psychometrician's approach is that when theory is expressed as a mathematical model, it can be tested in an unequivocal way.

It is possible to integrate the two approaches such that the strengths of both are combined. For the present example, this is possible by assuming that a (different) Rasch model holds for the generating and the evaluating process, and that a correct response will only be given if both processes resulted in a correct response. For the generating process, a correct response is a covert list containing a correct synonym, and for the evaluating process this is the (covert) evaluation of this particular candidate as a correct one. This model is characterized by the following probability of a correct response:

$$P(Y_{pi} = 1; \beta_{p1}, \beta_{p2}, \delta_{i1}, \delta_{i2})$$
$$= P(X_{pi1} = 1; \beta_{p1}, \delta_{i1}) \times P(X_{pi2} = 1; \beta_{p2}, \delta_{i2}). \tag{2}$$

In (2), X_{pi1} and X_{pi2} are nonobservable random variables (called *latent responses* in the following) that indicate, respectively, whether or not the generating and evaluating process resulted in a correct response. The right-hand side of (2) results from the assumption that both latent responses have to be correct for the observable response to be correct also, and the assumption of local stochastic independence (LSI) between X_{pi1} and X_{pi2}. In the following, this model will be denoted as the *conjunctive Rasch model*. The conjunctive Rasch model can be considered as a relaxation of Embretson's (1980, 1984) *multicomponent latent trait model* (MLTM), in the sense that it does not require the X_{pi1}'s and the X_{pi2}'s to be observed in order for it to be applicable to some set of Y_{pi}'s.

To test the cognitive theory, it is useful if the latent responses can be made observable. This is possible by making use of *subtasks*. The essential characteristic of subtasks is that the (observable) responses they yield to are assumed to conform to the model for the *latent* responses. For the present example, one can construct a *generating* and an *evaluating* subtask for every open synonym item. For example, a generating subtask for the open synonym item involving *woods* is the following: "Give a list of words that come to your mind as you are searching for a synonym for *woods*." And the evaluating subtask is a kind of multiple choice item in which a list of synonym candidates is presented and from which the correct synonyms have to be chosen. The link between the responses to the open synonym items, and the responses to the generating and evaluating subtasks is established by the assumption that the models for these latter *manifest* responses are also assumed to hold for the *latent* responses that underly the manifest responses to the open synonym items (i.e., $P(X_{pi1} = 1; \beta_{p1}, \delta_{i1})$ and $P(X_{pi2} = 1; \beta_{p2}, \delta_{i2})$ in the right-hand side of (2)).

Our approach to testing the cognitive hypothesis, which is also suggested by De Boeck (1988), assumes that we have at our disposal parameter estimates of the model for the open synonym items as well as for the models for the generating and the evaluating items. In essence, testing is a comparison of, on the one hand, the item and person parameter estimates obtained using the responses to the open synonym items, and on the other hand, the parameter estimates obtained using the responses to the generating and evaluating subtasks.

The use of subtasks to test a cognitive theory about responses to complex tasks has been introduced in psychometrics by Embretson (1980, 1984). Her approach is different from ours in that she does not estimate the parameters of the model for the complex items. Instead, she uses a two-step procedure in which she first estimates the Rasch model parameters of the subtasks, and then uses these estimates to *predict* the responses on the complex task. A comparison between these predictions and the responses that have actually been given, constitutes the core of her approach to testing the cognitive hypothesis. Performing comparisons on the level of the parameters, as in our approach, has some advantages, as will be shown in the final section.

The idea of modelling complex behavior such that this model's parameters are related in some well-specified way to the parameters of models for more elementary behavior, is not new. In particular, using response time as the variable to be modeled instead of correctness, this is also the basic idea in R. J. Sternberg's (1977) *componential analysis* (see also Goldman & Pellegrino, 1984; and Pellegrino & Kail, 1982).

Latent Response Models

We will use the conjunctive Rasch model to introduce the definition of *latent response models* (LRMs). The derivation of the conjunctive Rasch model is based on a distinction between two types of random variables: the observed (Y_{pi}'s) and the latent (X_{pi1}'s and X_{pi2}'s) random variables. The observed random variables are related to the latent ones through the following rule: Y_{pi} equals 1 (correct response) *if and only if* both X_{pi1} and X_{pi2} equal 1, and 0 otherwise. A rule that specifies the relationship between the latent and observed random variables will be called a *condensation rule*. The rule specified above, will be called the *conjunctive* condensation rule. Now, the conjunctive Rasch model simply follows from the assumption of a Rasch model for the X_{pi1}'s and X_{pi2}'s in combination with the conjunctive condensation rule.

A LRM is a model for a set of N variables Y_n ($n = 1, \ldots, N$). In psychometric applications (involving a complete design), N always equals $P \times I$, the number of persons times the number of items, and every n is associated with a particular person-

item pair (p, i). The N Y_n's are considered as a whole because the models for each of them separately are related by means of common parameters. In the LRM framework, there is no restriction on the type of Y_n's. In particular, they can be discrete and continuous, and they can be scalar or vector-valued.

The definition of probabilistic LRMs involves two key aspects: 1., the definition of latent random variables, and 2., the definition of a condensation rule. First, for every Y_n, a set of K latent random variables X_{nk} ($k = 1, \ldots, K$) is defined. In vector notation, $X_n = (X_{n1}, \ldots, X_{nK})^t$. In principle, not every X_{nk} has to be defined for a particular Y_n, but in order to keep the notation simple, it is assumed that for every Y_n a complete vector X_n is defined.

Second, a condensation rule is defined as a function that specifies the relationship between Y_n and X_n. Using C as a generic symbol for a condensation rule, this relationship can be expressed as follows:

$$Y_n = C(X_n).$$

Because a function is a many-to-one mapping, Y_n can be considered as a condensation of X_n.

We will now present three examples of condensation rules. Each of them is involved in a psychometric model that will be presented in the following. First, we consider the *conjunctive* condensation rule, which is defined for dichotomous (0/1) X_{nk}'s only. It can be written as follows:

$$C(X_n) = \prod_{k=1}^{K} X_{nk}.$$

Second, we consider the *disjunctive* condensation rule, which is also defined for dichotomous X_{nk}'s only. It can be written as follows:

$$C(X_n) = 1 - \prod_{k=1}^{K} (1 - X_{nk}).$$

And third, we consider the so-called *drop-off* condensation rule, which is also defined for dichotomous X_{nk}'s. This condensation rule results in a *polytomous* Y_n involving $K + 1$ categories. It can be written as follows:

$$C(X_n) = 0 \text{ if } X_{n1} = 0$$

$$= 1 \text{ if } X_{n1} = 1 \text{ and } X_{n2} = 0$$

$$= \cdot$$

$$= K \text{ if } X_{n1} = X_{n2} = \cdots = X_{nK} = 1.$$

This condensation rule is involved in a type of psychological process involving K components that are executed sequentially, but in such a way that passing the $(k + 1)$-th component is only possible if the k-th component was also passed. The application of a model involving this condensation rule assumes that it is possible to categorize the different responses according to the component at which the (first) error occurs. This condensation rule is denoted as the *drop-off* condensation rule, because at every level of Y_n a set of X_n-patterns drops off.

Many more interesting condensation rules exist. An overview of condensation

rules as they are involved in existing models is given by Maris (1992). Some examples of existing models that can be formulated as LRMs are the following: Thurstone's (1927) model of comparative judgment (with the condensation rule being *dichotomisation*), Verhelst and Verstralen's (1993) unfolding model (with the condensation rule being *collapsing*), sequential processing stages models (Luce, 1986; with the condensation rule being *addition*), and ML factor analysis as formulated by Rubin and Thayer (1982) (with the condensation rule being *omission*), which is a LRM for a vector-valued Y_n.

At this point, we did not yet specify a probability distribution function (PDF) for the Y_n's. However, given a particular condensation rule, this PDF is completely specified by a particular PDF for each of the X_n's. This can be expressed as follows:

$$f(Y_n; \mathbf{\theta}) = \int_{\Re_C} g(\mathbf{X}_n; \mathbf{\theta}) \, d\mu(\mathbf{X}_n). \qquad (4)$$

In (4), f and g denote the PDFs of Y_n and $\mathbf{X}_n$, respectively. These functions are usually different for different values of n, but for simplicity we will not index them. Instead, we let their arguments (Y_n and $\mathbf{X}_n$) indicate their particular form. The parameters of these PDFs are denoted by θ_t ($t = 1, \ldots, T$), and in vector notation, $\mathbf{\theta} = (\theta_1, \ldots, \theta_T)^t$. In this notation, no distinction is made between particular types of parameters like the item and person parameters of psychometric models. The essential point in (4) is the integral over the range $\Re_C$. This integral is used as a generic symbol for both summation over discrete and integration over continuous variables. This range $\Re_C$ is completely specified by the condensation rule C. The integration is with respect to a dominating measure μ (see Billingsley, 1986, p.442). In the discrete case, μ simply has the value 1 for all $\mathbf{X}_n$ in the sample space, and the integral can be replaced by a summation sign. The continuous case is more complicated, and will not be treated here.

In order for f to be a PDF, we make the technical assumption that C is a measurable function. And in order for this PDF to be nondegenerate, we assume that C is such that the complete probability mass is not concentrated in a single point, and for a continuous Y_n, that f is continuous on all interior points of its domain. This assumption holds for all condensation rules that are considered in this paper.

LRMs have two kinds of advantages: conceptual and technical. The conceptual advantages are twofold. First, the concept of LRMs is well suited for the development of new models based on particular psychological assumptions on the process that generates the observations. These psychological assumptions only have to be translated in a condensation rule that operates on a set of latent responses. A second conceptual advantage is that it may be clarifying to see that existing models that appear to be different at first sight, nevertheless belong to the same general class.

The technical advantage is that a relatively simple algorithm exists that can be used for the computation of the *maximum likelihood* (ML) and some *maximum a posteriori* (MAP) estimates of the parameters of LRMs.

Some LRMs with Applications to Psychometrics

From hereon, we will only consider psychometric LRM models. Thus, the index n has to be replaced by the double index pi. In particular, Y_n is replaced by Y_{pi}, X_{nk} by X_{pik} and $\mathbf{X}_n$ by $\mathbf{X}_{pi}$. The three LRMs to be presented are all based on Rasch models for the X_{pik}'s. These K models (for $k = 1, \ldots, K$) are characterized by the following probability:

$$P(X_{pik} = 1; \beta_{pk}, \delta_{ik}) = \frac{\exp{(\beta_{pk} - \delta_{ik})}}{1 + \exp{(\beta_{pk} - \delta_{ik})}}. \tag{5}$$

In the following, we will use π_{pik} as a shorthand notation for this probability.

The Conjunctive Rasch Model

The conjunctive Rasch model is characterized by the following probability:

$$P(Y_{pi} = 1; \beta_p, \delta_i) = \prod_{k=1}^{K} \pi_{pik}. \tag{6}$$

In (6), $\beta_p = (\beta_{p1}, \ldots, \beta_{pK})'$ and $\delta_i = (\delta_{i1}, \ldots, \delta_{iK})'$. In the following, we will use π_{pi} as a shorthand notation for this probability. Equation (6) follows from the conjunctive condensation rule and the assumption of LSI between the X_{pik}'s. Because the π_{pik}'s are within the range (0, 1), it follows from (6) that every π_{pik} (for $k = 1, \ldots, K$) imposes an *upper limit* on π_{pi}. This fact allows us to introduce the notion of *necessary ability levels*. In particular, if $\delta_{ik} = a$ for some k, then $\beta_{pk} = a$ is a necessary ability level for π_{pi} to be equal to 0.5.

The conjunctive Rasch model is a model for psychological processes that involve a number of events, occurring serially or in parallel, whose outcomes *all* have to be of a particular kind (e.g., correct) for the observable response (Y_{pi}) to be of a particular kind also. In other words, it is a model for psychological processes that involve *multiple requirements*.

Embretson (1980) was the first to present the conjunctive Rasch model. Later, Embretson (1984) also presented the following modification of the conjunctive Rasch model:

$$P(Y_{pi} = 1; \beta_p, \delta_i, a, g) = (a \times \pi_{pi}) + [g \times (1 - \pi_{pi})], \tag{7}$$

in which a and g are parameters. This model is a LRM with $(K + 1)$ X_{pik}'s, involving the conjunctive condensation rule. The first K X_{pik}'s and their joint PDF are the same as those of the conjunctive Rasch model. The PDF of $X_{pi(K+1)}$ is locally stochastically *dependent* on the first K X_{pik}'s. In particular, it can be defined as follows:

$$P(X_{pi(K+1)} = 1 | X_{pi1}, \ldots, X_{piK}; a, g) = a \text{ if } \forall k: X_{pik} = 1$$

$$= g \text{ if } \exists k: X_{pik} = 0.$$

The psychological interpretation of $X_{pi(K+1)}$ given by Embretson, is that it represents the outcome of the integration of the other X_{pik}'s (considering them as partial results) if $X_{pik} = 1$ for all $k = 1, \ldots, K$, and of an alternative method for solving the item if $X_{pik} = 0$ for some $k = 1, \ldots, K$. No inter-individual nor inter-item differences are assumed with respect to this component.

It is very easy to generalize the conjunctive Rasch model to cases in which, depending on the item type, different X_{pik}'s are assumed to be involved (e.g., items with many, and items with few X_{pik}'s). One such mixed model, resulting from this generalization, is the one in which responses to complex items and their subtasks are considered jointly.

The Disjunctive Rasch Model

The *disjunctive Rasch model* is characterized by the following probability:

$$P(Y_{pi} = 1; \beta_p, \delta_i) = 1 - \prod_{k=1}^{K} (1 - \pi_{pik}). \tag{8}$$

In the following, we will also use π_{pi} as a shorthand notation for this probability. Equation (8) follows from the disjunctive condensation rule and the assumption of LSI between the X_{pik}'s. Given the range of the π_{pik}'s, it follows from (8) that every π_{pik} (for $k = 1, \ldots, K$) imposes a *lower limit* on π_{pi}. This fact allows us to introduce the notion of *sufficient ability levels*. In particular, if $\delta_{ik} = a$ for some k, then $\beta_{pk} = a$ is a sufficient ability level for π_{pi} to be equal 0.5.

The disjunctive Rasch model is a model for psychological processes involving a number of events of which *at least one* outcome has to be of a particular kind (e.g., correct) for the observable response to be of a particular kind also. More specifically, it is a model for a psychological process that involves *multiple strategies* (multiple solution strategies, multiple sources of information, . . .). An example of such a process is character recognition in tachistoscopic presentations. This task involves the presentation of a screen with a fixation point, followed by a timed presentation of a word, of which the character on the fixation point has to be given as a response. Two concurrent strategies can be used to solve this task, a position that is also taken by McClelland and Rumelhart (1981) in their account of some basic findings on letter perception. One strategy involves the use of information on the physical properties of that particular character, and the other the use of context information (i.e., the information that becomes available by recognizing the neighbouring characters or the whole word). This hypothesis can be tested by examining the task difficulty parameters of the two strategies, the δ_{i1}'s and δ_{i2}'s. According to the hypothesis, the task difficulty parameters of the character information strategy should depend on the character itself, whereas the task difficulty parameters of the context information strategy should depend on the redundancy of the character on that position in that particular word.

The Drop-off Rasch Model

The *drop-off Rasch model* is characterized by the following set of probabilities:

$$P(Y_{pi} = 0; \beta_p, \delta_i) = 1 - \pi_{pi1}$$

$$P(Y_{pi} = 1; \beta_p, \delta_i) = \pi_{pi1}(1 - \pi_{pi2})$$

$$= \quad \cdot$$

$$P(Y_{pi} = K; \beta_p, \delta_i) = \prod_{k=1}^{K} \pi_{pik}. \tag{9}$$

Equation (9) follows from the drop-off condensation rule and the assumption of LSI between the X_{pik}'s.

This model is well suited as a model for a psychological process that involves multiple requirements, as is also the conjunctive Rasch model. The difference with the latter model is that the nature of the processing in a particular component (e.g., the one resulting in X_{pil}) may depend on the outcomes of the previous components (the X_{pik}'s with $k = 1, \ldots, (l - 1)$). The most obvious conceptualization of such a process is one

in which the information in the item is *mentally transformed* in each of the K components of the process (see Carpenter & Just, 1986; and Pellegrino & Kail, 1984, for examples of this type of process). Thus, every component takes as its input either the item as presented physically (for the first component), or a *mental representation* of it (for the remaining $K - 1$ components). And the output of every component is either a mental representation, or an observable response (last component).

Two versions exist of such a process of mental transformations. According to the first version, the components *always* result in a mental representation, which however may be correct or incorrect. It is important to see that this correctness can be considered in a *conditional* or an *absolute* way. Conditional correctness corresponds to *single* X_{pik}'s. It involves whether or not the mental transformation is performed correctly. Absolute correctness corresponds to the *whole set* of X_{pik}'s up to and including the one for the component being considered. It involves whether or not the current mental representation still is correct.

In the second version of the process, except for the last component, a component may or may not result in a mental representation. For obvious reasons, it is assumed that the last component always results in an observable response. And, in order to make a clear distinction with the first version of the process, it is also assumed that if a component results in a mental representation, this mental representation is always correct. According to this second version of the process, one cannot assume that the same model holds for the X_{pil}'s with $l > k$ when $X_{pik} = 0$ as when $X_{pik} = 1$. In fact, depending on one's hypothesis about what happens when no mental representation is available, for the case of $X_{pik} = 0$, one may specify a model for the X_{pil}'s with $l > k$ that is completely different from the one for the case of $X_{pik} = 1$. One may even assume that several of these X_{pil}'s do not exist.

An obvious hypothesis for this version of the process is that, if some component (say, the k-th) does *not* result in a mental representation, *none* of the following components in the sequence can result in a mental representation either. This hypothesis can be expressed as a (locally dependent) model in the following way:

$$P(X_{pi(k+1)} = 1 | X_{pik} = 1) = \pi_{pi(k+1)},$$

and

$$P(X_{pi(k+1)} = 1 | X_{pik} = 0) = 0.$$

However, considering the model for the Y_{pi}'s, if $X_{pik} = 0$ it makes no difference which X_{pil}'s with $l > k$ are assumed to exist and which model is formulated for them. This follows from the particular form of the drop-off condensation rule. Therefore, the different hypotheses about what happens when $X_{pik} = 0$ are empirically indistinguishable.

In the process as conceived up to now, each mental transformation corresponds to only one X_{pik}. However, this is not necessary. In particular, one can consider a process in which X_{pik} denotes whether or not a mental representation could be constructed, and $X_{pi(k+1)}$ whether or not this representation has a certain characteristic (e.g., whether it is correct or not). In this alternative view, it is clear that $X_{pi(k+1)}$ can only be defined if there is a mental representation available ($X_{pik} = 1$). As an example, we will now consider a model for correct, incorrect, and "don't know" responses to verbal arithmetic problems (see also Mayer, Larkin & Kadane, 1984). In Figure 2, a rough theory for these responses is presented.

It is assumed that persons first have to translate the problem into a covert symbolic format to which the rules of arithmetic can be applied. Second, in order to solve the
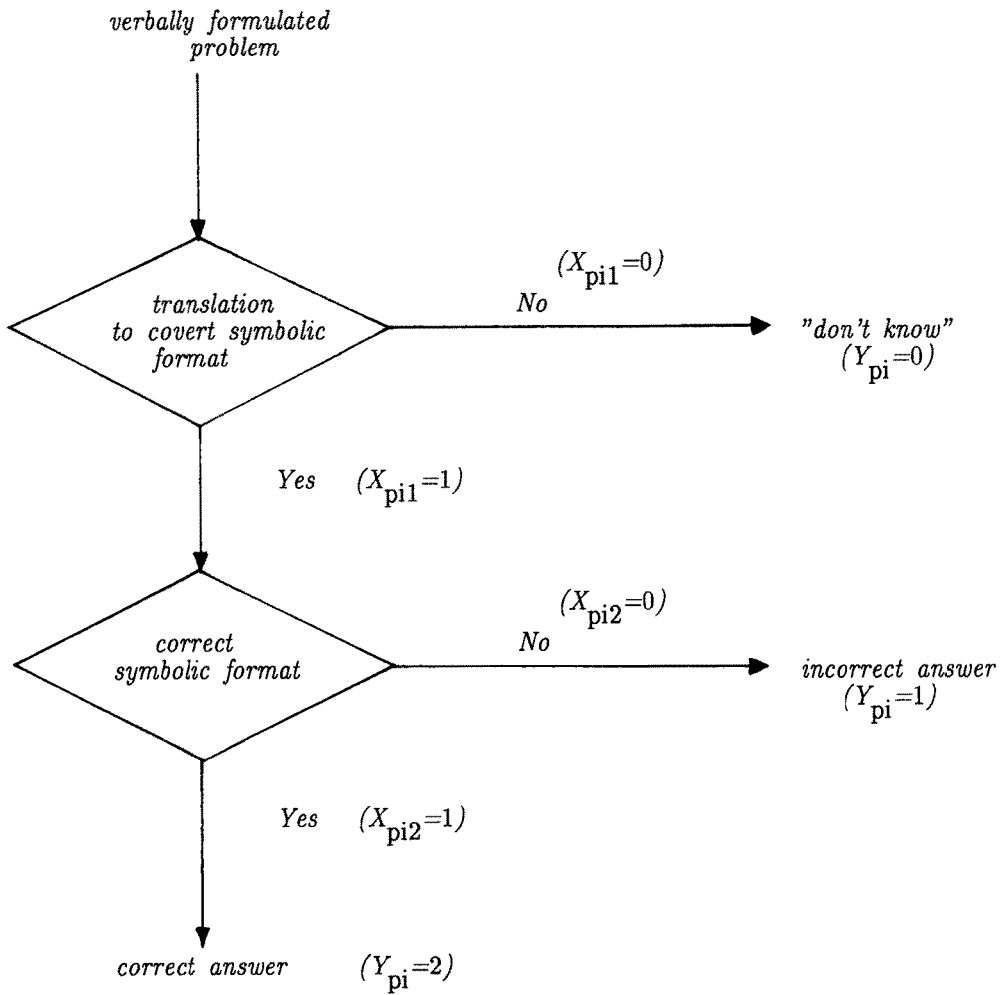
FIGURE 2.
Flow-chart of a theory of responses to verbal arithmetic items.

problem, these rules actually have to be applied to the numbers given in the problem. For simplicity, we assume that no errors are made in the application of the rules of arithmetic. Considering the translation part of the solution process, and assuming that guessing does not occur, it follows that an answer (i.e., a response different from "don't know") will only be given if the subject can perform the translation to a covert symbolic format. If the subject cannot, then a "don't know" response will be given. Further, if an answer is given, it will be a correct one if the subject has made a correct symbolic translation, and an incorrect one if he/she has made an incorrect translation.

Estimation

We will now consider an algorithm for the computation of the ML and some MAP estimates of the parameters of LRMs. We start by considering ML estimates. The PDF of $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, under the assumption of LSI between the Y_n's, is the following:

$$f(\mathbf{Y}; \boldsymbol{\theta}) = \prod_{n=1}^{N} f(Y_n; \boldsymbol{\theta}).$$

In developing an algorithm to maximize $f(\mathbf{Y}; \boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$, we start from the PDF of $\mathbf{X} = (\mathbf{X}_1^t, \ldots, \mathbf{X}_N^t)^t$, which, under the assumption of LSI between the $\mathbf{X}_n$'s, is the following:

$$g(\mathbf{X}; \boldsymbol{\theta}) = \prod_{n=1}^{N} g(\mathbf{X}_n; \boldsymbol{\theta}).$$

It is possible to write $f(\mathbf{Y}; \boldsymbol{\theta})$ as follows:

$$f(\mathbf{Y}; \boldsymbol{\theta}) = \int_{\mathfrak{R}_C} g(\mathbf{X}; \boldsymbol{\theta}) \, d\mu(\mathbf{X}). \tag{11}$$

Here, and in the following, $\mathfrak{R}_C$ will be used as a generic symbol for the range of integration that is specified by the condensation rule (as applied to a single $\mathbf{X}_n$, or as in the present case, to all N $\mathbf{X}_n$'s). Now, the conditional PDF $h(\mathbf{X}|\mathbf{Y}; \boldsymbol{\theta})$ can be written as follows:

$$h(\mathbf{X}|\mathbf{Y}; \boldsymbol{\theta}) = \frac{g(\mathbf{X}; \boldsymbol{\theta})}{f(\mathbf{Y}; \boldsymbol{\theta})}. \tag{12}$$

The essential point in (12) is the fact that the joint PDF of $\mathbf{X}$ and $\mathbf{Y}$ (which appears in the numerator of the formula of the conditional PDF) is the same as the marginal PDF of $\mathbf{X}$. This fact allows us to use the *expectation-maximization algorithm* (EM) (Dempster, Laird, & Rubin, 1977) to maximize $\ln f(\mathbf{Y}; \boldsymbol{\theta})$ (and therefore also $f(\mathbf{Y}; \boldsymbol{\theta})$). In the terminology of the EM-algorithm, $g(\mathbf{X}; \boldsymbol{\theta})$ is denoted as the *complete data likelihood*, and $f(\mathbf{Y}; \boldsymbol{\theta})$ as the *observed data likelihood*.

It is important to notice that the EM-algorithm is *not* being used here to maximize the loglikelihood of some *marginal* model, an application that was introduced to IRT by Bock and Aitkin (1981). In their application, the EM-algorithm was used to deal with the missingness of Bayesian person parameters, whereas in this case it is being used to deal with the missingness of latent responses.

We now consider the conjunctive Rasch model. For this model, the complete data likelihood is the following:

$$g(\mathbf{X}; \boldsymbol{\theta}) = \prod_{p=1}^{P} \prod_{i=1}^{I} \prod_{k=1}^{K} \pi_{pik}^{X_{pik}} (1 - \pi_{pik})^{1-X_{pik}},$$

in which the π_{pik}'s are defined as in (5). Thus, in this case, $\boldsymbol{\theta}$ consists of a matrix of person parameters $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^t, \ldots, \boldsymbol{\beta}_P^t)^t$, and a matrix of item parameters $\boldsymbol{\delta} = (\boldsymbol{\delta}_1^t, \ldots, \boldsymbol{\delta}_I^t)^t$. The k-th column of $\boldsymbol{\beta}$ will be denoted by $\boldsymbol{\beta}^k$, and the k-th column of $\boldsymbol{\delta}$ by $\boldsymbol{\delta}^k$. Further, it is easy to show that the complete data *log*likelihood can be written as follows:

$$\ln g(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\delta}) = \sum_{k=1}^{K} \left[q(\boldsymbol{\beta}^k, \boldsymbol{\delta}^k) + \sum_{p=1}^{P} T_{pk} \beta_{pk} + \sum_{i=1}^{I} U_{ik} \delta_{ik} \right], \tag{15}$$

in which

$$q(\beta^k, \delta^k) = -\sum_{p=1}^{P} \sum_{i=1}^{I} \ln \left[1 + \exp (\beta_{pk} - \delta_{ik}) \right].$$

Equation (15) simply is the sum of K loglikelihoods, each of which corresponds to one ordinary Rasch model. The T_{pk}'s and U_{ik}'s are the sufficient statistics, and they are defined as follows:

$$T_{pk} = \sum_{i=1}^{I} X_{pik},$$

and

$$U_{ik} = -\sum_{p=1}^{P} X_{pik}.$$

Because of its exponential family membership, the maximization of $\ln g(X; \beta, \delta)$ is known to have a unique solution (under certain regularity conditions) that can be found using an appropriate optimization method, as for example, a modified Newton-Raphson algorithm (see, Gill, Murray, & Wright 1981). However, this fact only holds if an identifiability restriction is imposed on the parameters (e.g., $\delta_{1k} = 0$, for all k). In the following, it will be tacitly assumed that such identifiability restrictions can be easily found and implemented.

The same optimization method that maximizes $\ln g(X; \beta, \delta)$ can also be used as a step in the maximization of $\ln f(Y; \beta, \delta)$, the loglikelihood of the conjunctive Rasch model, namely as an M-step in an EM-algorithm. In particular, in the M-step of the $(p + 1)$-th EM-cycle we maximize $E[\ln g(X; \beta, \delta)|Y; \beta^{(p)}, \delta^{(p)}]$, the conditional expected value of the complete data loglikelihood given the observed data and the parameter values of the p-th EM-cycle (denoted by $\beta^{(p)}$ and $\delta^{(p)}$). Since $\ln g(X; \beta, \delta)$ is linear in the sufficient statistics, this conditional expected value differs from $\ln g(X; \beta, \delta)$ only in the fact that these sufficient statistics are replaced by their conditional expected values. In particular, in the right-hand side of (15) with the T_{pk}'s and U_{ik}'s replaced by, respectively:

$$E(T_{pk}|Y, \beta^{(p)}, \delta^{(p)}) = \sum_{i=1}^{I} E(X_{pik}|Y_{pi}, \beta_p^{(p)}, \delta_i^{(p)}), \qquad (16)$$

and

$$E(U_{ik}|Y, \beta^{(p)}, \delta^{(p)}) = -\sum_{p=1}^{P} E(X_{pik}|Y_{pi}, \beta_p^{(p)}, \delta_i^{(p)}). \qquad (17)$$

The right-hand sides of (16) and (17) follow from their corresponding left-hand sides because of the assumption of LSI between the Y_{pi}'s. The computation of these conditional expected values is the E-step of the EM-algorithm.

The right-hand sides of (16) through (17) are sums of conditional probabilities that are defined by the PDF of the X_{pi}'s and the condensation rule. Using the formula of a conditional probability, it is easy to show the following:

$$E(X_{pik}|Y_{pi}, \beta_p, \delta_i) = 1 \text{ if } Y_{pi} = 1$$

$$= \frac{\pi_{pik} - \pi_{pi}}{1 - \pi_{pi}} \text{ if } Y_{pi} = 0, \tag{18}$$

with π_{pik} and π_{pi} being defined in (5) and (6), respectively.

We still have to deal with the question whether the EM-algorithm does what it is being used for, namely maximizing $\ln f(Y; \theta)$. In this respect, the EM-algorithm neither is better nor worse than the existing algorithms (e.g., steepest ascent, Newton-Raphson, Davidon-Fletcher-Powell). In particular, under certain regularity conditions (see Wu, 1983), which are fulfilled for all models that are considered in this paper, it can be proved that the EM-algorithm converges to a stationary point (i.e., a solution of the likelihood equations) of $\ln f(Y; \theta)$. Whether or not this stationary point is also a local or global maximum, depends on the particular form of $\ln f(Y; \theta)$ and the starting values (i.e., $\theta^{(0)}$).

We will now consider the computation of MAP estimates. The reason for considering MAP estimates is that, depending on the particular set of observations, finite ML estimates may not exist (see, e.g., Albert & Anderson, 1984, for the general case of logistic regression; and Fischer, 1981, for the special case of the Rasch model). In the Bayesian framework, this problem does not exist. The problem of the arbitrariness of the prior PDF is not considered to be serious because, except for a constant, the likelihood function and the posterior PDF are asymptotically equivalent.

Although ML and MAP estimates are defined in a different statistical framework, their actual computation may be very similar. In particular, the choice of a particular prior PDF may be formally equivalent to adding a *prior sample* within the ML framework (see, Novick & Jackson, 1974; and Jannarone, Yu, & Laughlin, 1990).

As our prior PDF, we consider the *logistic* one, which is defined as follows:

$$f(W; \alpha, \gamma) = \frac{\exp [\gamma(W - \alpha)]}{\{1 + \exp [\gamma(W - \alpha)]\}^2}, \tag{19}$$

in which α and γ $(\gamma > 0)$ are the two parameters of this PDF. The expected value of W is equal to α, and its variance is proportional to γ. We assume that α equals 0, and that γ equals 1. Further, it is easy to show that the cumulative PDF of W, which will be denoted by $F(W; 0, 1)$, is the following:

$$F(W; 0, 1) = \frac{\exp (W)}{1 + \exp (W)}. \tag{20}$$

Then, it follows that

$$f(W; 0, 1) = F(W; 0, 1) \times [1 - F(W; 0, 1)].$$

For any model in which the probability of a 1-response is expressed as the function $\exp(\ )/[1 + \exp(\ )]$ of some other function (e.g., linear) of parameters (e.g., logistic regression, Rasch model, item factor analysis), taking $f(W; 0, 1)$ as the *prior* PDF for a parameter W is equivalent to a prior sample of two observations (a 1-response and a 0-response) on a particular hypothetical data point. The probability of a 1-response on this hypothetical data point is given by (20).

Making use of the prior sample interpretation of the prior PDF, it is clear that computing MAP estimates is the same as computing ML estimates using an extended sample. This extended sample involves both Y and the prior sample, which will be

denoted by $\mathbf{Z}$. The array $\mathbf{Z}$ is of order $(T \times 2)$, and contains one pair of observations, Z_{t1} and Z_{t2}, for every value of t ($t = 1, \ldots, T$), with T denoting the total number of parameters. For the conjunctive Rasch model (and ignoring the identifiability restrictions) T equals $K \times (P + I)$.

These MAP estimates can be computed by means of the EM-algorithm. The complete data are $\mathbf{X}$ and $\mathbf{Z}$. The Z_{tj}'s are considered as a special type of latent random variables because they are mapped in observed random variables by means of the identity function (which is the condensation rule in this case). For the conjunctive Rasch model, the function to be maximized in the M-step is the conditional expected value of $\ln g(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\delta})$, as defined for the ML estimates, *plus* the loglikelihood of $\mathbf{Z}$. Denoting the PDF of $\mathbf{Z}$ by $m(\mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\delta})$, this loglikelihood can be shown to be the following:

$$\ln m(\mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\delta}) = \sum_{k=1}^{K} \left\{ -2 \times \sum_{p=1}^{P} \ln [1 + \exp (\beta_{pk})] \right.$$

$$\left. - 2 \times \sum_{i=1}^{I} \ln [1 + \exp (\delta_{ik})] + \sum_{p=1}^{P} \beta_{pk} + \sum_{i=1}^{I} \delta_{ik} \right\}.$$

As is the case for $\ln g(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\delta})$, $\ln m(\mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\delta})$ is also an exponential family loglikelihood. Therefore, the sum of the conditional expected value of $\ln g(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\delta})$ and $\ln m(\mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\delta})$ has a unique maximum that can be found using the same optimization method as the one that maximizes $\ln g(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\delta})$.

Example: Testing a Cognitive Theory About Responses to Open Synonym Items

The Data

The data that were analyzed are a subset of a larger data set that was collected by Janssen, Hoskens and De Boeck (1993). These data are responses to the three types of items that were presented in the first section, namely, the generating, evaluating, and open synonym items. From the larger data set, 22 stimulus words (Dutch language) were selected, each of which was associated with one open synonym, one generating and one evaluating item. These 22 stimulus words are one of six sets of stimulus words. They were chosen because, as a group, their fit to the Rasch model (for the generating and the evaluating items) was slightly better than for the other sets. Each set was responded to by a different group of subjects. All stimulus words were verbs (e.g., to understand, to change, to erase, to hug, to shine).

The subjects were allowed to give *multiple* synonyms in response to the open synonym items, and to evaluate *multiple* synonym candidates as being correct synonyms in response to the evaluation items. The items were formulated in this way to stimulate the subjects to evaluate every synonym candidate in an absolute way instead of comparing them with each other (De Boeck, personal communication, August 26, 1992). In order to have an idea about the effect of this somewhat unusual formulation of the items, we calculated the average number of synonym candidates generated (for the generating items), synonym candidates evaluated as being correct synonyms (for the evaluating items), and synonyms given (for the open synonym items). These numbers were 2.914, 1.581, and 1.293, respectively. Thus, despite the fact that the difference in instructions for the generating and the open synonym items is rather small

(writing down all synonym candidates vs. only the correct ones) it is clear that the subjects responded to them differently.

The correct synonym for each word was determined by means of a dictionary. The answers were scored as being correct if the correct synonym was given by the subject (either as a synonym candidate, a multiple choice option, or an actual synonym for, resp., a generating, evaluating, or open synonym item). The alternatives in the evaluating items were the most frequently given answers to the corresponding generating items plus the correct synonym.

The subjects were fifth and sixth graders (mostly 17–18 years of age) from six schools of general secondary education in the Dutch speaking part of Belgium. A total of 258 subjects responded to 66 items (three sets of 22) in total. Two subjects were removed from the present sample: one subject did not follow the instructions properly, and another subject responded correctly to all evaluating items.

All subjects responded to the generating items first. After a delay of approximately six weeks the open synonym items were administered. About half of the subjects were given the evaluating items about six weeks after the open synonym items. Due to practical circumstances, the other half of the subjects were given the evaluating items right after the open synonym items in the same session.

Method of Analysis

Estimation. The responses to the open synonym items were analyzed by means of the two-component conjunctive Rasch model. The parameter estimates were computed by means of the EM-algorithm described previously. With respect to this analysis, four points have to be mentioned.

First, in the analysis the subjects were *grouped*, and it was assumed that every subject within a group has the same pair of person parameters. The grouping was performed assuming that only the abilities involved in the generating and the evaluating subtasks may be involved in the open synonym items. As such, the approach is a confirmatory one. Given this assumption, grouping of the subjects according to their sum scores on the generating and evaluating items, which are the minimal sufficient statistics for the person parameters, is asymptotically equivalent (for increasing number of items) to the case of no grouping. Since the grouping has to be performed according to *two* sum scores jointly, every subject was assigned to a group according to his or her sum score *pair*. There were 14 different sum scores in both the generating and the evaluating data set, and there were 98 different sum score pairs (out of a possible number of $14 \times 14 = 196$). Because of this grouping, only 14 generating and 14 evaluating person parameters were estimated; the responses by the persons in each of the 98 groups were assumed to be governed by person parameter *pairs* of which the first element was one of the 14 generating person parameters and the second was one of the 14 evaluating person parameters. There are two reasons for this grouping. First, it reduces the number of data points, and the number of parameters per component, both of which are too large to be handled by the computer program in which the EM-algorithm was implemented. Second, it results in more reliable parameter estimates, since the amount of information per parameter is larger.

Second, no finite ML estimates of the conjunctive Rasch model parameters could be found. Therefore, we will only consider MAP estimates.

The third point concerns the uniqueness of the solution of the likelihood equations (better denoted as *posterior PDF equations* in the case of MAP estimates). As yet, no conclusive evidence exists for the uniqueness of this solution. Therefore, in order to get

some reassurence on this point, the MAP estimates were computes 50 times[1], each time using different random starting values. The algorithm always converged to the same solution point.

The fourth and last point concerns the goodness-of-recovery of the true parameter values. Goodness-of-recovery is related to consistency in that consistency involves that asymptotically goodness-of-recovery is perfect. Because we are dealing with a *fixed-effects* (unconditional, nonmarginal) conjunctive Rasch model, we cannot rely on the standard consistency property of ML estimates (and MAP estimates, as their approximations). Moreover, even if consistency could be proved, we still would not know how large the standard errors and the biases are. Therefore, a small simulation study was performed, the results of which are presented in the following.

Besides the parameters of the conjunctive Rasch model, the parameters of the ordinary Rasch model also had to be estimated (i.e., for the generating and the evaluating subtasks). For these parameters, also MAP estimates were computed, using the same prior as for the parameters of the conjunctive Rasch model.

A goodness-of-recovery study. In order to get results that are relevant for the present data, we generated data sets of the same structure as the original data set (with respect to the number of subjects in each group and the number of items), and used the MAP estimates from the original data set as the true parameter values to generate them. A total of 100 data sets was generated.

For every true parameter value, the mean and the standard deviation (further denoted as *standard error*) of its 100 estimates was computed. As an estimate of each parameter's bias, we computed the absolute value of the difference between the true parameter value and the mean of its estimates. In order to evaluate the sizes of the estimated biases and standard errors, for each replication, we computed the standard deviation of the parameter estimates, separately for each of the four types of parameters (i.e., generating-person, generating-item, evaluating-person, evaluating-item). Further, for each replication, we computed the product-moment correlation between the true and the estimated parameter values, again separately for each of the four types. This statistic was computed because in the following we will consider correlations between different sets of parameter estimates (i.e., using subtask and open synonym data). Now, in order to make an overall evaluation of the results, the following averages were computed: the average bias and standard error, separately for each of the four parameter types (over the parameters of each type), and the average standard deviations and correlations (over the replications). These statistics are shown in Table 1.

For the generating component, parameter recovery is good. The only weak point is the bias in the item parameter estimates. For the evaluating component, the results are less good, although not unacceptable. Again, person parameter recovery is better than item parameter recovery. With respect to the latter, both bias and standard error are substantial. As will be shown in the following, the inferior results for the evaluating component are due to the relative easiness of this component, resulting in less statistical information.

Results and Discussion

Goodness-of-fit. Goodness-of-fit was examined by means of generalized likelihood ratio (GLR) and Pearson chi-square statistics (Mood, Graybill, & Boes, 1974,

---

[1] Actually, an algorithm that is numerically more efficient than the algorithm described in the previous paragraph, was used for this purpose. However, this is of no importance for the problem being examined.

TABLE 1

Goodness–of–Recovery Statistics for the Four Types of

Conjunctive Rasch Parameter Estimates (see text for explanation)

	Generating Component		Evaluating Component	
	Person Parameters	Item Parameters	Person Parameters	Item Parameters
Bias	0.00774	0.13809	0.11795	0.46315
Standard Error	0.13916	0.31279	0.37819	0.61567
Standard Deviation	0.42468	1.84625	0.83246	0.81231
Correlation	0.94092	0.98386	0.88185	0.62708

p. 419, p. 440–448). With respect to the Pearson chi-square statistic, we had to consider the fact that most of the sum score pair and some single sum score groups only contain a small number of subjects (e.g., 1, 2, 4), because for this reason the chi-square distribution cannot be expected to be approximated well. Therefore, sum score pair (for the open synonym items) and single sum score (for the subtasks) groups were joined. Each of the three data sets was arranged in one vector. For the generating and evaluating items, the data were arranged according to, respectively, generating and evaluating sum score (in descending order). For the open synonym items, the data were first arranged according to generating sum score, and within every generating sum score according to evaluating sum score. Then, separately for the responses to each of the items, joining occurred across adjacent sum score (pair) groups. This can be expressed as follows:

$$Q = \sum_{i=1}^{I} \sum_{g=1}^{G_i} \left[\frac{(o_{ig+} - e_{ig+})^2}{e_{ig+}} + \frac{(o_{ig-} - e_{ig-})^2}{e_{ig-}} \right]. \tag{21}$$

In (21), Q denotes the Pearson chi-square statistic. The groups are indexed by g, and their total number is denoted by G_i (indicating that the grouping may be different for different items). Further, o_{ig+} and o_{ig-} denote, respectively, the observed number of correct and incorrect responses to item i by the persons in group g. Their expected values, computed at the MAP estimates, are denoted by e_{ig+} and e_{ig-}, respectively. Finally, the criterion of the grouping was that, for all i and g, both e_{ig+} and e_{ig-} had to be larger than 5.

The Pearson chi-square statistic will be evaluated by locating its value in the

chi-square distribution with degrees of freedom equal to the number of free cells (i.e., the sum of the G_i's over i) minus the number of estimated parameters.

We will first consider the results on the goodness-of-fit of the Rasch model for the generating and the evaluating items. For the generating items, the Q-statistic equals 195.608 (df = 184), which corresponds to a p-value of 0.265. And for the evaluating items, the Q-statistic equals 222.59 (df = 203), which corresponds to a p-value of 0.165. Thus, for both the generating and the evaluating items, the Rasch model cannot be rejected. The same conclusion was reached when these data were analyzed by means of the computer program RIDA (Glas, 1990), which computes some Pearson chi-square statistics using conditional maximum likelihood estimates (see Glas, 1988).

We will now consider the main hypothesis of this study, namely the hypothesis that the conjunctive Rasch model involving the two ordinary Rasch models for the generating and the evaluating items as models for its latent random variables, holds for the open synonym items. The GLR-statistic can be used here, because it is possible to formulate this hypothesis as a comparison between two hierarchically related models for the joint set of responses to the three types of items. The non-restricted model involves two ordinary Rasch models for the generating and the evaluating items, and a non-restricted conjunctive Rasch model for the open synonym items. The restricted model differs from this latter model in that the person and item parameters of the two components of the conjunctive Rasch model are restricted to be identical to the corresponding parameters of the two ordinary Rasch models for the subtasks. Computing the ML or MAP estimates of the parameters of this restricted model involves formula's that are somewhat different from the ones presented in the previous section, but not in an essential way. Now, the GLR-statistic equals 100.2 (df = 70), which corresponds to a p-value of 0.0104.

We can test the same hypothesis by means of the chi-square statistic defined in (21). In this case the number of items (I) equals 66. Now, for the unrestricted model the Q-statistic was 631.725 (df = 557), which corresponds to a p-value of 0.013 (using asymptotic normality). And for the restricted model the Q-statistic was 986.918 (df = 611), which corresponds to a highly significant p-value of almost 0.0 (using asymptotic normality). Thus, whereas the GLR-statistic resulted in a p-value on the boundary of significance, the Pearson chi-square statistic clearly rejects the null hypothesis.

A Comparison of parameter estimates.   It is possible to examine to main hypothesis of this study in more detail than by means of the goodness-of-fit statistics described above, namely by comparing the parameter estimates from the ordinary Rasch analyses of the subtask data with the parameter estimates from the conjunctive Rasch analysis of the open synonym data. In the following, we will consider comparisons between five sets of Rasch model parameter estimates: two from the ordinary Rasch analyses of the subtask data, two from the two-component conjunctive Rasch analysis of the open synonym data, and one from the ordinary Rasch analysis of these latter data (see further for the reason why this latter analysis was performed). Then, since each comparison can be made both with respect to the person and the item parameters, a total of $2 \times \binom{5}{2} = 20$ comparisons can be made.

The most obvious statistic for describing the relation between two sets of parameter estimates is the product-moment correlation coefficient. However, the hypothesis of identical parameter values for the observed and the latent responses also implies that the means and the standard deviations (SDs) of the two sets of parameter estimates have to be the same. Comparing to the means of the item parameter estimates only makes sense if the same identifiability restriction is imposed on the two sets of *person*

TABLE 2

Correlations Between the Item Parameter Estimates

Computed on the Responses to the Generating and the Evaluating

(row entries) and the Open Synonym Items (column entries)

Item Type			Open Synonym Items		
	Type of Analysis		Conjunctive Rasch		Ordinary Rasch
		Parameter Set	Generating	Evaluating	
Generating Items	Ordinary Rasch		0.931	0.377	0.956
Evaluating Items	Ordinary Rasch		0.428	0.257	0.448

parameter estimates. In all five sets of parameter estimates, the mean of the person parameter estimates was restricted to zero.

In Tables 2 to 5 we present correlations between the different sets of item and person parameter estimates. And in Table 6 we present their means and SDs.

Distinguishability. In order for the conjunctive to be well-distinguishable from the ordinary Rasch model, it is necessary that the two *latent* ordinary Rasch models are well-distinguishable, which means that the item and the person parameter estimates of the two models have to be different. This appears to be the case. In particular, the generating and evaluating person parameters of the latent ordinary Rasch models correlate 0.175 (see Table 5), and the corresponding item parameters 0.169 (see Table 3). The corresponding correlations obtained using the subtask data are also low; 0.262 for the person, and 0.451 for the item parameters (not in the tables).

Item parameters—Ordinary Rasch estimates of subtask data versus conjunctive Rasch estimates of open synonym data. The crucial comparisons in this study are the ones between the parameter estimates from the ordinary Rasch analyses of the subtask data and the parameter estimates from the conjunctive Rasch analysis of the open synonym data. In the building up of an overall interpretation, it is useful to consider first the results on the item parameters. For the generating component the crucial correlation is 0.931, and for the evaluating component it is 0.257 (see Table 2). The means of the ordinary Rasch and conjunctive Rasch generating item parameter estimates are 0.947 and 0.587, respectively (see Table 6), which involves a difference that is significantly (with respect to the asymptotic sampling distribution) different from zero ($p =$

TABLE 3

Correlations Between the Item Parameter Estimates

Computed on the Responses to the Open Synonym Items

Item Type			Open Synonym Items		
	Type of Analysis		Conjunctive Rasch		Ordinary Rasch
		Parameter Set	Generating	Evaluating	
Open Synonym Items	Conjunctive Rasch	Generating	1.0	0.169	0.984
		Evaluating	0.169	1.0	0.311
	Ordinary Rasch		0.984	0.311	1.0

0.001, two-sided). Further, the SDs of these estimates are 1.676 and 1.893 for the ordinary Rasch and conjunctive Rasch parameters, respectively (see Table 6).

For the evaluating item parameters, the differences are much larger. In particular, the means of the item parameter estimates are -1.244 and -2.382, for the ordinary Rasch and conjunctive Rasch parameters, respectively (see Table 6), which is significantly different from zero ($p \approx 0.0$, two-sided). Further, the SDs of the item parameter estimates are 1.231 and 0.779, respectively (see Table 6).

From the results on the item parameters, and ignoring the small difference between the mean generating item parameter estimates, it appears that the psychological process underlying the responses to the open synonym items involves a generating process that is the same as the one that is made observable in the generating items. And the evaluating process to be made observable in the evaluating items, appears to be not involved at all in the responses to the open synonym items. In this, we take into account the fact that the evaluating item parameter recovery was probably less than generating item parameter recovery (i.e., we can expect a correlation of about 0.627, see Table 1).

An ordinary Rasch model for the open synonym data? From the results discussed above, it would be obvious to formulate the hypothesis that open synonym items are no more than generating items which have to be responded to only by giving the correct synonyms (if available). This way of processing the items is possible if the correctness of a synonym candidate is immediately obvious to the subject Then, it would follow that the ordinary Rasch model for the generating items should also hold for the open synonym items.

TABLE 4

Correlations Between the Person Parameter Estimates

Computed on the Responses to the Generating and the Evaluating

(row entries) and the Open Synonym Items (column entries)

Item Type			Open Synonym Items		
	Type of Analysis		Conjunctive Rasch		Ordinary Rasch
		Parameter Set	Generating	Evaluating	
Generating Items	Ordinary Rasch		0.946	0.171	0.573
Evaluating Items	Ordinary Rasch		0.246	0.696	0.299

Considering the fit of the ordinary Rasch model for the open synonym items, one has to keep in mind that the ordinary Rasch model is a special case of the conjunctive Rasch model. In particular, it holds that an ordinary Rasch model is the limiting case of a two-component conjunctive Rasch model in which the person parameters of the other *latent* ordinary Rasch model increase relative to the item parameters. In this case, the (latent) response probabilities of this latter ordinary Rasch model converge to 1.0. Now, for our data, the mean estimated probabilities for the latent generating and evaluating components are 0.431 and 0.875, respectively. Thus, one can say that the conjunctive Rasch model for the open synonym items is *almost* an ordinary Rasch model.

The most obvious way to test the hypothesis that the open synonym items are merely generating items in disguise, is by performing an ordinary Rasch analysis of the open synonym data, and by comparing the resulting estimates to the ones obtained using the subtask data. Now, the correlation between the ordinary Rasch open synonym item parameter estimates and the ordinary Rasch generating item parameter estimates is 0.956 (see Table 2). The average item parameter estimates are 0.947 and 0.935 (see Table 6), for the generating and the open synonym items, respectively (no significant difference; $p = 0.902$). Further, the SDs of these two sets of estimates are 1.676 and 1.797, respectively (see Table 6). The correlation with the Rasch evaluating item parameter estimates is 0.448 (see Table 2), and as can be seen from the values given above, the means and the SDs are substantially different. Thus, these results are completely in line with the hypothesis that open synonym items are no more than generating items which have to be responded to by only giving the correct synonyms (if available).

TABLE 5

Correlations Between the Person Parameter Estimates

Computed on the Responses to the Open Synonym Items

Item Type			Open Synonym Items		
	Type of Analysis		Conjunctive Rasch		Ordinary Rasch
		Parameter Set	Generating	Evaluating	
Open Synonym Items	Conjunctive Rasch	Generating	1.0	0.175	0.598
		Evaluating	0.175	1.0	0.313
	Ordinary Rasch		0.598	0.313	1.0

Person parameters—Ordinary Rasch estimates of subtask data versus conjunctive Rasch estimates of open synonym data. The discussion of the results here is completely parallel to the one for the item parameters. In particular, the important correlations are the ones between the ordinary Rasch and conjunctive Rasch person parameter estimates, which are 0.946 and 0.696 for the generating and the evaluating parameters, respectively (see Table 4). The SDs of the ordinary Rasch and conjunctive Rasch generating person parameter estimates are 0.763 and 0.4, respectively (see Table 6). It is not clear how this difference should be interpreted, and in the following we will ignore it. Further, for the evaluating person parameter estimates, the corresponding SDs are 0.835 and 0.881 (see Table 6). No comparisons between mean person parameter estimates could be made, because of the particular identifiability restriction that was imposed.

From these results it follows that the process underlying the responses to the open synonym items involves a generating process that not only involves the same item parameters, but also the same latent ability as the one that is made observable in the generating items, and an evaluating process that involves different item parameters, but approximately the same latent ability as the one that is made observable in the evaluating items. As follows from the mean conjunctive Rasch generating and evaluating item parameter estimates (0.587 and −2.382, respectively) and from the mean estimated generating and evaluating probabilities (0.431 and 0.875, respectively), this evaluating process is much easier than the generating process.

Actually, it should not be too surprising that the ordinary Rasch and conjunctive Rasch evaluating item parameters are different. In fact, the synonym candidates presented in the evaluating items are probably very different from the lists of synonym

TABLE 6

Means and Standard Deviations of the Item and Person Parameter Estimates

Computed on the Responses to the Generating, Evaluating, and Open Synonym Items

Item Type	Type of Analysis	Parameter Set	Item Parameters		Person Parameters
			Mean	SD	SD
Generating Items	Ordinary Rasch		0.947	1.676	0.763
Evaluating Items	Ordinary Rasch		−1.244	1.231	0.835
Open Synonym Items	Conjunctive Rasch	Generating	0.587	1.893	0.4
		Evaluating	−2.382	0.779	0.881
	Ordinary Rasch		0.935	1.797	0.666

candidates generated and evaluated covertly by most subjects. The small variation in the conjunctive Rasch evaluating item parameter estimates suggests that the evaluating difficulties of the covertly generated lists of synonym candidates are unrelated to the stimulus words. This obviously has to result in the fact that the variance of the conjunctive Rasch evaluating item parameter estimates is nothing but error variance.

Ordinary Rasch estimates of open synonym data from the perspective of the choice for the wrong model. Given the evidence for an evaluating process underlying the responses to the open synonym items, it is good to reconsider the fact that the latent generating process is a little bit *easier* than the observable one, whereas the mean *ordinary* Rasch open synonym item parameter estimate is almost *equal* to the mean ordinary Rasch generating item parameter estimate. This near-equality probably is an artifact resulting from the fact the easier (than in the subtask) generating process involved in the open synonym items compensates for the (relatively rare) incorrect responses on the second component. This compensation mechanism does not have a strong effect on the correlation between the ordinary Rasch generating and open synonym item parameter estimates, because the SD of the conjunctive Rasch evaluating item parameter estimates is small in comparison with the SD of the conjunctive Rasch generating item parameter estimates (SDs of 0.779 versus 1.893).

In a similar way as for the item parameter estimates, the results of the comparisons involving the ordinary Rasch open synonym *person* parameter estimates can be ex-

plained as resulting from the choice of the wrong model for the open synonym data. In particular, the correlation between the ordinary Rasch open synonym and generating person parameter estimates is only 0.573 (see Table 3). This low correlation is in accordance with the relatively large SD of the conjunctive Rasch evaluating person parameter estimates (0.88 versus 0.4 for the conjunctive Rasch generating person parameter estimates). Thus, whereas the small SD of the conjunctive Rasch evaluating *item* parameter estimates could not keep down the correlation between the ordinary Rasch open synonym and generating item parameter estimates, the same thing *was* possible for the corresponding *person* parameter estimates.

Finally, the low correlation between the ordinary Rasch open synonym and evaluating person parameter estimates (i.e., 0.299; see Table 3) can be explained in a similar way. In particular, given that there is so little variation in the correctness of the latent evaluating responses in comparison with the variation in the correctness of the latent generating responses, one cannot expect the evaluating ability to show up in an ordinary Rasch analysis of the open synonym data. This small variation in the correctness of the latent evaluating *responses* can exist despite the large variation in the evaluating *person parameters*, because the latent evaluating process is so easy.

Conclusion. Concluding, we can put forward a psychological theory about the process involved in solving open synonym items, which accounts for almost all of the available evidence. Solving an open synonym item involves a process of generating synonym candidates. This process involves the same latent ability as the one involved in responding to generating items. However, the item difficulties of these two generating processes (one involved in solving open synonym item, and the other in responding to generating items) differ by a small constant quantity (for all items); the latent generating process is a little bit easier than the observable one.

The second process involved in solving open synonym items is evaluating the correctness of the synonym candidates. This process involves approximately the same latent ability as the one involved in the evaluating items. However, the item difficulties of these two evaluating processes are completely different. This is most likely due to the fact that the synonym candidates presented in the evaluating items are different from the lists of synonym candidates generated covertly by most subjects.

Comments on the Method of Analysis

One has to keep in mind that the subjects were grouped according to their sum scores on the generating and evaluating items. Because of this fact, the approach is a confirmatory one, whereas an exploratory approach is to be preferred (preferred with respect to the analysis of the open synonym data; comparing the resulting parameter estimates with those from the analyses of the subtask data, which is also a confirmatory approach, is *not* considered to be less preferrable than a study without such a comparison). This exploratory approach would involve an application of the conjunctive Rasch model to the open synonym data *without* making use of external information on the underlying dimensions. Because of technical (too many parameters and data points) and statistical (incidental parameters, consistency) problems, using the fixed-effects (unconditional) conjunctive Rasch model in this exploratory way is not advisable. Clearly, what is needed here is a marginal version of the conjunctive Rasch model, which solves both of the above problems but confronts us with the technical problems involved in multivariate numerical integration.

Using the conjunctive Rasch model in this exploratory way does not preclude making use of subtask data in order to test hypotheses about the relations between the underlying dimensions, but this second part is simply not needed. Without making use

of subtask data, this exploratory approach is, in essence, *conjunctive item factor analysis*, and as such an alternative for *compensatory* item factor analysis (Bock & Aitkin, 1981; Muthén, 1978). An example of this approach is given by Embretson (in press) who also presents a method for examining the nature of the underlying dimensions that does not make use of subtask data.

With respect to an analysis by means of the conjunctive Rasch model, either in its fixed-effects or its marginal version, there are two more points that still require some study. First, there is the question of the uniqueness of the solution of the likelihood equations (or posterior PDF equations). Although multiple analyses using random starting values (as in this example) can give some reassurence on this point, it is clear that only analytical evidence is fully satisfactory. An interesting question in this respect is whether the conjunctive Rasch model suffers from a similar weakness as the so-called *rotational invariance* of the item factor analysis model, which is for a great deal responsible for the fact that the psychological interpretation of the underlying dimensions is largely a subjective matter.

Second, there is the question of the goodness-of-recovery of the true parameter values (involving consistency, bias and sampling variance). Working with some marginal version of the conjunctive Rasch model gives us consistent ML estimates under some weak regularity conditions (see Kiefer & Wolfowitz, 1956). But even in this case, we still need information on the estimates' biases and sampling variances for sample sizes encountered in practice. A large-scale simulation study would give this information. An important factor in the design of such a simulation study is the component's difficulty, which determines the amount of statistical information with respect to its parameters. In particular, if this component is either very easy or very difficult, this amount of statistical information will be small (because large changes in the parameter values result in small changes in the loglikelihood). This leads to estimates having large asymptotic sampling variances. This was the case in the simulation study on which is reported in this paper: the mean conjunctive Rasch generating and evaluating item parameter estimates (which served as the true parameter values in the simulation study) were, respectively, 0.587 and −2.382 (see Table 6), which correspond to mean estimated latent response probabilities for the generating and evaluating components of, respectively, 0.431 and 0.875. Therefore, it is no surprise that the goodness-of-recovery for the parameters of the evaluating component was worse than for the generating component (see Table 1).

References

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika, 71*, 1–10.
Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM-algorithm. *Psychometrika, 46*, 443–459.
Billingsley, P. (1986). *Probability and measure*. New York: Wiley.
Carpenter, P. A., & Just, M. A. (1986). Spatial ability: An information processing approach to psychometrics. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 3, pp. 221–253). Hillsdale, NJ: Lawrence Erlbaum.
De Boeck, P. (1988). *A psychometric study of cognitive components in verbal intelligence tasks* (Research proposal OT-89-9). Leuven: University of Leuven, Research Group on Quantitative Methods.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1–38.
Embretson, S. (1980). Multicomponent latent trait models for ability tests. *Psychometrika, 45*, 479–494.
Embretson, S. (1984). A general latent trait model for response processes. *Psychometrika, 49*, 175–186.
Embretson, S. (in press). Working memory versus general control processes in intelligence. *Intelligence.*
Fischer, G. H. (1974). *Einfuhrung in die Theorie psychologischer Tests* [Introduction to the theory of psychological tests]. Bern: Huber.

Fischer, G. H. (1981). On the existence and uniqueness of maximum likelihood estimates in the Rasch model. *Psychometrika, 46,* 59–77.

Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization.* New York: Academic Press.

Glas, C. A. W. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika, 53,* 525–546.

Glas, C. A. W. (1990). *Rasch incomplete design analysis* [Computer program]. Arnhem, The Netherlands: National Institute for Educational Measurement (CITO).

Goldman, S. R., & Pellegrino, J. W. (1984). Deductions about induction: Analyses of developmental and individual differences. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 149–198). Hillsdale, NJ: Lawrence Erlbaum.

Jannarone, R. J., Yu, K. F., & Laughlin, J. E. (1990). Easy Bayes estimates for Rasch-type models. *Psychometrika, 55,* 449–460.

Janssen, R., Hoskens, M., & De Boeck, P. (1993). An application of Embretson's multicomponent latent trait model to synonym tests. In R. Steyer, K. F. Wender, & K. F. Widaman (Eds.), *Proceedings of the 7th European Meeting of the Psychometric Society* (pp. 187–190). Stuttgart and New York: Gustav Fischer Verlag.

Kiefer, J., & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics, 27,* 887–903.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Maris, E. (1992). *Psychometric models for psychological processes and structures.* Unpublished doctoral dissertation, University of Leuven, Belgium.

Mayer, R. E., Larkin, J. H., & Kadane, J. B. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 231–273). Hillsdale, NJ: Lawrence Erlbaum.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics.* London: McGraw-Hill.

Muthén, B. (1978). Contributions to factor analysis of dichotomized variables. *Psychometrika, 43,* 551–560.

Novick, M. R., & Jackson, P. H. (1974). *Statistical methods for educational and psychological research.* New York: Mc-Graw-Hill.

Pellegrino, J. W., & Kail, R. (1982). Process analysis of spatial aptitude. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 311–366). Hillsdale, NJ: Lawrence Erlbaum.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: The University of Chicago Press.

Rubin, D. B., Thayer, D. T. (1982). EM algorithms for factor analysis. *Psychometrika, 47,* 69–76.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ: Lawrence Erlbaum.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–286.

Verhelst, N. D., & Verstralen, H. H. F. M. (1993). A stochastic unfolding model derived from the partial credit model. *Kwantitatieve Methoden, 42,* 73–92.

Wu, C. F. J. (1983). On the convergence properties of the EM-algorithm. *The Annals of Statistics, 11,* 95–103.