7

Application of the Saltus Model to Stagelike Data: Some Applications and Current Developments

Karen Draney and Mark Wilson

University of California, Berkeley

7.1 Background of the Saltus Model

The saltus model was developed in dichotomous form by Wilson (1989), and expanded to polytomous form by Draney (1996) as a method for detecting and analyzing discontinuities in performance that are hypothesized to occur as a result of rapidly occurring person growth (e.g.,Fischer, Pipp, & Bullock, 1984). Such discontinuities are often theorized to occur as the result of progression through developmental stages or levels. The most influential such theory was developed by Jean Piaget (e.g., Piaget, 1950; Inhelder & Piaget, 1958). Although Piagetian theory has been somewhat controversial of late (e.g., Lourenço & Machado, 1996), there is still a strong interest in stage-like development in a number of areas, including moral and ethical reasoning (e.g., Dawson, 2002; Kohlberg & Candee, 1984), evaluative reasoning (e.g., Dawson-Tunik, 2002; Armon, 1984), adult development (e.g., Commons et al., 1998; Fischer, Hand, & Russel, 1984), and cognitive development (e.g., Bond, 1995b,a; Bond & Bunting, 1995; Demetriou & Efklides, 1989, 1994; Hiele, 1986).

The work of Piaget describes the cognitive developmental stages through which children progress as they grow. In particular, school-age children progress from the preoperational stage, through the concrete operational stage, to the formal operational stage. In the preoperational stage, children are able for the first time to produce mental representations of objects and events, but unable to consistently perform logical mental operations with these representations. In the concrete operational stage, children are able to perform logical operations, but only on representations of concrete objects. In the formal operational stage, which starts to occur around the beginning of adolescence, children are able to perform abstract operations on abstractions as well as concrete objects.

According to Piaget, progress from stage to stage is characterized by more than simple linear growth in reasoning ability. The transition from one stage to another involves a major reorganization of the thinking processes used by children to solve various sorts of problems. Theories with similar structure, but perhaps different substantive focus, are described by the many neo-Piagetian researchers, and by other researchers who use stage-based theories.

Researchers in the Piagetian tradition are using increasingly complex statistical and psychometric models to analyze their data. Béland & Mislevy (1996) analyze proportional reasoning tasks using Bayesian inference networks. Noelting et al. (1995) discuss the advantages of Rasch scaling for the understanding of Piagetian tasks. Bond (1995b,a) discusses the implications of RMs for Piagetian theory and philosophy.

In addition, researchers in psychometrics have begun wrestling with the problem of developing and applying models with sufficient complexity to address such substantive issues. For example, the three-parameter model has been used diagnostically by researchers such as Yen (1985). She describes patterns of problematic item fit that are sometimes observed in analyzing complex data and asserts that these may be indicators for increasing item complexity. Differences in item complexity such as she describes could potentially be indicative of a set of items that represent more than one developmental stage.

Another approach to the problem of incorporating different response patterns and their associations with classes is given by latent-class modeling. For example, Dayton & Macready (1976) applied this approach to behavioral hierarchies of the type often seen in developmental theories. In this approach, each underlying class is represented by a set of response probabilities to the items in question. Whereas Yen's (1985) research might be considered exploratory, latent-class theory can be used in a more confirmatory way. Additional examples of such models and their uses are given in Rijmen & De Boeck (2003), Formann (1992), and Croon (1990),

However, Rost (1988) states that the defining feature of latent-class models is the characteristic that all persons within a latent class have the same probabilities of answering a set of items correctly, and thus (if considered in an educational context) the same ability or proficiency. It is plausible that children within a given developmental stage might vary in overall proficiency within that stage.

The saltus model was developed to combine the advantages of the RM, including varying person proficiency, and latent-class modeling, including differing patterns of response probability across different latent subgroups of persons. In this way, it is similar in its origins to the HYBRID model (Yamamoto, 1989; compare also Chapter 4 in this volume), a latent-class model that included as one of the classes a "catch-all" latent-trait model for those persons who did not fit well into one of the other classes. However, unlike the HYBRID model, the saltus model provides a latent-trait model within each of the latent classes identified.

7.2 The Saltus Model

The saltus model is based on the assumption that there are C classes, representing developmental stages or levels. Each level is represented by a set of items, which are constructed such that only persons at or above the developmental stage represented by those items are fully equipped to answer them correctly, and once persons enter that developmental stage, they should gain a substantial advantage in answering those items.

In the discussion to follow, the terms "person class" and "item group" will be used. This is merely a device used for clarity when it is necessary to differentiate between classes of persons and groups of items, and does not have any particular substantive significance.

The saltus model assumes that all persons in class c answer all items in a manner consistent with membership in that class. However, persons within a class may differ by proficiency. In a Piagetian context, this means that a child in, say, the concrete operational stage is always in that stage, and answers all items accordingly. The child does not show formal operational development for some items and concrete operational development for others. However, some concrete operational children may be more proficient at answering items than are other concrete operational children.

In the saltus model, two parameters describe a person v: a unidimensional proficiency parameter θ_v, and an indicator vector for class membership $\boldsymbol{\phi}_v$. If there are C latent person classes, then $\boldsymbol{\phi}_v = (\phi_{v1}, \ldots, \phi_{vC})$, where ϕ_{vc} takes the value of 1 if person v is in class c and 0 if not. Note that only one ϕ_{vc} is theoretically nonzero; however, since it is a latent parameter, it must in practice be estimated.

Just as persons are members of only one class, items are associated with one and only one group. In a developmental context, an item's group would be said to be the first developmental stage at which a child would have all of the skills necessary to perform that item correctly. It is, of course, possible for children at lower developmental stages to perform items correctly from time to time; however, this usually occurs because of guessing or a poorly developed strategy that happens to produce the correct answer in some cases. Unlike person-class membership, however, which is unknown and must be estimated, item-group membership is known a priori, based on the theory that was used to produce the items. It will be useful to denote item-group membership by the indicator vector $\mathbf{b}_i$. As with person classes indicated by the $\boldsymbol{\phi}_v$, we assume that there are C item groups, and each item is member of exactly one group, i.e., $\mathbf{b}_i = (b_{i1}, \ldots, b_{iC})$, when $\mathbf{b}_{ic}$ takes the value of 1 if item i belongs to item class k, and 0 otherwise. The set of all $\mathbf{b}_i$ is denoted by $\mathbf{b}$.

The equation

$$P\left(X_{vij}=j|\theta_v, \phi_{vc}=1, \beta_i, \tau_{ck}\right)=\frac{\exp\sum_{s=0}^{j}\left(\theta v-\beta_{is}+\tau_{ck}\right)}{\sum_{t=0}^{m_i}\exp\sum_{s=0}^{t}\left(\theta_v-\beta_{is}+\tau_{ck}\right)} \tag{7.1}$$

defines the probability of response j to item i, with step difficulty β_{ij}. This defines a polytomous item response model that has been augmented by the introduction of the saltus parameter τ_{ck} as an additive element of the logistic argument. The saltus parameter describes the additive effect—positive or negative—for people in class c on the item parameters of all items in group k. Typically, in developmental contexts involving stages, this has taken the form of an increase in probability of success at higher levels as the person achieves the stage at which an item is located, indicated by $\tau_{ck}>0$ when $c \geq k$ (although this need not be the case). The saltus parameters can be represented as a C x C matrix **T**.

The probability that a person with parameters ϕ_v and θ_v will respond in category j to item i is given by

$$P\left(X_{vij}=j|\theta_v, \phi_v, b_i, \beta_i, T\right)=\prod_h\prod_k P(X_{vij}=j|\theta_v, \phi_{nh}=1, b_i, \tau_{hk})^{\phi_{vh}b_{ik}}. \tag{7.2}$$

Note that for only one combination of c and k do the product terms have a nonzero exponent $\phi_{vc}b_{ik}$. Since item responses are assumed to be independent given θ_v, ϕ_v, and all of the item and saltus parameters, the model-based probability of a response vector is

$$P\left(X_v =_v |\theta_v, f_v, b_i, \beta_i, T\right)=\prod_h\prod_k\prod_i P(X_{vij}=x_{ij}|\theta_v, \phi_{vh}=1, b_i, \tau_{hk})^{\phi_{vh}b_{ik}}. \tag{7.3}$$

The saltus model requires a number of constraints on the parameters. For item-step parameters, we use two traditional constraints: first, $\beta_{i0}=0$ for every item, and second, the sum of all the β_{ij} across all items is set equal to zero. Some constraints are also necessary on the saltus parameters. This could be accomplished in several ways, but once parameters have been estimated with one set of restrictions, they can be translated to corresponding values under another set. The set of constraints we have chosen is the same as that used by Mislevy & Wilson (1996), and will allow us to interpret the saltus parameters as changes relative to the first (lowest) developmental stage. Two sets of constraints are used. First $\tau_{c1}=0$; thus, the difficulty of the first (lowest) group of items is held constant for all person classes; changes in the difficulty of groups of items for $k>1$ are interpreted with respect to this first group of items for all person classes. Also $\tau_{1k}=0$; thus, items as seen by

person classes with $c>1$ will be interpreted relative to the difficulty of those items as seen by person class 1.

The saltus model is a special case of the more general *mixed RM* described by Rost (1990, compare also Chapter 6 in this volume). The estimation of polytomous mixed RMs with and without constraints using conditional maximum likelihood methods is discussed in von Davier & Rost (1995). This model is itself a member of the class of finite mixture-distribution models (e.g.,Titterington et al., 1985; Everitt & Hand, 1981). Perhaps the most general of such models to have been discussed in an educational context is the mixture multidimensional random coefficients multinomial logit (M^2RCML) model described by Pirolli & Wilson (1998).

7.3 An Example Application

An example of the application of the saltus model will be based on a set of responses to Noelting's (1980a; 1980b) orange juice mixtures test for assessing proportional reasoning. The items in this test consist of pictures of a certain number of glasses of juice and glasses of water, representing a mixture. In each item, the child is shown two such mixtures and asked which would taste more strongly of juice, or if they would taste the same. A representation of such an item is shown in Figure 7.1.

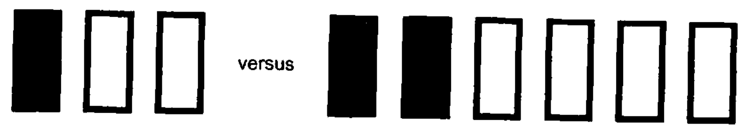

Fig. 7.1. Representation of Noelting juice mixture item. Dark indicates juice, light indicates water.

Noelting postulates a Piagetian stage hierarchy consisting of three stages—the intuitive, the concrete operational, and the formal operational—for persons solving these items. Noelting develops juice mixture problems to represent the skills that differentiate between each developmental stage.

In the intuitive stage, the child can additively compare the relative quantity of an attribute (e.g., more glasses of juice or more glasses of water), but tends to pay attention only to one attribute or the other. In the concrete operational stage, the child begins to learn the concept of ratio and proportionality. Rather than simply comparing the number of glasses of juice or water between the two mixtures, the child is able to recognize the concept of "one glass of juice for every glass of water" or "twice as much juice as water." In the formal operational stage, the child learns to deal formally with fractions, ratios, and percentages. Here, the child begins to master the formal mathematical rules for comparing two arbitrary mixtures. For this example, we will consider the items developed for the first two stages (the intuitive and the concrete operational).

Noelting postulates three problem types (representing ordered substages) within a stage and develops between one and four replications of each of these substage problem types. These problem types and replications are described in Table 7.1. The items were administered to a sample of 460 subjects ranging in age from 5 to 17 years. The number of persons at each age is given in Table 7.2.

Table 7.1. Noelting items

Item	Stage	Mixture 1	Mixture 2
1	intuitive	4 Juice, 1 Water	1 Juice, 4 Water
2	intuitive	1 Juice, 2 Water	2 Juice, 1 Water
3	intuitive	1 Juice, 0 Water	1 Juice, 1 Water
4	intuitive	1 Juice, 2 Water	1 Juice, 3 Water
5	intuitive	2 Juice, 3 Water	1 Juice, 1 Water
6	intuitive	2 Juice, 1 Water	3 Juice, 4 Water
7	concrete	1 Juice, 1 Water	2 Juice, 2 Water
8	concrete	2 Juice, 2 Water	3 Juice, 3 Water
9	concrete	1 Juice, 2 Water	2 Juice, 4 Water
10	concrete	2 Juice, 4 Water	3 Juice, 6 Water
11	concrete	4 Juice, 3 Water	8 Juice, 6 Water
12	concrete	3 Juice, 1 Water	6 Juice, 2 Water
13	formal	3 Juice, 1 Water	5 Juice, 2 Water
14	formal	8 Juice, 3 Water	3 Juice, 1 Water
15	formal	5 Juice, 2 Water	7 Juice, 3 Water
16	formal	3 Juice, 5 Water	5 Juice, 8 Water
17	formal	1 40 %, 0 10 %, 1 Water	0 40 %, 2 10 %, 0 Water
18	formal	0 40 %, 2 10 %, 1 Water	2 40 %, 0 10 %, 4 Water
19	formal	1 40 %, 1 10 %, 1 Water	1 40 %, 0 10 %, 2 Water
20	formal	1 40 %, 1 10 %, 1 Water	2 40 %, 1 10 %, 2 Water

The saltus model to be fit to these data will be a two-stage model, comparing the intuitive and the concrete items. In this model, saltus class 1 should include the youngest children in the intuitive stage, saltus class 2 should include middle-aged children in the concrete operational stage, as well as the oldest children in the formal operational stage. In this model, there will be one between-class saltus parameter, for the older children taking the concrete operational items. This parameter is expected to be positive.

Parameter estimates and standard errors for this model are given in Table 7.3. Approximately 60% of the sample is classified into saltus class 1, and 40% into class 2. Class 1 is lower in mean proficiency than class 2. This is not surprising, since older children are in general higher in proficiency on the

Table 7.2. Ages of subjects in the Noelting sample

Age	Frequency	Percent
5	3	0,01
6	26	0,06
7	40	0,09
8	53	0,12
9	45	0,10
10	51	0,11
11	60	0,13
12	40	0,09
13	48	0,10
14	26	0,06
15	29	0,06
16	28	0,06
17	11	0,02
Total	460	

intuitive items (i.e., less prone to errors) than are younger children, in addition to having the skills necessary to solve developmentally more complex groups of items. Also as predicted, the saltus parameter τ_{22} is statistically different from zero (with magnitude more than twice its standard error), indicating that there is some systematic effect of class membership on item performance for concrete operational items.

Recall that item difficulties as shown in Table 7.3 are interpreted relative to the lowest person class. The following is an example of how the τ parameter may be interpreted. For the lowest person class, intuitive item 1 has difficulty parameter −7.08, while concrete item 1 has difficulty parameters 1.89. For person class 2, intuitive item 1 retains the same difficulty parameter (although the mean proficiency of person class 2 is higher than for class 1, and thus the probability of correct responses to the intuitive items is higher for person class 2). However, the difficulty parameter for concrete item 1 is adjusted by τ when seen by person class 2, and thus becomes $1.89 - 5.66 = -3.77$. Not only are persons in class 2 more likely to answer items correctly than are persons in class 1 (because of the higher average proficiency of class 2), the difference between the difficulties of intuitive and concrete items is greater for persons in class 1 than it is for persons in class 2. In probability terms, this means that an average person in class 1 has a .10 probability of scoring 1 on concrete item 1. If τ had been zero (and the difficulties of concrete items had been the same for person class 2 as for person class 1), an average person in class 2 would have scored 1 with a probability of approximately .58. However, because of the size of τ, the average person scores 2 on this item with a probability greater than .99.

The interpretation of item difficulties and mean abilities for classes is often easier when these parameters are displayed in a graphical form sometimes referred to as a Wright map (Wilson, 2005) in honor of its creator, Benjamin

Table 7.3. Parameter estimates and standard errors

Parameter	Estimate	S. E.
β_1	-7.08	0.532
β_2	-6.61	0.446
β_3	-4.10	0.232
β_4	-3.75	0.218
β_5	-1.97	0.168
β_6	-2.38	0.177
β_7	1.89	0.183
β_8	1.49	0.179
β_9	5.30	0.181
β_{10}	5.89	0.164
β_{11}	5.66	0.171
β_{12}	5.66	0.171
τ_{22}	5.66	0.131
μ_1	-0.26	
σ_1	2.61	
π_1	0.59	
μ_2	2.21	
σ_2	1.72	
π_2	0.41	

D. Wright, of the University of Chicago. Maps have long been used with RMs such as the partial-credit and rating-scale models, and are incorporated into many estimation software packages for these models. These maps are not meant to decide whether a particular model fits the item response data, but they are very useful in describing which persons (here which classes) and which groups of items are close when comparing (average) ability estimate and item difficulty.

A Wright map of the mean class abilities and the item difficulties as seen by each class is given in Figure 7.2. In this figure, the units of the logit scale (the scale in which parameters for this model are estimated) are shown on the extreme left side of the page. The column to the right of this contains the mean abilities of the person classes, with a range of one standard deviation on either side of each class mean. The mean of each class is represented by the letter M followed by the class number (e.g., M1 for the mean of class 1). Similarly, the upper and lower limits of the standard deviation range are represented by the letter S and the class number; these limits are connected by dashed lines to the class mean.

The difficulty levels for the various item steps as seen by each class are shown in the remaining columns. The difficulty levels for the items as seen by class 1 are shown in the column labeled "Item difficulty," under the heading

	Group 1		Group 2	
Logits	proficiency	item difficulty	proficiency	item difficulty
6.5				
6.0		C4		
5.5		C3, C5, C6		
5.0				
4.5				
4.0			S2	
3.5			|	
3.0			|	
2.5			|	
2.0	S1	C1	M2	
1.5	|	C2	|	
1.0	|		|	
0.5	|		|	C4
0.0	|		S2	C3, C5, C6
−0.5	M1			
−1.0	|			
−1.5	|			
−2.0	|	I5		I5
−2.5	|	I6		I6
−3.0	S1			
−3.5		I4		I4,C1
−4.0		I3		I3, C2
−4.5				
−5.0				
−5.5				
−6.0				
−6.5		I2		I2
−7.0		I1		I1
−7.5				

Fig. 7.2. Wright map of person distributions and item difficulties for two groups

for class 1, and similarly for class 2. More-difficult item steps and more able persons are toward the top of the page, and less-difficult item steps and less able persons are toward the bottom of the page.

The effect of the saltus parameters can be seen quite clearly in this figure: The gap between the difficulties of the intuitive items and the concrete items is substantial for class 1. While the difficulty of the intuitive items is held fixed for both classes, the difficulty of the concrete items drops for class 2, such that the easier of the concrete items are nearly identical in difficulty to the harder intuitive items for this class.

Model-based response probabilities by a person whose proficiency was equal to the mean of each class, using the estimated parameter values, are given in Table 7.4. For the intuitive items, both classes are most likely to score 1. Response probabilities for the concrete items are quite different for the two classes, with Class 1 most likely to answer incorrectly to all of the items, while Class 2 is most likely to answer correctly. Even class 2, however, has about a 10% chance of answering all but the first two concrete items incorrectly.

Table 7.4. Item difficulty and probability of correct response by item for the average ability level of two saltus groups

	Group 1		Group 2	
Item	Difficulty	$P(X = 1)$	Difficulty	$P(X = 1)$
I1	-7.08	1.00	-7.08	1.00
I2	-6.61	1.00	-6.61	1.00
I3	-4.10	0.98	-4.10	1.00
I4	-3.75	0.97	-3.75	1.00
I5	-1.97	0.85	-1.97	0.98
I6	-2.38	0.89	-2.38	0.99
C1	1.89	0.10	-3.77	1.00
C2	1.49	0.15	-4.17	1.00
C3	5.30	0.00	-0.36	0.93
C4	5.89	0.00	0.23	0.88
C5	5.66	0.00	0.00	0.90
C6	5.66	0.00	0.00	0.90

An example set of person-response vectors, classification probabilities, ability estimates, and standard errors is given in Table 7.5. Classification probabilities are in fact estimates of the person-class-indicator parameters ϕ_c, which range from zero to one, and which sum to one for $c = 1, \ldots, C$, and thus are interpretable as probabilities. Persons such as A and B, who respond correctly only to intuitive items, are classified solidly into class 1. Even persons such as C and D, who respond correctly to all of the intuitive items, and one or two of the easier concrete items, are still most likely to be in class 1, although person D has a small probability of being in class 2. Persons such as G and H, who respond correctly to all of the intuitive items and most of the concrete items, including some of the most difficult of these, are classified into class 2, although person G, who misses two of the concrete items, still has nearly a 1 in 5 probability of being in class 1.

Persons such as E and F are more difficult to classify. These persons answer some but not all of the concrete items, including some of the more difficult ones. In addition, person F misses one of the intuitive items. These persons have probabilities between .4 and .6 of being in either of the two classes; essentially, they do not fall clearly into either class.[1]

Table 7.5. Example person-response strings with proficiency and classification

		Group 1			Group 2		
Person	Responses	Probability	Ability	SE	Probability	Ability	SE
A	111000 000000	1.00	-3.60	.99	.00	-3.50	.76
B	111111 000000	1.00	-0.15	1.17	.00	-1.78	.76
C	111111 010000	1.00	1.23	1.17	.00	-1.19	.77
D	111111 110000	.93	2.56	.08	.08	- .59	.79
E	111111 010101	.58	3.76	1.04	.43	0.06	.82
F	110111 000110	.43	1.23	1.17	.57	-1.19	.77
G	111111 110011	.18	4.67	.64	.82	.79	.90
H	111111 111110	.03	5.24	.65	.97	1.74	1.06

The response vectors given in this table are typical of most of the response vectors in the data set; in particular, most of the persons classified as intuitive responded either like person B (responding correctly to all of the intuitive items and to none of the concrete items), or by missing only one or two of the intuitive items and still missing all of the concrete items—a total of 136 persons, or about 30%. Most of the persons classified as concrete responded like persons F and G, responding correctly to all of the intuitive and all or nearly all of the concrete items. Such persons accounted for 170, or about 37%. Relatively few persons (19 in all, or 4%) missed one or more of the intuitive items while responding correctly to some of the concrete items—persons such as F. The remaining persons either solved all of the intuitive, and one to three of the concrete items, correctly (99, or about 22% of the data set), or solved only a small number of the intuitive, and none of the concrete. items correctly (36, or about 7% of the data set).

7.4 Discussion

The use of the saltus model has allowed us to learn some interesting things about the example data set. For instance, it would seem that the saltus model is more suitable for use with these data than a latent-class model. Latent-class models are similar to mixture IRT models such as the saltus model and the mixed RM, in that they assume that the observed population is composed of latent subpopulations; however, in contrast to the latter, latent-class models include no quantitative person parameters. In latent-class models;

[1] In order to determine whether persons such as E and F can be fitted by the model appropriately, fit diagnostics (Molenaar, 1983) such as person-fit statistics may be used. von Davier & Molenaar (2003) present person-fit statistics that can be used with latent-class and discrete-mixture RMs.

class membership accounts for all explained variation between persons, and within-class variation is considered random variation. However, as seen in Table 7.3, person classes have relatively large standard deviations (between 1.7 and 2.6 logits), indicating that there is substantial and more importantly, systematic, within-class variability in these data.

Various types of fit analysis might prove useful. For example, it might be useful to develop a saltus-like model with variable item slopes, since models with equal slopes for all items are often too restrictive to fit well. In addition, it might be the case that models that included saltus parameters indexed by individual item (and perhaps by step in polytomous items), rather than simply associating saltus parameters with items as a whole, and estimating a single parameter across all items within an item class, might yield interesting differences by item and/or step.

One promising method for estimating parameters for such models is through their expression as generalized nonlinear mixed models. Statistical software packages are being developed that can estimate a wide variety of such models. An example of how this could be done using SAS was given by Fieuws et al. (2004); other software packages could also be used.

The saltus model has shown potential for aiding researchers, especially in the fields of cognitive science and Piagetian or neo-Piagetian theory, as do other extended models able to reflect the complexities of polytomous data and latent classes. For example, Commons and his colleagues have begun investigating the use of the saltus model for Commons's general stage theory of development (see, for example, Dawson et al., 1997). Other promising applications should follow as researchers in psychometrics continue their collaboration with educational and psychological researchers.