Effects of varying magnitude and patterns of local dependence in the unidimensional Rasch model

Ida Marais

and

David Andrich

Murdoch University, Western Australia

Mailing address

Ida Marais

Murdoch University

Murdoch 6150

Western Australia

Acknowledgements

The research reported in this paper was supported in part by an Australian Research Council grant with the Australian National Ministerial Council on Employment, Education, Training and Youth Affairs (MCEETYA) Performance Measurement and Reporting Task Force, UNESCO's International Institute for Educational Planning (IIEP), and the Australian Council for Educational Research (ACER) as Industry Partners*.

*Report No. 7 ARC Linkage Grant LP0454080: Maintaining Invariant Scales in State, National and International Level Assessments. D Andrich and G Luo Chief Investigators, Murdoch University

Effects of varying magnitude and patterns of local dependence in the unidimensional Rasch model

Abstract

By adding items with responses identical to a selected item, Smith (2005) investigated the effect that the local dependence has on person and item parameter estimates in the dichotomous Rasch model. By varying the magnitude of local dependence among selected items, rather than their having perfect dependence, this paper provides additional insights into the effects that local dependence has on the same estimates in the same model. Two sets of simulations are reported. In the first set, responses to all items except the first were dependent on either the first item or on the immediately preceding item; in the second set, subsets of items were formed first, and then within each of these subsets, responses to all items in a subset except the first were dependent on either the first item or on the immediately preceding item. The effects of dependence were noticeable in all the statistics reported. In particular, the fit statistics and the parameter estimates showed increasing discrepancies from their theoretical values as a function of the magnitude of the dependence. In some cases, however, two related statistics gave the impression of improvement as a function of increased dependency; first the standard deviation of person estimates showed an increase, and second the index analogous to traditional reliability showed relative improvement. In addition to the estimates and depending on the structure and magnitude of the dependence, the person distribution was affected systematically, ranging from becoming skewed to becoming bimodal. The effects on the distribution help explain some of the effects on the statistics reported. In the case of the second set of simulations in which the dependence is within subsets of items, it is possible to take account of the local dependence. This is done by summing the responses of the items within each subset to form a polytomous item and then to analyse the data in terms of a smaller number of polytomous items. This way of accounting for dependence, in which the maximum score for the test as a whole remains the same, gives a more accurate value of the reliability and a more realistic distribution of the person estimates than when the dependence within subsets of items is not taken into account.

Effects of varying magnitude and patterns of local dependence in the unidimensional Rasch model

1. Introduction

The unidimensional Rasch model for more than two ordered categories can be expressed in the form

$$\Pr\{X_{ni} = x\} = [\exp(x\ (\beta_n - \delta_i) - \sum_{k=1}^{x} \tau_{ki})] / \sum_{x=0}^{m_i} [\exp(x\ (\beta_n - \delta_i) - \sum_{k=1}^{x} \tau_{ki})] \quad (1)$$

where $x \in \{0,1,2...m_i\}$ is the integer response variable for person n with ability β_n responding to item i with difficulty δ_i, and $\tau_{1i}, \tau_{2i,\ldots}\ \tau_{mi}$ are thresholds between $m_i + 1$ ordered categories where m_i is the maximum score of item i (Andrich, 2005). This implies a single dimension with values of β, δ and τ located additively on the same scale.

The special case of Eq. (1) for dichotomous responses is

$$\Pr\{X_{ni} = x\} = [\exp(x(\beta_n - \delta_i))] / [1 + \exp(\beta_n - \delta_i)] \quad (2)$$

where $x \in \{0,1\}$ and there is only one threshold, δ_i.

The model implies independence of responses in the sense that

$$\Pr\{((x_{ni}))\} = \prod_n \prod_i \Pr\{x_{ni}\} \quad (3)$$

where $((x_{ni}))$ denotes the matrix of responses $X_{ni} = x$, n = 1…N, I = 1…I. A specific expression for the independence of the responses between two items i and j is that

$$\Pr\{X_{nj} = x_j \mid X_{ni} = x_i\} = \frac{\Pr\{x_i \cap x_j\}}{\Pr\{x_i\}} = \Pr\{x_j\}. \quad (4)$$

That is, $\Pr\{x_i \cap x_j\} = \Pr\{x_i\}\Pr\{x_j\}$, which is a special case of Eq. (3).

We use Eq. (4) later in the paper as a basis to construct responses which are not independent. This independence is generally termed "local independence" (Andrich, 1991).

No data fit any model perfectly and one concern regarding the violation of the Rasch model is the independence defined in Eqs. (3) and (4). This concern includes the effect that dependence has on the person and item parameter estimates, how to detect local dependencies, and when present, how best to account for them.

Smith (2005) addressed the effect that local dependence has on Rasch person and item parameter estimates and showed its effects in a very specific simulation study. He simulated dependence by adding redundant items to an existing item set. By a redundant item is meant an item whose responses are an exact copy of the responses to an existing item. Because the vector of responses $(x_{nj}) = (x_{ni})$ when item j is dependent on item i in this way, it follows that

$$\Pr\{X_{nj} = x_j \mid X_{ni} = x_i\} = \Pr\{X_{ni} = x_i\}, \tag{5}$$

which violates Eq. (4).

To evaluate the effect of this kind of dependence, Smith studied data sets varying in numbers of persons (50, 150, 250, 500, 1000 and 2000) responding to varying numbers of items (10, 30 and 50). He reported the effect on the standard deviations of person and item estimates and item and person reliabilities. He also reported the root mean squared differences and mean signed differences between a baseline condition with no dependence and the dependence conditions, as well as the correlation between person estimates in those conditions. In addition, he reported the percentages of person estimates in dependence conditions shifting by more than 0.5 logits from the baseline

estimates with independence. Smith concluded that the amount of dependence of the kind he generated needs to be considerable before person and item estimates were seriously affected.

However, adding redundant items in the way Smith did, while instructive, is an extreme example of local dependence and unlikely to be seen in practice. He made suggestions for further simulation work to which his results can be compared. This paper addresses some of these suggestions.

Adding more redundant items increased the *magnitude of dependence in the whole data* set in the Smith study but the *magnitude of dependence between an item and the redundant items* remained the same. In contrast, in the study reported in this paper the magnitude of dependence between pairs of items is systematically varied. Such non-extreme dependence is more likely to reflect practical situations.

In practice, local dependence is found under different circumstances. For example, where judges make judgements on a set of criteria and a halo effect operates, local independence among criteria is readily violated (Heldsinger & Humphry, 2006). It is also violated when a correct answer on a question gives a clue to the answer to one or more subsequent questions. Finally, it can be violated when questions have some feature in common, such as the case when questions arise from the reading of a single stimulus. Different *patterns* of local dependence are found in these different circumstances. In this study, not only are different magnitudes of dependence between items simulated, but also different patterns of dependence among items.

Because a redundant item is an exact copy of another item, it also has the same *difficulty* as the item copied. In the Smith study all the redundant items were redundant with respect to just one item, that is, all the dependent items were dependent on the same item and therefore their responses were also identical to each other. Therefore, as more redundant items were added, more items of the same difficulty were added. Not only did the total number of items then change, but the distribution of the item locations also

changed, specifically it reduced their standard deviation. In the study reported in this paper, a dependent item need not be of the same difficulty as the item it is dependent upon, and another item does not need to be added to increase dependency. The only constraint is that an item's responses can only be dependent on those of a previous item. The algorithm for generating such dependence is described in the next section.

To provide an overview of the contrasting designs of Smith's study and the study of this paper, Table 1 summarises their respective features.

Table 1.Comparison of research designs in the present and Smith(2005) studies

Smith(2005)	**This study**
1) Number of persons varied	*1) Number of persons* remained the same (1000)
2) Total number of items varied	*2) Total number of items* remained the same (30)
3) Magnitude of dependence	*3) Magnitude of dependence*
between dependent items same	*between dependent items* varied
4) Structure/Pattern of dependence same	*4) Structure/Pattern of dependence* varied

In order to be comparable to the Smith study the statistics for the same effects are reported in this paper. As in that study, all items are dichotomous.

In addition to addressing the effects that dependence has on Rasch person and item parameter estimates, the effects of a possible way of dealing with the dependence are also investigated. This involves combining dichotomous dependent items in a data set into a polytomous item (Andrich, 1985; Wang, Bradlow, & Wainer, 2002; Wilson, & Adams, 1995; Zenisky, Hambleton & Sireci, 2002). A series of simulation studies, parallel to the first set, are carried out to investigate this approach to studying the effects of local dependence.

2. Method

Two data sets which were structurally different were simulated. Section 2.1 describes the simulation algorithm. Section 2.2 describes Simulation set 1 which involved the items

having various degrees of dependence but no structure among subsets of items. Section 2.3 describes Simulation set 2 which involved a structure among subsets of items. In section 2.4 the statistics that are reported in this study are described.

2.1 A simulation algorithm for local dependency

Local dependence is simulated by making a person's response on an item be a function of the person's response to a previous item. Specifically, local dependence is simulated by making the probability of a person's correct response on an item increase as a function of the correct response, and decrease as a function of the incorrect response, on a previous item on which it depends. How much the probability increased or decreases can be determined by a constant, more specifically, by adding or subtracting a constant from the difficulty of the dependent item.

Equation 6 formalises this construction for item j dependent on item i:

$$\Pr\{X_{nj}=1 \mid X_{ni}=1\}=[\exp(\beta_n-(\delta_j-d))]/[1+\exp(\beta_n-(\delta_j-d))] \qquad (6)$$

and

$$\Pr\{X_{nj}=1 \mid X_{ni}=0\}=[\exp(\beta_n-(\delta_j+d))]/[1+\exp(\beta_n-(\delta_j+d))],$$

which reduces to

$$\Pr\{X_{nj}=1 \mid X_{ni}=x_i\}=[\exp(\beta_n-\delta_j-(1-2x_i)d)]/[1+\exp(\beta_n-\delta_j-(1-2x_i)d)]$$

where d is the constant used to increase or decrease the magnitude of dependence.
The general equation which includes responses $X_{nj}=1$ as well as $X_{nj}=0$ takes the form

$$\Pr\{X_{nj}=x_j \mid X_{ni}=x_i\}=$$

$$[\exp(x_j(\beta_n-\delta_j-(1-2x_i)d))]/[1+\exp(x_j(\beta_n-\delta_j-(1-2x_i)d))] \qquad (7)$$

It is evident from Eq. (7) that the response of item j depends on the response of item i and therefore violates Eq. (4). Specifically, if a person's response to item i was $x_{ni} = 1,$ then the dependent item j's difficulty is changed to $\delta_j - d$ for that person. Hence the dependent item has been made easier, thus also increasing the probability of a response of $x_{nj} = 1,$ of the person to item j. Conversely, if a person's response to item i was 0 $x_{ni} = 0,$ then the dependent item j's difficulty is changed to $\delta_j + d$ for that person, thus also increasing the probability of a response $x_{nj} = 0,$ of that person to that item.

It is readily shown that the probabilities according to Eq. (7) constitute a complete response space. Thus from the basic expression of conditional probability (Ross, 1976) shown in the first part of Eq. (4),

$$\Pr\{X_{nj} = x_j \mid X_{ni} = x_i\} = \frac{\Pr\{x_i \cap x_j\}}{\Pr\{x_i\}},$$

giving

$$\Pr\{x_i \cap x_j\} = \Pr\{X_{nj} = x_j \mid X_{ni} = x_i\}\Pr\{x_i\}. \tag{8}$$

Eq. (7) is used to construct the elements of Eq. (8). Table 2 shows all the possible outcomes and their probabilities. The sum of these probabilities in Table 2 is 1, as required of elements of an outcome space.

Table 2. Joint probabilities of responses of pairs of items in the presence of dependence

Response pattern Item i	Item j	Joint probability
0	0	$\frac{1}{1+e^{\beta_n-\delta_i}} \frac{1}{1+e^{\beta_n-\delta_j-d}}$
0	1	$\frac{1}{1+e^{\beta_n-\delta_i}} \frac{e^{\beta_n-\delta_j-d}}{1+e^{\beta_n-\delta_j-d}}$
1	0	$\frac{e^{\beta_n-\delta_i}}{1+e^{\beta_n-\delta_i}} \frac{1}{1+e^{\beta_n-\delta_j+d}}$
1	1	$\frac{e^{\beta_n-\delta_i}}{1+e^{\beta_n-\delta_i}} \frac{e^{\beta_n-\delta_j+d}}{1+e^{\beta_n-\delta_j+d}}$
		Sum of probabilities = 1

In both data sets, data were generated for 1000 people and 30 dichotomous items. The choice of item difficulties and the choice of the distribution of person abilities were meant to make the targeting of persons to items ideal. The distribution of person abilities was $N(0,2^2)$. The distribution of item difficulties was uniform, ranging between –3.5 and 3.5.

The *magnitude of dependence* was varied by setting d in Eq. (7) to 0, 1, 2, 3 and 4. The case of d = 0, no dependence, provided the frame of reference for interpretation of the analyses and its values were taken as theoretical values from which the estimates in the other simulations might deviate.

2.2 Simulation set 1

The following *patterns of dependence* were simulated in Simulation set 1:

- Pattern 1: All items were dependent on the first item in the data set where the first item is the easiest item.
- Pattern 2: All items were dependent on the first item in the data set where the first item was of average difficulty.

- Pattern 3: All items were dependent on the item preceding them where the successive items were of increasing difficulty, as might arise from an intelligence or achievement test.

Table 3 summarises the three patterns of dependence and the item difficulties for each pattern. Because no items precede it, item 1 is not dependent on any item. For pattern 2, and in order to make the first item of average difficulty whilst keeping the same distribution of difficulties for the item set, the difficulties for the first item and the middle item (item 16) were exchanged.

Table 3. Simulation set 1: Patterns of dependence. δ is item difficulty, DItem is the item dependent upon.

	Pattern 1		Pattern 2		Pattern 3	
Item	**δ**	**DItem**	**δ**	**DItem**	**δ**	**DItem**
1	-3.50	-	0.12	-	-3.50	-
2	-3.26	1	-3.26	1	-3.26	1
3	-3.02	1	-3.02	1	-3.02	2
4	-2.78	1	-2.78	1	-2.78	3
5	-2.53	1	-2.53	1	-2.53	4
6	-2.29	1	-2.29	1	-2.29	5
7	-2.05	1	-2.05	1	-2.05	6
8	-1.81	1	-1.81	1	-1.81	7
9	-1.57	1	-1.57	1	-1.57	8
10	-1.33	1	-1.33	1	-1.33	9
11	-1.09	1	-1.09	1	-1.09	10
12	-0.84	1	-0.84	1	-0.84	11
13	-0.60	1	-0.60	1	-0.60	12
14	-0.36	1	-0.36	1	-0.36	13
15	-0.12	1	-0.12	1	-0.12	14
16	0.12	1	-3.50	1	0.12	15
17	0.36	1	0.36	1	0.36	16
18	0.60	1	0.60	1	0.60	17
19	0.84	1	0.84	1	0.84	18
20	1.09	1	1.09	1	1.09	19
21	1.33	1	1.33	1	1.33	20
22	1.57	1	1.57	1	1.57	21
23	1.81	1	1.81	1	1.81	22
24	2.05	1	2.05	1	2.05	23
25	2.29	1	2.29	1	2.29	24
26	2.53	1	2.53	1	2.53	25
27	2.78	1	2.78	1	2.78	26
28	3.02	1	3.02	1	3.02	27
29	3.26	1	3.26	1	3.26	28
30	3.50	1	3.50	1	3.50	29
SD	2.13		2.13		2.13	

2.3 Simulation set 2

In practice dependence between items is often found within a *subset* of items in the whole set, for example within a subset of items that arise from the reading of a single stimulus. There can be more than one such subset of dependent items within a whole set. Simulation set 2 attempts to capture this structure. The 30 items were divided into 6 subsets of 5 items each. The items in each subset had difficulties distributed from –3.5 to 3.5.

Three *patterns of dependence*, analogous to those in Simulation set 1, were generated:

- Pattern 1: All items were dependent on the first item in the subset where the first item is the easiest item.
- Pattern 2: All items were dependent on the first item in the subset where the first item is of average difficulty.
- Pattern 3: All items were dependent on the item preceding them in the subset where the successive items were of increasing difficulty.

Table 4 summarises the three patterns of dependence and the item difficulties for each pattern. The last column in the Table shows how the items are divided into 6 subsets. In order to make the first item in the *subset* of average difficulty in pattern 2, the difficulties for the *first item in a subset* and the *middle item in the subset* were exchanged as in Simulation set 1.

Table 4. Simulation set 2: Patterns of dependence. δ is item difficulty, DItem is the item dependent upon.

	Pattern 1		Pattern 2		Pattern 3		
Item	**δ**	**DItem**	**δ**	**DItem**	**δ**	**DItem**	
1	-3.50	-	0.00	-	-3.50	-	**Subset 1**
2	-1.75	1	-1.75	1	-1.75	1	
3	0.00	1	-3.50	1	0.00	2	
4	1.75	1	1.75	1	1.75	3	
5	3.50	1	3.50	1	3.50	4	
6	-3.50	-	0.00	-	-3.50	-	**Subset 2**
7	-1.75	1	-1.75	1	-1.75	6	
8	0.00	1	-3.50	1	0.00	7	
9	1.75	1	1.75	1	1.75	8	
10	3.50	1	3.50	1	3.50	9	
11	-3.50	-	0.00	-	-3.50	-	**Subset 3**
12	-1.75	1	-1.75	1	-1.75	11	
13	0.00	1	-3.50	1	0.00	12	
14	1.75	1	1.75	1	1.75	13	
15	3.50	1	3.50	1	3.50	14	
16	-3.50	-	0.00	-	-3.50	-	**Subset 4**
17	-1.75	1	-1.75	1	-1.75	16	
18	0.00	1	-3.50	1	0.00	17	
19	1.75	1	1.75	1	1.75	18	
20	3.50	1	3.50	1	3.50	19	
21	-3.50	-	0.00	-	-3.50	-	**Subset 5**
22	-1.75	1	-1.75	1	-1.75	21	
23	0.00	1	-3.50	1	0.00	22	
24	1.75	1	1.75	1	1.75	23	
25	3.50	1	3.50	1	3.50	24	
26	-3.50	-	0.00	-	-3.50	-	**Subset 6**
27	-1.75	1	-1.75	1	-1.75	26	
28	0.00	1	-3.50	1	0.00	27	
29	1.75	1	1.75	1	1.75	28	
30	3.50	1	3.50	1	3.50	29	

2.4 Analyses

The generated data sets were analysed with the RUMM2020 software (Andrich, Sheridan, & Luo, 1997 - 2005). The following statistics from the analyses are reported in the Results section:

Person distribution: The effects of dependence on the *mean* and *standard deviation* of the person distribution are reported.

Person separation index (PSI): The effects of dependence on person reliability, specifically the PSI, are reported. The PSI is based on the traditional true score reliability formula (Gulliksen, 1950)

$$r_{xx} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_e^2} = \frac{\sigma_x^2 - \sigma_e^2}{\sigma_x^2},$$

where σ_x^2 is the variance of the observed total scores, σ_τ^2 is the variance of the true scores, σ_e^2 is the error variance of each measurement and $\sigma_x^2 = \sigma_\tau^2 + \sigma_e^2$. Thus the reliability is a function of both the variance of the observed estimates and the error of measurement. A similar index is constructed with Rasch measurement, termed the person separation,

$$r_{\beta\beta} = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_e^2} = \frac{\sigma_{\hat{\beta}}^2 - \sigma_{\hat{e}}^2}{\sigma_{\hat{\beta}}^2}$$

with $\hat{\sigma}_{\hat{\beta}}^2$ is the estimated variance of the locations of the persons, and $\hat{\sigma}_{\hat{e}}^2$ is the average squared standard error of measurement for each person. Again, it is a function of both the variance of the estimates of the person locations and the error of measurement variance, and is a relevant statistic to consider in relation to specific violations of the

model. In general and for complete data, the values of this index and the traditional index, Cronbach's alpha, are virtually identical (Andrich, 1982).

The range and Standard Deviation of the scale: The effects on the range and standard deviation of the scale are reported. The range is the difference between the person ability estimates for a total score of 1 and $\sum_i m_i - 1$ where $\sum_i m_i$ is the maximum score.

In the Rasch models, the total score is a sufficient statistic for the person ability. Thus for each total score on a set of items, and irrespective of the pattern of correct responses across the items, there is a single ability estimate. Each total score is transformed non linearly to give the ability estimate. Thus irrespective of taking or not taking into account the dependence, the analysis will have the same raw score range with distribution of raw scores. One concern is the range of values of the transformed scores and the degree to which violations of the model affect these values, and therefore the scale.

Deviations of person estimates: The *correlations* between person estimates in the baseline (no dependence) condition and the conditions with dependence are reported as well as the *root mean squared difference (RMSD)* and *mean signed difference (MSD)* between them (Smith, 2005).

$$RMSD = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(b_n - e_n)^2} \tag{9}$$

and

$$MSD = \frac{1}{N}\sum_{n=1}^{N}(b_n - e_n) \tag{10}$$

where b_n refers to a person n's estimate when $d = 0$, e_n refers to person n's estimate when $d > 0$ and N refers to the total number of persons (1000 in all simulations in this

study). Also reported is the *percentage of person estimates with a logit shift* of more than 0.5.

Item distribution: The effect of dependence on the *standard deviation* of the item distribution is reported.

Item residual correlations: The standardized residual for a person interacting with an item is the difference between the actual response and the expected value, divided by the square root of the variance. The standardized residual for person n interacting with item i is given by

$$Z_{ni} = \frac{x_{ni} - E[X_{ni}]}{\sqrt{V[X_{ni}]}}. \tag{11}$$

If the data and model accord with each other, then the standardized person-item residuals should show no systematic relationships. A high correlation of residuals between pairs of items can indicate that they are more dependent in their responses than can be accounted for by the locations of the persons and the relative locations of the item estimates according to the model assuming independence. Specifically, if one item has a correct response then the other item is more likely to have a correct response than can be accounted for by the locations of the persons and items. As will be seen, when there is no dependence (d = 0) the correlations of residuals between most pairs of items are less than 0.1. Therefore, for the simulated data in this study the percent of correlations between standardized residuals of pairs of items greater than 0.1 are reported. As the magnitude of dependence increased one would expect the percentage of items with inter item residual correlations greater than 0.1 to increase.

Item fit residual: The Item fit residual is a statistic that provides information on the fit of the data to the model from the perspective of the items. For each item, this statistic is based on the standardised residuals of the responses of all persons to the item. To obtain

the overall index for an item, the residuals are calculated as in Eq. (11), then squared and summed over the persons to give

$$Y_i^2 = \sum_{n=1}^{N} z_{ni}^2,$$

and then transformed to be similar to a standard normal deviate

$$T_i = \frac{Y_i^2 - E[Y_i^2]}{\sqrt{V[Y_i^2]}}. \tag{12}$$

Because this distribution is not symmetrical, a logarithmic transformation is made to T_i to make the distribution more symmetrical. This is done by forming the mean square ratio

$$Y_i^2 / f_i$$

where f_i is the degrees of freedom, and then taking its natural logarithm. Then T_i becomes

$$T_i = \frac{f_i(\log Y_i^2 - \log f_i)}{\sqrt{V[Y_i^2 / f_i]}}$$

which is a more symmetrical distribution with $E[T_i] = 0$ and $V[T_i] = 1$. This index, as reported in RUMM2020, is reported in the results that follow.

Item fit residuals that lie within the range from approximately -2.5 to 2.5 would be considered as fitting the model based on this one criterion. If an item is over-discriminating the item fit residual will be very negative (<-2.5), and if it is under-

discriminating, it will be very positive (>2.5). When there is dependence between items as described in this paper, items typically over-discriminate. Therefore, as dependence is increased in the simulated data in this study, one would expect more items to over-discriminate, and hence have very negative item fit residuals (<-2.5). The percentage of items with an item fit residual more negative than the most negative item fit residual in the no dependence condition (*d*=0) for a pattern is reported. This is calculated by first identifying, for the no dependence condition of each pattern, the most negative item fit residual. Then, for each of the conditions where *d*>0, the percentage of items with an item fit residual more negative than this value is reported.

3. Results

3.1 Simulation set 1

Table 5 shows the results for Simulation set 1 and Figure 1 shows the distributions of person and item estimates graphically as produced in RUMM2020.

Person distribution:

Data were simulated with a person mean of 0 and a SD of 2. When d = 0 (no dependence) the *person mean* was 0 or very close to 0 and the *person SD* was 2 or very close to 2 for all the patterns. However, as the magnitude of dependence (d) increased the mean increased for patterns 1 and 3 but not for pattern 2. The person SD also increased as the magnitude of dependence (d) increased for all patterns.

These results can be understood from the distributions of person locations shown in Figure 1. In the case of patterns 1 and 3, where items are dependent on an easy item, there is a high probability of a correct response on that easy item. It follows that the dependent items will have a high probability of a correct response as well, hence persons will get more items correct resulting in higher ability estimates and an increased mean. Figure 1 shows how the distributions became skewed for patterns 1 and 3 with increased dependence with the result of a change of mean.

In contrast, in pattern 2, where items were dependent upon an item of average difficulty, the distribution became bimodal with increased dependence. Although the distribution changed, the mean remained relatively constant. This is explained from the effect that if a person gets the first item correct, then as a result of dependence the person will tend to get all items correct; likewise if a person gets the first item incorrect, the person will tend to get all items incorrect.

Out of the three patterns, pattern 2 is the pattern most like the one simulated by Smith (2005), in that items were dependent on the *same* item, the item being of *average difficulty*. Smith did not report the person mean but did report the person SD, which increased as redundant items were added. From Table 1 it is clear that the person SD in this study also increased as the magnitude of dependence increased. Smith also noted that the distribution was bimodal when redundant items were added.

PSI:

In Smith (2005) the PSI was 0.84 for N=1000 and 30 items in the no dependence condition. The PSI was 0.93 in the no dependence condition in this study. The different PSI's reflect the different person and item standard deviations in the two studies. In the Smith study the person SD was 1.0 and items ranged from -1 to +1. In this study data were simulated with a person SD of 2 and items ranged from -3.5 to +3.5 with a SD of 2 to 2.5. Because the range of items targeted the range of persons more completely in this study, it resulted in a higher PSI. The PSI increased as d increased for all patterns.

Range and SD of the scale:

The *range* of the scale and *SD* of the scale increased as d increased for all patterns.

Deviations of person estimates:

The *Root mean Squared Difference (RMSD)* increased as d increased for all patterns. The RMSD values reported by Smith were smaller than the values reported here. The

RMSD values for pattern 2, the pattern most like the one simulated by Smith, were the smallest for all the patterns.

The *Mean Signed Difference (MSD)* became an increasingly higher negative number as d increased for patterns 1 and 3 and not for pattern 2. The MSD values for pattern 2 remained relatively constant, as in the Smith study. Once again, this can be understood from the distributions of person locations shown in Figure 1. For patterns 1 and 3 the person estimates were likely to be greater than the baseline estimates when dependence is present. Since the MSD is computed by subtracting this greater estimate from the baseline estimate for each person it will result in a negative number.

The c*orrelation between person estimates when d=0 and when d>0* decreased as d increased for all patterns and was on the whole smaller than in Smith (2005). The *percent of person estimates with a logit shift greater then 0.5* increased as d increased for all patterns. These values were generally greater than in Smith (2005).

Item distribution:

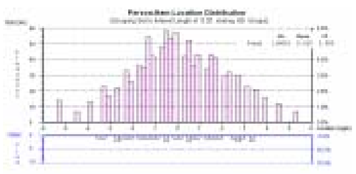
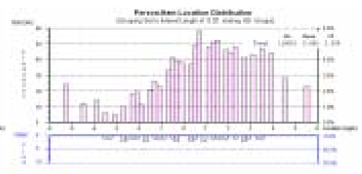
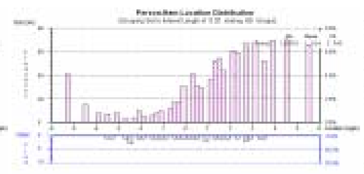
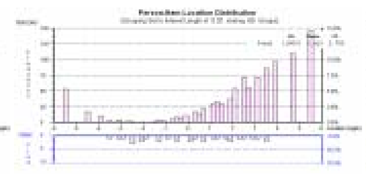
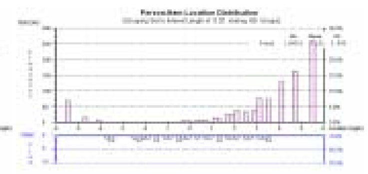
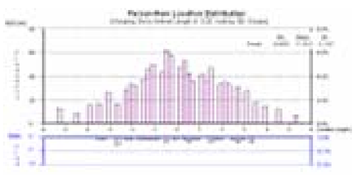
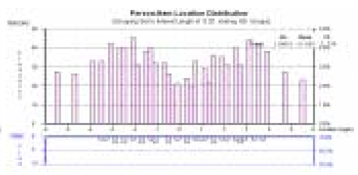
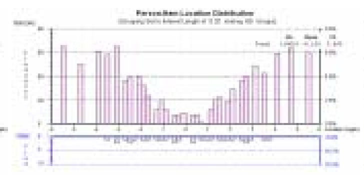
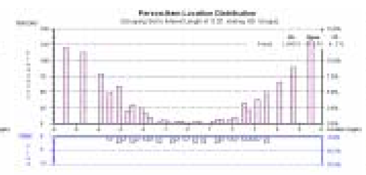
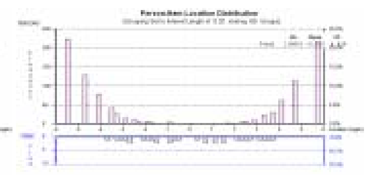
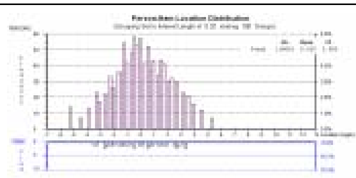
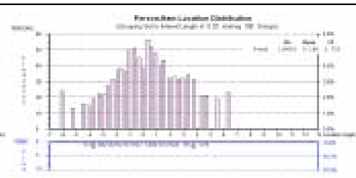
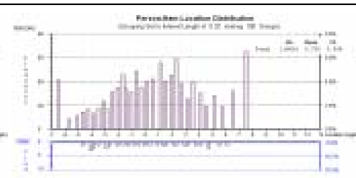
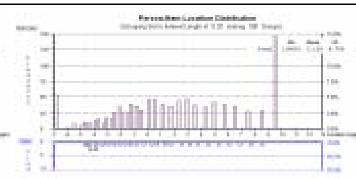
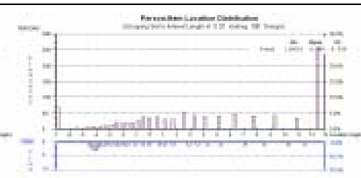
The *item SD* increased as d increased for all patterns. Because the redundant items were all of the same difficulty in Smith (2005), the item SD reduced with each redundant item added. With the simulation algorithm used in this paper dependent items were not of the same difficulty. As the results in Table 1 show, the item SD increased as the magnitude of dependence increased. This is also evident in Figure 1, especially for pattern 3.

The *percent of items with item residual correlations greater than 0.1* increased as d increased for all patterns. The *percent of items with fit residuals more negative than baseline* increased as d increased for all patterns. For pattern 3 the value first increased then decreased.

Table 5. Simulation set 1: Person mean, SD and PSI for all patterns as a function of d. Root mean squared difference (RMSD), Mean signed difference (MSD), Correlation (r) and Percent of person estimates with shift more than 0.5 logits from baseline (% log shift) as a function of d. Item SD, scale range, scale SD, percent of items with residual correlations > 0.1(% item residual r > 0.1) and percent of items with fit residuals more negative than baseline (% < fit residual) for all patterns as a function of d.

Pattern			Magnitude(d) 0	1	2	3	4
1 (Item 1 easiest)	Persons	**Person mean**	0.02	0.84	1.61	2.41	3.13
		Person SD	2.00	2.27	2.54	2.78	3.00
		PSI	0.93	0.94	0.95	0.95	0.96
		Scale range	9.02	9.10	9.10	9.19	9.40
		Scale SD	2.42	2.44	2.43	2.48	2.53
		RMSD	0.00	1.10	2.02	2.93	3.75
		MSD	0.00	-0.82	-1.60	-2.40	-3.12
		r	1.00	0.95	0.88	0.80	0.72
		% log shift	0	84	97	98	98
	Items	**Item SD**	2.08	2.11	2.10	2.16	2.23
		% item residual r>0.1	1	1	2	4	5
		% < fit residual	0	3	13	13	20
2 (Item 1 average)	Persons	**Person mean**	0.01	-0.04	-0.12	-0.18	-0.23
		Person SD	2.02	2.74	3.49	4.17	4.82
		PSI	0.93	0.95	0.97	0.98	0.98
		Scale range	9.04	9.23	9.25	9.28	9.48
		Scale SD	2.43	2.49	2.50	2.51	2.59
		RMSD	0.00	1.11	2.04	2.84	3.61
		MSD	0.00	0.05	0.13	0.19	0.24
		r	1.00	0.94	0.86	0.8	0.74
		% log shift	0	84	70	98	98
	Items	**Item SD**	2.09	2.17	2.19	2.20	2.29
		% item residual r>0.1	1	1	1	3	7
		% < fit residual	0	10	13	17	20
3 (Preceding item)	Persons	**Person mean**	0.02	0.20	0.73	2.12	4.38
		Person SD	2.00	2.73	3.54	4.76	6.34
		PSI	0.93	0.95	0.97	0.98	0.99
		Scale range	9.02	10.65	12.16	14.46	16.99
		Scale SD	2.42	2.97	3.47	4.13	4.58
		RMSD	0.00	0.97	2.02	3.84	6.52
		MSD	0.00	-0.18	-0.71	-2.11	-4.37
		r	1.00	0.97	0.91	0.86	0.81
		% log shift	0	60	81	88	93
	Items	**Item SD**	2.08	2.76	3.35	4.10	4.67
		% item residual r>0.1	1	7	22	34	35
		% < fit residual	0	3	17	13	0

Figure 1. Simulation set 1: Person and item distributions for all patterns as a function of d

	Magnitude(d)				
Pattern	**0**	**1**	**2**	**3**	**4**
1 (Item 1 easiest)					
2 (Item 1 average)					
3 (Preceding item)					

3.2 Simulation set 2

In Simulation set 1 all items were analysed as dichotomous items. In Simulation set 2 dependence was simulated within subsets of items. The items were then analysed as dichotomous first, followed by a second analysis in which the items belonging to a subset were combined into a polytomous item. As indicated earlier, this takes into account the dependence within a subset (Andrich, 1985).

3.2.1 First analysis: dichotomous items

Table 6 shows the results for the first analysis of Simulation set 2 and Figure 2 shows the distributions of person and item estimates graphically.

Person distribution and PSI:
As the magnitude of dependence increased the *person mean* increased for patterns 1 and 3 but not for pattern 2. The *person SD* also increased, then decreased as the magnitude of dependence increased for all patterns. This decrease is due to a ceiling effect, when high dependence results in more persons getting the maximum score. For those patterns dependent on easy items (patterns1 and 3), Figure 2 shows that the means increased and that the distributions became unimodal, as in patterns 1 and 3 of Simulation set 1. For pattern 2, where items were dependent upon an item of average difficulty, the distributions became bimodal again as in pattern 2 of Simulation set 1. Of particular interest is that the PSI increased as dependence increased for pattern 2 but stayed roughly similar for patterns 1 and 3. The initial high value of the PSI meant that its increase was constrained.

Range and SD of the scale:
The ceiling effects also appeared in the values of the *range* and *SD* of the scale . For example the range and SD decreased as d increased for patterns 1 and 2. For pattern 3 the range and SD first increased then decreased. This effect can be understood by more

scores reaching a ceiling of the maximum and minimum scores as a result of the dependence.

Deviations of person estimates:

The *Root mean Squared Difference (RMSD)* increased as d increased for all patterns. The *Mean Signed Difference (MSD)* became more negative as d increased for patterns 1 and 3. The c*orrelation between person estimates when d=0 and when d>0* decreased as d increased for all patterns. The *percent of person estimates with a logit shift greater than 0.5* increased as d increased for all patterns.

Item distribution:

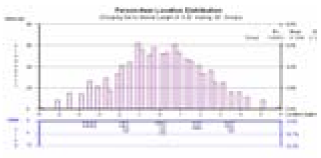
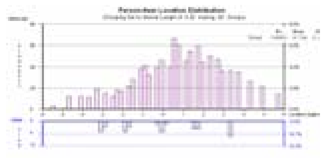
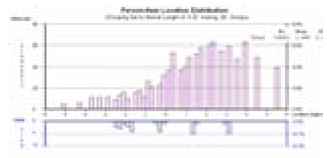
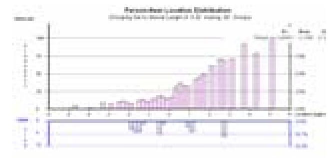
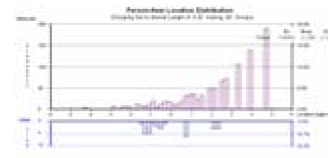
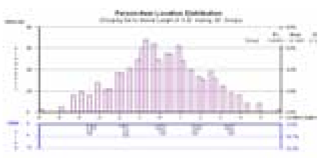
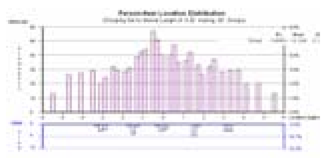
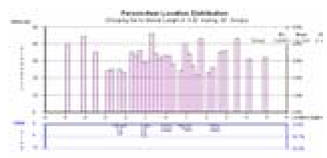
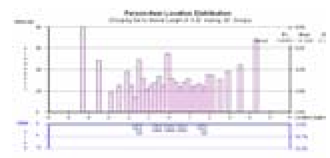
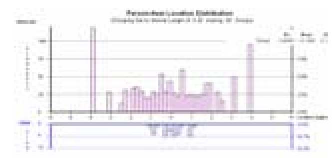
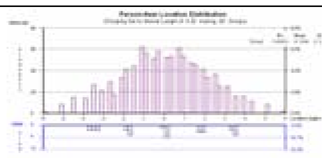
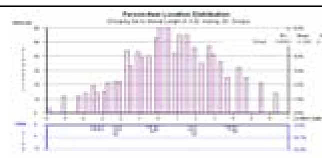
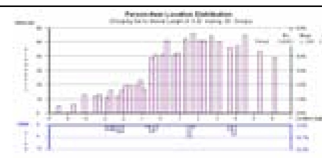
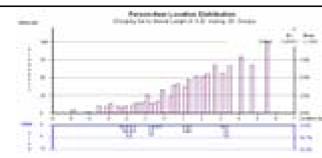
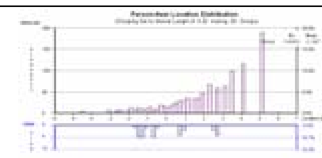
The *item SD* decreased as d increased for patterns 1 and 2. Figure 2 shows this effect. For pattern 3 the SD first increased then decreased as d increased.

The *percent of items with item residual correlations greater than 0.1* increased as d increased for all patterns. The *percent of items with fit residuals more negative than baseline* increased as d increased for all patterns.

Table 6. Simulation set 2 – first analysis: Person mean, SD and PSI for all patterns as a function of d. Root mean squared difference (RMSD), Mean signed difference (MSD), Correlation (r) and Percent of person estimates with shift more than 0.5 logits from baseline (% log shift) as a function of d. Item SD, scale range, scale SD, percent of items with residual correlations > 0.1(% item residual r > 0.1) and percent of items with fit residuals more negative than baseline (% < fit residual) for all patterns as a function of d.

			Magnitude(d)				
Pattern			**0**	**1**	**2**	**3**	**4**
1 (Item 1 easiest)	**Persons**	**Person mean**	0.00	0.71	1.40	1.94	2.33
		Person SD	2.06	2.18	2.17	2.09	1.95
		PSI	0.93	0.93	0.93	0.92	0.91
		Scale range	10.06	9.52	8.84	8.12	7.41
		Scale SD	2.78	2.63	2.39	2.15	1.91
		RMSD	0.00	0.88	1.55	2.10	2.50
		MSD	0.00	-0.71	-1.40	-1.93	-2.33
		r	1.00	0.97	0.95	0.92	0.90
		% log shift	0	64	91	96	97
	Items	**Item SD**	2.53	2.33	2.04	1.70	1.34
		% item residual r>0.1	1	3	9	16	32
		% < fit residual	0	10	23	40	53
2 (Item 1 average)	**Persons**	**Person mean**	0.00	-0.02	-0.02	-0.01	-0.01
		Person SD	2.03	2.39	2.43	2.39	2.33
		PSI	0.93	0.94	0.95	0.96	0.96
		Scale range	9.96	9.39	8.07	6.95	6.31
		Scale SD	2.75	2.55	2.12	1.77	1.58
		RMSD	0.00	0.59	0.76	0.83	0.87
		MSD	0.00	0.01	0.01	0.01	0.00
		r	1.00	0.98	0.96	0.94	0.93
		% log shift	0	35	48	50	51
	Items	**Item SD**	2.50	2.25	1.65	1.08	0.63
		% item residual r>0.1	1	9	30	57	83
		% < fit residual	0	7	13	27	20
3 (Preceding item)	**Persons**	**Person mean**	0.00	0.58	1.36	1.96	2.37
		Person SD	2.06	2.45	2.52	2.37	2.15
		PSI	0.93	0.94	0.94	0.94	0.93
		Scale range	10.06	10.41	9.83	8.84	7.88
		Scale SD	2.78	2.92	2.73	2.38	2.06
		RMSD	0.00	0.88	1.60	2.16	2.55
		MSD	0.00	-0.58	-1.35	-1.96	-2.36
		r	1.00	0.97	0.95	0.92	0.90
		% log shift	0	51	84	96	98
	Items	**Item SD**	2.53	2.70	2.46	2.02	1.57
		% item residual r>0.1	1	8	16	23	38
		% < fit residual	0	7	20	33	43

Figure 2. Simulation set 2 – initial analysis: Person and item distributions for all patterns as a function of d

	Magnitude(d)				
Pattern	**0**	**1**	**2**	**3**	**4**
1 (First subset Item easiest)					
2 (First subset Item average)					
3 (Preceding item in subset)					

3.2.2 Subtest analysis: polytomous items

In the subtest analysis the dependent dichotomous items in a subset were analysed by combining items within a subset into a polytomous item. Since there were 6 subsets of 5 dichotomous items each, which resulted in 6 polytomous items with 5 thresholds each, the standard deviations of the thresholds τ_{ni} of Eq. (1), rather than the standard deviations of the items, are reported. Table 7 shows the results for the subtest analysis of Simulation set 2. Two statistics reported in previous tables were not applicable here and thus not reported: the percent of items with residual correlations greater than 0.1 and the precent of items with fit residuals more negative than the baseline. Figure 3 shows the distributions of person and threshold estimates graphically.

Person distribution and PSI:
As the magnitude of dependence (d) increased the *person mean* increased for patterns 1 and 3 but not pattern 2. The *person SD* decreased as the magnitude of dependence (d) increased for all patterns. Figure 3 shows this effect. The PSI decreased as d increased for all patterns.

Range and SD of the scale:
The *range* of the scale and *SD* of the scale decreased as d increased for all patterns.

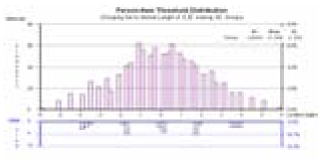
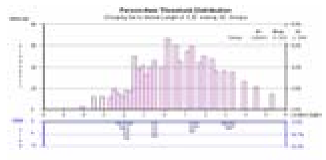
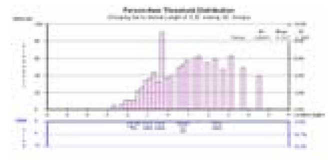
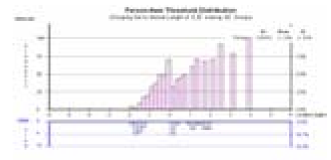
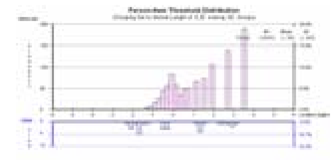
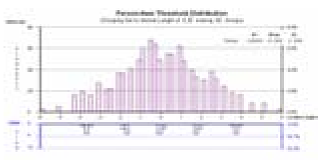
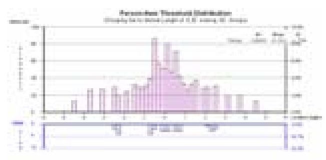
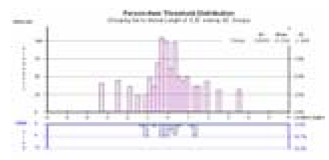
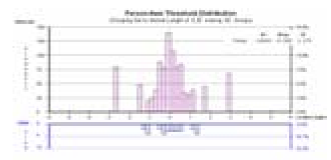
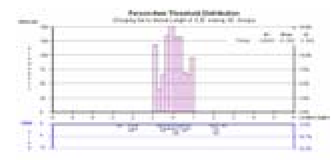
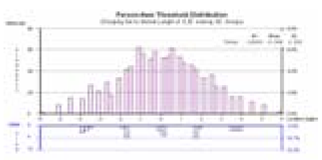
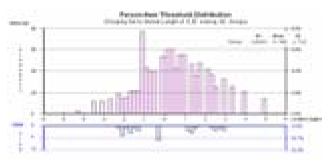
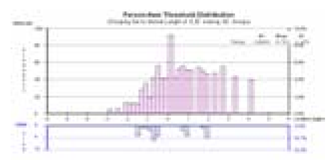
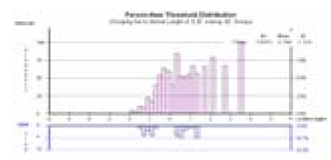
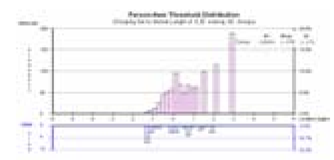
Deviations of person estimates:
The *Root mean Squared Difference (RMSD)* increased as d increased for all patterns. The *Mean Signed Difference (MSD)* became more negative as d increased for patterns 1 and 3 but not pattern 2. The c*orrelations between person estimates when d=0 and when d>0* decreases as d increased for all patterns. The *percent of person estimates with a logit shift greater than 0.5* increased as d increased for all patterns.

Item distribution:
The *SD of the thresholds* decreased as d increased for all patterns. Figure 3 shows this effect.

Table 7. Simulation set 2 – subtest analysis: Person mean, SD and PSI for all patterns as a function of d. Root mean squared difference (RMSD), Mean signed difference (MSD), Correlation (r) and Percent of person estimates with shift more than 0.5 logits from baseline (% log shift) as a function of d. Threshold SD, scale range, scale SD, percent of items with residual correlations > 0.1(% item residual r > 0.1) and percent of items with fit residuals more negative than baseline (% < fit residual) for all patterns as a function of d.

			Magnitude(d)				
Pattern			**0**	**1**	**2**	**3**	**4**
1 (Item 1 easiest)	**Persons**	**Person mean**	-0.01	0.52	0.92	1.19	1.38
		Person SD	2.05	1.91	1.68	1.55	1.43
		PSI	0.92	0.92	0.90	0.89	0.86
		Scale range	10.19	7.94	6.04	4.74	3.78
		Scale SD	2.80	2.21	1.63	1.22	0.90
		RMSD	0.00	0.74	1.18	1.51	1.75
		MSD	0.00	-0.53	-0.92	-1.20	-1.39
		r	1.00	0.97	0.94	0.91	0.87
		% log shift	0	64	91	96	97
	Items	**Threshold SD**	2.67	2.01	1.46	1.42	1.82
		% item residual r>0.1	-	-	-	-	-
		% < fit residual	-	-	-	-	-
2 (Item 1 average)	**Persons**	**Person mean**	0.00	-0.01	-0.02	-0.01	-0.02
		Person SD	2.05	1.71	1.41	1.27	0.56
		PSI	0.92	0.91	0,90	0,92	0.83
		Scale range	10.10	7.34	5.02	3.18	1.51
		Scale SD	2.78	1.80	1.08	0.64	0.37
		RMSD	0.00	0.56	0.87	1.06	1.54
		MSD	0.00	0.01	0.02	0.01	0.00
		r	1	0.97	0.94	0.90	0.93
		% log shift	0	35	48	50	51
	Items	**Threshold SD**	2.65	1.56	0.80	0.82	1.42
		% item residual r>0.1	-	-	-	-	-
		% < fit residual	-	-	-	-	-
3 (Preceding item)	**Persons**	**Person mean**	-0.01	0.38	0.72	0.94	1.08
		Person SD	2.05	1.72	1.47	1.31	1.17
		PSI	0.92	0.91	0.89	0.88	0.85
		Scale range	10.19	7.51	5.63	4.26	3.27
		Scale SD	2.79	2.02	1.47	1.07	0.77
		RMSD	0.00	0.69	1.10	1.39	1.59
		MSD	0.00	-0.39	-0.73	-0.95	-1.08
		r	1.00	0.97	0.94	0.91	0.88
		% log shift	0	51	84	96	98
	Items	**Threshold SD**	2.66	1.79	1.19	0.98	1.14
		% item residual r>0.1	-	-	-	-	-
		% < fit residual	-	-	-	-	-

Figure 3. Simulation set 2 – subtest analysis: Person and threshold distributions for all patterns as a function of d

	Magnitude(d)				
Pattern	**0**	**1**	**2**	**3**	**4**
1 (First subset Item easiest)					
2 (First subset Item average)					
3 (Preceding item in subset)					

In general, dependence is suggested from a range of statistical indices and the shape of the person distribution, but none is unequivocal in its own right in identifying local dependence. However, in the case of a subtest structure within which there is local dependence, the PSI seems to be relatively conclusive regarding dependence. To summarise this conclusion, Table 8 shows the PSI when items were analysed as dichotomous, and when dependent items were combined in a subtest analysis for Simulation set 2. In the subtest analysis the PSI decreased as a function of d. Even when d=0 there was a slight decrease for the subtest analysis even though there is no dependence in that condition. It can be shown algebraically that this is most probably due to a capitalisation on any chance dependence between items in a subtest.

Table 8. Simulation set 2: PSI for the first and subtest (st) analysis for all patterns as a function of d.

		Magnitude(d)				
Pattern		**0**	**1**	**2**	**3**	**4**
1 (Item 1 easiest)	**PSI**	0.93	0.93	0.93	0.92	0.91
	PSI(st)	0.92	0.92	0.90	0.89	0.86
2 (Item 1 average)	**PSI**	0.93	0.94	0.95	0.96	0.96
	PSI(st)	0.92	0.91	0.90	0.92	0.83
3 (Preceding item)	**PSI**	0.93	0.94	0.94	0.94	0.93
	PSI(st)	0.92	0.91	0.89	0.88	0.85

4. Discussion

This paper investigated the violation of local dependence in the dichotomous Rasch model in two main test designs. In the first, all items were discrete; in the second, items were identified within a subtest structure. Within each, three patterns of dependence were simulated. In addition to the structure and patterns of dependence, the magnitude of dependence was also varied.

The effects of dependence were noticeable in all the statistics reported. In particular, the fit statistics and the parameter estimates showed increasing discrepancies from their

theoretical values as a function of the magnitude of the dependence. In some cases, however, two related statistics gave the impression of improvement as a function of increased dependency; firstly, the standard deviation of person estimates showed an increase, and, secondly, the PSI, analogous to the traditional reliability index, showed relative improvement. In the presence of dependence these two related results appear more favourable than they really are.

The effects on the statistics were different for the different patterns of dependence, for example when items were dependent on an easy item the mean increased and the person distribution was skewed and unimodal. When items were dependent on an item of moderate difficulty the mean did not increase significantly and the person distribution was bimodal. The effects on the distribution help explain some of the effects on the statistics reported. To diagnose evidence of possible dependence, we conclude that it is necessary to study not only one, but all the statistics and their related effects in conjunction with each other, including the graphical distribution of person and item estimates.

The second test design permitted a closer study of the effect of accounting for dependence by combining dependent items into a polytomous item. The combining of items in a subtest analysis resulted in more realistic reliability. In practice, the combining can be done based on a priori evidence, for example when the known test structure shows which items belong together. It is important to appreciate that statistics like the PSI only give clues regarding the presence of dependence in a data set. Studying the test format and marking keys can then provide more specific information as to which items might show local dependence. Alternatively, a post hoc analysis, for example the correlations in the item residual matrix, can provide clues as to which items are dependent, but again, these are not unequivocal on their own.

Smith (2005) found person and item estimates to be fairly robust with regards to violations of dependence as simulated in that study. The person estimates in this study

were found not to be as robust. The differences between estimates when no dependence is present and estimates when dependence is present were large and of concern.

This paper investigated the effects of two main factors on the *dichotomous* Rasch model. Further simulation studies are currently being carried out to investigate the effects of local dependence on the Rasch model for ordered categories.

References

Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR.20 index, and the Guttman scale response pattern. *Education Research and Perspectives, 9* (1), 95-104.

Andrich, D (1985). A latent trait model for items with response dependencies: Implications for test construction and analysis. In S.E. Embretson (Ed.), *Test design* (pp.245-275). New York: Academic Press.

Andrich, D., Sheridan, B. & Luo, G. (1997-2005). RUMM2020. RUMM Laboratory, Perth, Australia.

Andrich, D. (1991). Essay review of Rolf Langeheine and Jurgen Rost, Latent Trait and Latent Class Analysis, New York, 1988. Plenum Press. *Psychometrika, 56*, 155-168.

Andrich, D. (2005) Georg Rasch: Mathematician and Statistician. In Kimberly Kempf-Leonard (Editor-in-Chief). *Encyclopedia of Social Measurement,* Academic Press, Amsterdam: Volume 3. 299- 306.

Gulliksen, H. (1950). *Theory of Mental Tests.* New York: Wiley.

Heldsinger, S and Humphry, S (2006). The violation of local independence in the measurement of writing. ARC Linkage Research Report No. 11, Murdoch University.

Ross, S, (1976). A first course in probability. New York. Collier Macmillan

Smith, E (2005). Effect of Item redundancy on Rasch Item and Person Estimates. Journal of Applied Measurement, 6(2), 147-163.

Wang, X., Bradlow, E.T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied Psychological Measurement*, *26 (1)*, 109-128.

Wilson M. & Adams R. J. (1995). Rasch models for item bundles. Psychometrika, 60, 181 – 198.

Zenisky, A.L., Hambleton, R.K., & Sireci, S.G., (2002). Identification and Evaluation of local item dependencies in the medical college admissions test. Journal of Educational Measurement, 39(4), 291-309.