Formalising dimension and response violations of local independence in the unidimensional Rasch model

Ida Marais

David Andrich

University of Western Australia

Mailing address
Ida Marais
Graduate School of Education
University of Western Australia
35 Stirling Highway
CRAWLEY WA 6009
Mailbox M428
Email: ida.marais@uwa.edu.au

Acknowledgements

The research reported in this paper was supported in part by an Australian Research Council Linkage 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.

Abstract

Local independence in item response theory models can be violated in two ways that are difficult to distinguish empirically and are generally not distinguished clearly in the literature. In this paper we distinguish between a violation of unidimensionality, which we call *trait dependence*, and a specific violation of statistical independence, which we call *response dependence*, both of which violate local independence. Distinct algebraic formulations for trait and response dependence are developed as violations of the dichotomous Rasch model, data are simulated with varying degrees of dependence according to these formulations, and then analysed according to the Rasch model assuming no violations. Relative to the case of no violation it is shown that trait and response dependence result in opposite effects on the unit of scale as manifested in the range and standard deviation of the scale and the standard deviation of person locations. In the case of trait dependence the scale is reduced; in the case of response dependence it is increased. Again, relative to the case of no violation, the two violations also have opposite effects on the person separation index (analogous to Cronbach's α reliability index of traditional test theory in value and construction): it decreases for data with trait dependence; it increases for data with response dependence. A standard way of accounting for dependence is to combine the dependent items into a higher-order polytomous item. This typically results in a decreased person separation index index and Cronbach's α, compared with analysing items as discrete, independent items. This occurs irrespective of the kind of dependence in the data, and so further contributes to the two violations not being distinguished clearly. In an attempt to begin to distinguish between them statistically this paper articulates the opposite effects of these two violations in the dichotomous Rasch model.

Formalising dimension and response violations of local independence in the unidimensional Rasch model

1. Introduction

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

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

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 $\tau_{1i}, \tau_{2i,...} \tau_{mi}, \sum_{x=0}^{m_i} \tau_{xi} = 0$ are thresholds between $m_i + 1$ ordered categories where m_i is the maximum score of item i, $\tau_0 \equiv 0$ (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 ordered dichotomous responses is

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

where $x \in \{0,1\}$ and there is only one threshold, δ_i. Both cases are used in this paper.

The model also implies statistical *independence of responses* in the sense that

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

where $((x_{ni}))$ denotes the matrix of responses $X_{ni} = x, n = 1....N, i = 1....I$.

The holding of Eq. (2) and (3) together is generally referred to as *local independence* or sometimes *conditional independence*. The term *local* refers to the idea that all the variation among responses to an item is accounted for by the variable β, and therefore

that for the same value of β, there is no further relationship among responses (Andrich, 1991; Lazarsfeld, & Henry, 1968).

Local independence defined in this way can be violated in two ways. First, there may be person parameters other than β that are involved in the response. This is a violation of unidimensionality and therefore statistical independence relative to the model of Eq. (2). To reflect the presence of more than one trait, in this paper the violation of unidimensionality is called *trait dependence*.

Second, for the same person and therefore the same value of β, the response to one item might depend on the response to a previous item. This is clearly a violation of statistical independence relative to Eq. (3). To distinguish this latter violation of Eq. (3) from the violation of unidimensionality, it will be referred to as *response dependence*. Both are formalised algebraically in Section 2. We note that it is possible to have a multidimensional model with response independence and a unidimensional model with response dependence. However, we are concerned with studying the effects of these violations relative to the unidimensional Rasch model which also implies statistical independence as defined in Eq (3).

Trait and response dependence, although conceptually different, are known to be difficult to distinguish empirically. The two violations are also generally not distinguished clearly in the literature with the term *multidimensionality* used for trait dependence and the generic term *local dependence* used for both trait and response dependence. Nonetheless, there have been concerns about the various effects of trait and response dependence in the Rasch model of Eq. (1) (e.g. Smith, 1996, Smith and Miao, 1994; Smith, 2002; Smith, 2005).

In practice, trait dependence is found in many scales in social measurement that are constructed to measure a single variable, but are composed of subsets of items which measure somewhat different aspects of the variable. Nevertheless, the responses across all items are intended to be summed. An example is the Functional Independence Measure (FIM™) motor scale (Keith, Granger, Hamilton, & Sherwin, 1987), which consists of 13 items, ranging from bladder management to climbing

stairs. These items can be grouped into subsets, for example S*phincter Control* can comprise *Bowel Management* and *Bladder Management*. Although the presence of subsets captures better the complexity of a variable and increases its validity, it compromises its unidimensionality. Another example is the Australian Scholastic Aptitude test (ASAT) where items, which are summed, are grouped into subsets representing mathematics, science, humanities and social science (Bell, Pattison and Withers, 1988).

Trait dependence is also found in items that are linked by attributes such as common stimulus materials, common item stems, common item structures or common item content. These have been described as *subtests* (Andrich, 1985), *testlets* (Wang et al, 2002) or *item bundles* (Rosenbaum, 1988; Wilson & Adams, 1995).

Response dependence is found when a person's response to an item depends on the response to a previous item, for example in assessments where a correct answer on a question gives a clue or determines the answer to one or more subsequent questions. In contrast to trait dependence, in general no one deliberately constructs items with response dependence.

An example of response dependence in practice is in the physical functioning subscale of the SF-36, a widely used rating scale in health research (Ware, Snow, Kosinski, & Gandek, 1993). Kreiner and Christensen (2007) show that the items *Climbing one flight of stairs* and *Climbing several flights of stairs* are dependent in this way. Similarly, the items *Walking one block*, *Walking several blocks* and *Walking more than a mile* are dependent in this way.

Response dependence is also found in satisfaction questionnaires in an item requesting an overall level of satisfaction following several other satisfaction items, as in the Course Experience Questionnaire used in Australian Universities (Wilson, Lizzio, & Ramsden, 1997). It is also found where judges make judgements on a set of criteria and a halo effect operates (Heldsinger & Humphry, 2006).

One way of detecting dependence, trait or response, in a data set is by comparing traditional reliability estimates or their modern test theory counterpart, from two

separate analyses of the data (e.g. Andrich, 1985; Marais & Andrich, 2005; Zenisky, Hambleton & Sireci, 2002). The first estimate uses the original items and assumes all items are statistically independent. In a second analysis, items hypothesized to be trait dependent or response dependent are combined into polytomous items and the data reanalysed as polytomous items. If the reliability estimate from the second analysis is lower than the reliability estimate from the first, trait or response dependence is confirmed. That trait and response dependence can have such an identical effect on a statistic further contributes to them not being clearly distinguished.

In this paper simulation algorithms for these two violations are developed from the algebraic formulations in Section 2. Data are simulated with varying degrees of trait or response dependence and then analysed according to the Rasch model to show typical effects in the analyses from these specific violations of the model. The effects of interest are on (i) the unit of scale as manifested in the range and standard deviation of the estimates for the possible scores and the standard deviation of person locations, on (ii) the person separation index (PSI), which is analogous to Cronbach's α of traditional test theory in both values and construction (Andrich, 1982), and on (iii) the standard deviation of the item locations.

It has been shown that response dependence often results in increased reliability and person estimates (e.g. Smith, 2005). We expect, therefore, that in the data with response dependence, the range and standard deviation of the scale, the standard deviation of person locations, and the PSI will increase compared to data with no dependence. On the other hand, we expect that in data with trait dependence opposite effects will be found: a decrease in the range and standard deviation of the scale, the standard deviation of person locations, and the PSI compared to data with no dependence (Marais & Andrich, 2005).

However, when data are analysed by combining trait or response dependent items into polytomous items, we expect a decrease in the PSI but also in the standard deviation of the person locations, and in the range and standard deviation of the scale, compared to data with no dependence (Andrich, 1985; Zenisky, Hambleton & Sireci, 2002).

2. Method

This section formalises the presence of trait and response dependence as violations of the Rasch model and describes the methods used to investigate their effects. The focus in the rest of the paper is on the model for dichotomous responses, but the principles can be generalised.

2.1 The formalisation of trait dependence – violation of unidimensionality

As indicated earlier, many scales in psychology, education and social measurement in general, which are constructed to measure a single variable, are nevertheless composed of subsets of items which measure different but related aspects of the variable. Consider a scale composed of $s = 1,2,...,S$ subsets, and let

$$\beta_{ns} = \beta_n + c_s \beta_{ns}^{'} , \qquad (4)$$

where $c_s > 0$, β_n is the common trait for person *n* among subsets and is the same variable as in Eq. (1), $\beta_{ns}^{'}$ is the *distinct* trait characterized by subset *s* and is uncorrelated with β_n, that is, $COV[\beta \ , \beta_s^{'}] = 0$. Therefore, β_n is the value of the main, common, variable or trait among subsets, and $\beta_{ns}^{'}$ is the variable or trait unique to each subset. The value c_s characterizes the magnitude of the variable of subset *s* relative to the common variable among subsets. The common variable β might be considered a *higher order* variable relative to the variables $\beta_s^{'}$ of the subsets. Further, variables of the subsets are considered to be mutually uncorrelated, that is $COV[\beta_s , \beta_t] = 0$ for all subsets *s* and *t*. Table 1 shows this subset design. In generating data with trait dependence, β_n of Eq. (2) is replaced by β_{ns} of Eq. (4) to give

$$\Pr\{X_{ni}^{(s)} = x\} = [\exp(x(\beta_{ns} - \delta_i))]/[1 + \exp(\beta_{ns} - \delta_i)] \qquad (5)$$

where the superscript s indicates subset $s = 1,2,...S$. Clearly, responses according to Eq. (5) violate Eq. (2).

Table 1 Summary of subset design

Subsets					
Items	1	2	.	.	*S*
1	$\beta_{n1} = \beta_n + c_1\beta'_{n1}$	$\beta_{n2} = \beta_n + c_2\beta'_{n2}$	.	.	$\beta_{nS} = \beta_n + c_S\beta'_{nS}$
2	$\beta_{n1} = \beta_n + c_1\beta'_{n1}$	$\beta_{n2} = \beta_n + c_2\beta'_{n2}$	.	.	$\beta_{nS} = \beta_n + c_S\beta'_{nS}$
.	.	.	.	.	.
.	.	.	.	.	.
I	$\beta_{n1} = \beta_n + c_1\beta'_{n1}$	$\beta_{n2} = \beta_n + c_2\beta'_{n2}$	.	.	$\beta_{nS} = \beta_n + c_S\beta'_{nS}$

Because each subset is composed of a distinct variable β'_{ns} as well as the common variable β_n, the correlation among the subsets is *not* 1. However, because it has the common variable β_n, and depending on the sizes of c_s and c_t, the correlation will generally be greater than 0. Within the above constraints, any correlation among any two subsets of a scale can be specified by setting

$$V[\beta] = V[\beta_s] = \sigma^2 \tag{6}$$

for all subsets *s* and *t*.

Further modifications of the mean and variance of the variables measured by the scale can be generated by transforming β_{ns} according to

$$B_{ns} = a_s + b_s\beta_{ns}, \tag{7}$$

where a_s is the mean and b_s is the standard deviation of B_s.

The Appendix shows that the theoretical, latent correlation ρ_{st} between two items in different subsets *s* and *t* from the construction in Table 1 is given by

$$\rho_{st} = \frac{1}{\sqrt{1+c_s^2]}\sqrt{1+c_t^2}} \quad . \tag{8}$$

Assuming that $c_s = c_t = c$ gives

$$\rho_{st} = \frac{1}{1+c^2} . \quad (9)$$

Clearly, the larger the value of c, the smaller the correlation between two items of different subsets: e.g. if $c = 0$, then $\rho_{st} = 1$, and if c =10, then $\rho_{st} = 0.00999...$. Therefore, in generating responses, the number of subsets and the value of c_s are specified so that a person *n* has a location β_{ns} in responding to the items of each subset $s = 1,2,...,S$. The appendix also shows that ρ_{st} , the correlation between two items from different subsets *s* and *t*, is also the latent correlation among the subsets themselves.

2.2 The formalisation of response dependence – violation of statistical independence

Response dependence is formalised by making a person's response on an item be a function of the person's response to a previous item. Specifically, the probability of a person's positive response on an item increases as a function of the positive or correct response, and decreases as a function of the negative or incorrect response, on a previous item on which it depends. How much the probability increases or decreases can be determined by a constant, more specifically, by adding or subtracting a constant from the location or difficulty of the dependent item.

Eq. (10) formalises this construction for item *j* dependent on item *i* by modifying Eq. (2) according to:

$$\Pr\{X_{nj} = 1 \mid X_{ni} = 1\} = [\exp(\beta_n - (\delta_j - d))]/[1 + \exp(\beta_n - (\delta_j - d))] \quad (10)$$

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 $d > 0$ is the constant used to increase or decrease the magnitude of dependence.

The general equation which includes both responses $X_{nj} = 1, X_{nj} = 0$ takes the form

$$\Pr\{X_{nj} = x_j \mid X_{ni} = x_i\} =$$
$$[\exp(x_j(\beta_n - \delta_j - (1-2x_i)d))]/[1+\exp(x_j(\beta_n - \delta_j - (1-2x_i)d))]. \quad (11)$$

It is evident from Eq. (11) that the response of a person to item *j* depends on the person's response to item *i*. Specifically, and taking items as achievement items with correct or incorrect responses, if a person's response to item *i* is $x_{ni} = 1,$ then the dependent item *j*'s difficulty is effectively changed to $\delta_j - d$ for that person, making the dependent item easier, thus increasing the probability of the same response of $x_{nj} = 1$. Conversely, if a person's response to item i is $x_{ni} = 0$**,** then the dependent item *j*'s difficulty is changed to $\delta_j + d$ for that person, making the dependent item more difficult, thus increasing the probability of the same response $x_{nj} = 0$. In summary, responses according to Eq. (11) violate Eq. (2).

2.3 The simulation design structure for trait and response dependence

Using the formulations described in Sections 2.1 and 2.2, data sets were generated with trait dependence or response dependence of varying degrees. No condition had both trait and response dependence at the same time.

Many different parameters and patterns could potentially be developed when simulating trait and response dependence respectively. To make effective comparisons between the effects of the two types of dependence, *the basic item and person design was made the same for both*. Data were generated for 1000 persons and 30 items. Item locations and the distribution of person locations were chosen to make the targeting of persons to items relatively ideal. The distribution of person locations was $N(0,1^2)$ and the items had locations distributed uniformly from –2 to 2. The 30 items were divided into 6 subsets of 5 items each. Dependence, trait or response, was kept within subsets. Table 2 shows the item difficulties and the subdivision of items of varying difficulty into 6 subsets.

Table 2 Item difficulties

Subset 1		Subset 2		Subset 3		Subset 4		Subset 5		Subset 6	
Item no	**Item** δ	**Item no**	**Item** δ	**Item no**	**Item** δ	**Item no**	**Item** δ	**Item no**	**Item** δ	**Item no**	**Item** δ
1	-2.00	**6**	-1.31	**11**	-0.62	**16**	0.07	**21**	0.76	**26**	1.45
2	-1.86	**7**	-1.17	**12**	-0.48	**17**	0.21	**22**	0.9	**27**	1.59
3	-1.72	**8**	-1.03	**13**	-0.34	**18**	0.34	**23**	1.03	**28**	1.72
4	-1.59	**9**	-0.9	**14**	-0.21	**19**	0.48	**24**	1.17	**29**	1.86
5	-1.45	**10**	-0.76	**15**	-0.07	**20**	0.62	**25**	1.31	**30**	2.00
Mean	-1.72		-1.03		-0.34		0.34		1.03		1.72
SD	0.22		0.22		0.22		0.22		0.22		0.22
Mean: 0.00; SD=1.21 for all items											

2.3.1 Trait dependence

For trait dependence the *latent variables underlying the subsets* were varied systematically by setting the value of the constant c in Eq. (4) to 0, 1, or 2. These values have corresponding correlations among subsets of 1.0, 0.5 and 0.2 respectively. The case where $c = 0$, which gives a correlation of 1.0 among subsets and is the case of data fitting the unidimensional dichotomous Rasch model perfectly, provides a frame of reference for the analyses when $c \neq 0$.

2.3.2 Response dependence

The example of response dependence between items that was constructed was based on a judge making judgements on a set of criteria where responses to these criteria were all dependent on an overall judgement, in effect a halo item. *The latent variables underlying the subsets were perfectly correlated but all the responses of all items within a subset were made dependent on the halo item.* Six halo items, one for each subset, were therefore generated and responded to first. The responses to the five remaining items within each of the six subsets were governed by their response to this halo item in the subset. The location of the halo item was the same as the difficulty of the middle item of a subset, which was the average difficulty in the

subset. Following the data simulation, these halo items were deleted from the analyses, leaving only 30 items in 6 subsets as in the trait dependence condition. The magnitude of dependence was varied by setting d in Eq. (11) to 0, 1, or 2 respectively. The case of $d = 0$, in which the data fit the Rasch model perfectly, again provides a frame of reference for the analyses when $d \neq 0$.

Thus the *basic design structure was the same* for trait and response dependence: the *difference between them was that for trait dependence an extra, unique dimension for all items in a subset governed the responses to items in that subset; for response dependence the response to a halo item for each subset governed the responses to all items in that subset.*

2.4 Analyses

The generated data sets were analysed with the RUMM2020 software (Andrich, Sheridan and Luo, 2005). Two analyses were performed on each data set. Firstly, all items were analysed as discrete, dichotomous items according to the dichotomous Rasch model assuming unidimensionality and statistical independence. In order to account for the dependence, responses of dependent dichotomous items within a subset were summed to give a polytomous item. Then a second analysis was performed on the polytomous items according to Eq. (1) assuming unidimensionality and statistical independence (Andrich, 1985; Wang etal, 2002; Wilson, & Adams, 1995; Zenisky et al, 2002).

The range and SD of the scale: 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 taking or not taking account of dependence, there is a single ability estimate for each raw score. This estimate is a non linear transformation of the total score which is a function of the item locations. Thus irrespective of taking or not taking into account the dependence, the analysis will have the same raw score range with the same distribution of raw scores. However, if there is dependence in the data, estimates of abilities will be different for the same 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. These effects are referred to as effects on the scale.

Therefore, the effects on the range and standard deviation of the scale are of interest and 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. Scores of 0 and $\sum_i m_i$, which in theory have infinite estimates and in practice require some extrapolation, are not used for calculating the range and SD of the scale.

Person distribution: One consequence of change of scale described above is the impact on the person distribution. Therefore the effects of dependence on the *standard deviation* of the person distribution are reported.

Person Separation Index: 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)

$$r_{xx} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_e^2} = \frac{\sigma_x^2 - \sigma_e^2}{\sigma_x^2}, \qquad (12)$$

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. In parallel to Eq. (12) the PSI is defined according to

$$r_{\beta\beta} = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_e^2} = \frac{\sigma_{\hat{\beta}}^2 - \sigma_{\hat{e}}^2}{\sigma_{\hat{\beta}}^2} \qquad (13)$$

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 of the error of measurement, 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 reliability index, Cronbach's α, are virtually identical (Andrich, 1982).

Item distribution: The change of scale is a reflection of the effect of violations of the model on item locations. Therefore the effects of dependence on the *standard deviation* of the item locations are also reported.

3. Results

3.1 First analysis – as dichotomous items

As indicated previously, in the first analysis all items were analysed as discrete, dichotomous items according to the dichotomous Rasch model assuming unidimensionality and statistical independence. Table 3 shows the results for this analysis and Figure 1 shows the distributions of person and item estimates graphically.

Table 3 First analysis – as dichotomous items. Person SD, PSI, scale range and SD and Item SD for trait dependence as a function of c and for response dependence as a function of d.

Type of dependence				**Magnitude of dependence**		
			c:	**0**	**1**	**2**
Trait dependence	**Persons**	**Person SD**		1.11	0.79	0.62
		PSI		0.83	0.70	0.55
		Scale range		7.48	6.99	6.88
		Scale SD		1.90	1.78	1.75
	Items	**Item SD**		1.25	1.11	1.04
			d:	**0**	**1**	**2**
Response dependence	**Persons**	**Person SD**		1.11	1.42	1.59
		PSI		0.83	0.88	0.90
		Scale range		7.48	7.78	7.72
		Scale SD		1.90	2.03	2.01
	Items	**Item SD**		1.25	1.53	1.50

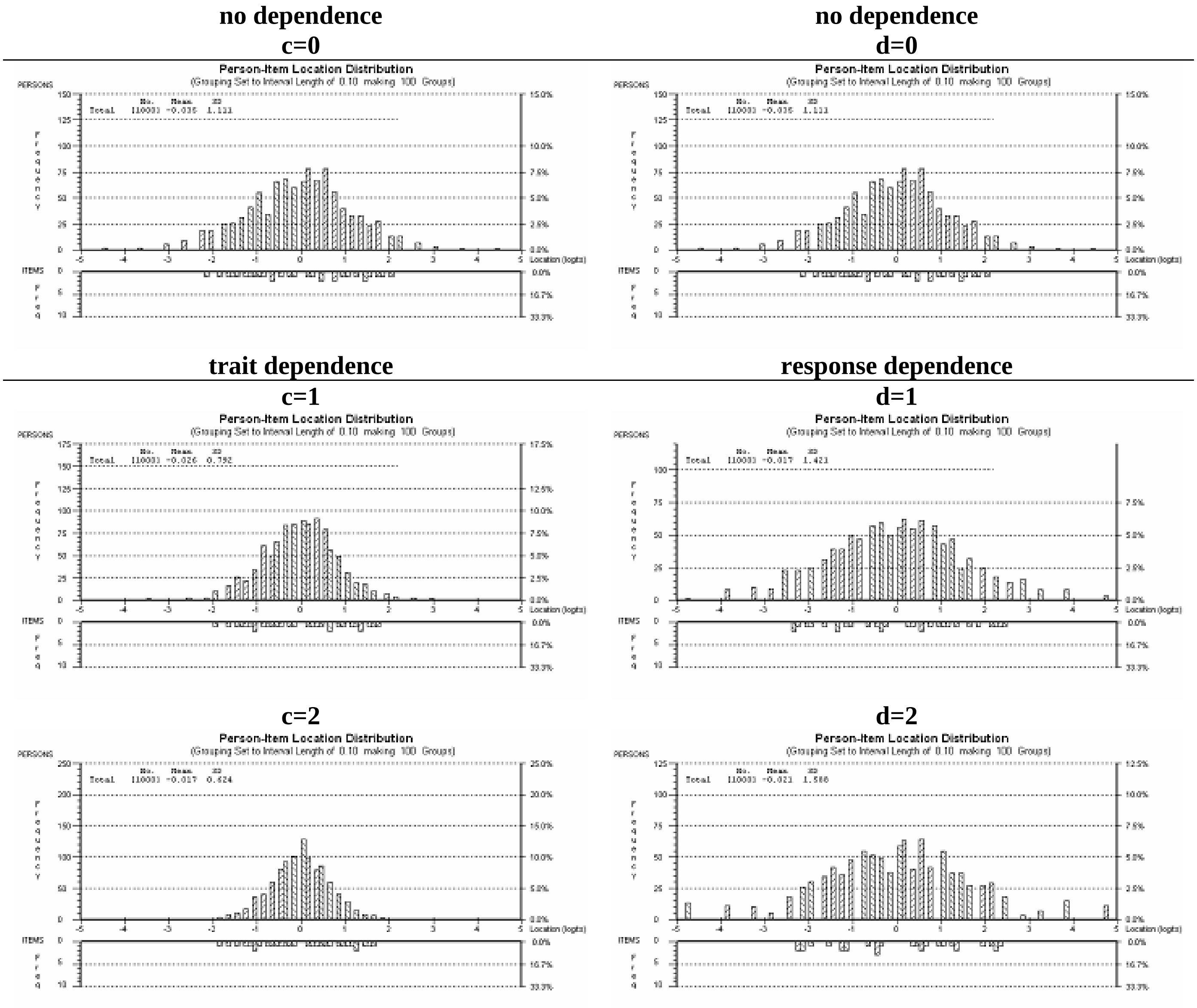

Figure1. First analysis – as dichotomous items: Person and item distributions for both types of dependence as a function of c and d.

In order to better compare the effects of trait and response dependence relative to no violation of the Rasch model, Figure 2 shows selected results from Table 3 in graphical form. Specifically, it shows the SD of person locations, PSI, range and SD of the scale as a function of c or d.

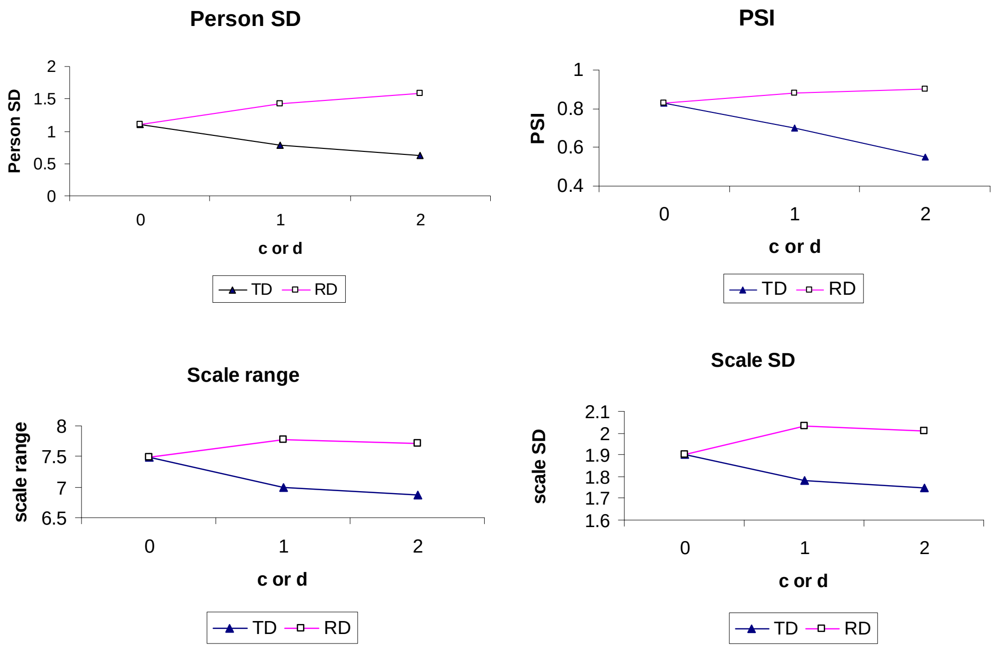

Figure 2. First analysis - All dichotomous items: Person SD, PSI, range and SD of the scale as a function of c and d for the first analysis. TD = Trait dependence condition, RD = Response dependence condition.

It is evident from Table 1 and Figures 1 and 2 that when data are analysed not accounting for the violations, the *person SD* decreased as the magnitude of trait dependence (c) increased. In contrast, the *person SD* increased as the magnitude of response dependence (d) increased.

It has been shown that violations of statistical independence often result in increased reliability (e.g. Smith, 2005). As is evident from Table 3 and Figure 2 again, the analysis of the simulated data in this paper confirms that result. The *PSI* increased as the magnitude of response dependence (d) increased. In contrast, the *PSI* decreased as the magnitude of trait dependence (c) increased.

A similar effect is evident in the *range and SD of the scale*: The range and SD decreased as trait dependence (c) increased; in contrast, the range and SD increased, then remained relatively constant, as response dependence (d) increased. It may seem anomolous that the range and SD increase then stay relatively constant but the Person SD increased consistently as d increased. An explanation of this anomaly is that the impact of response dependence on the scale may have reached a limit, but the Person SD still increased as the frequencies in the extremities of the score range continued to increase as d increased.

3.2 Second analysis – as polytomous items

It will be recalled that in the second analysis, and in order to account for dependence, items within subsets were summed to form a polytomous item. The polytomous items were analysed according to Eq (1) assuming unidimensionality and statistical independence. Because 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_{xi}, $x = 1,2,....5, i = 1,2,....6$ of Eq. (1), rather than the standard deviations of the items, are reported. Table 4 shows the results for this analysis and Figure 3 shows the distributions of person and item estimates graphically. Figure 4 shows the comparative effects graphically.

Table 4 Second analysis – as polytomous items. Person SD, PSI, scale range and SD and Threshold SD for trait dependence as a function of c and for response dependence as a function of d.

Type of dependence				**Magnitude of dependence**		
			c:	**0**	**1**	**2**
Trait dependence	**Persons**	**Person SD**		1.08	0.59	0.41
		PSI		0.83	0.60	0.30
		Scale range		6.95	5.69	5.17
		Scale SD		1.80	1.38	1.23
	Items	**Threshold SD**		1.60	1.13	0.95
			d:	**0**	**1**	**2**
Response dependence	**Persons**	**Person SD**		1.08	0.86	0.50
		PSI		0.83	0.80	0.68
		Scale range		6.95	4.98	2.42
		Scale SD		1.80	1.24	0.63
	Items	**Threshold SD**		1.60	1.04	0.89

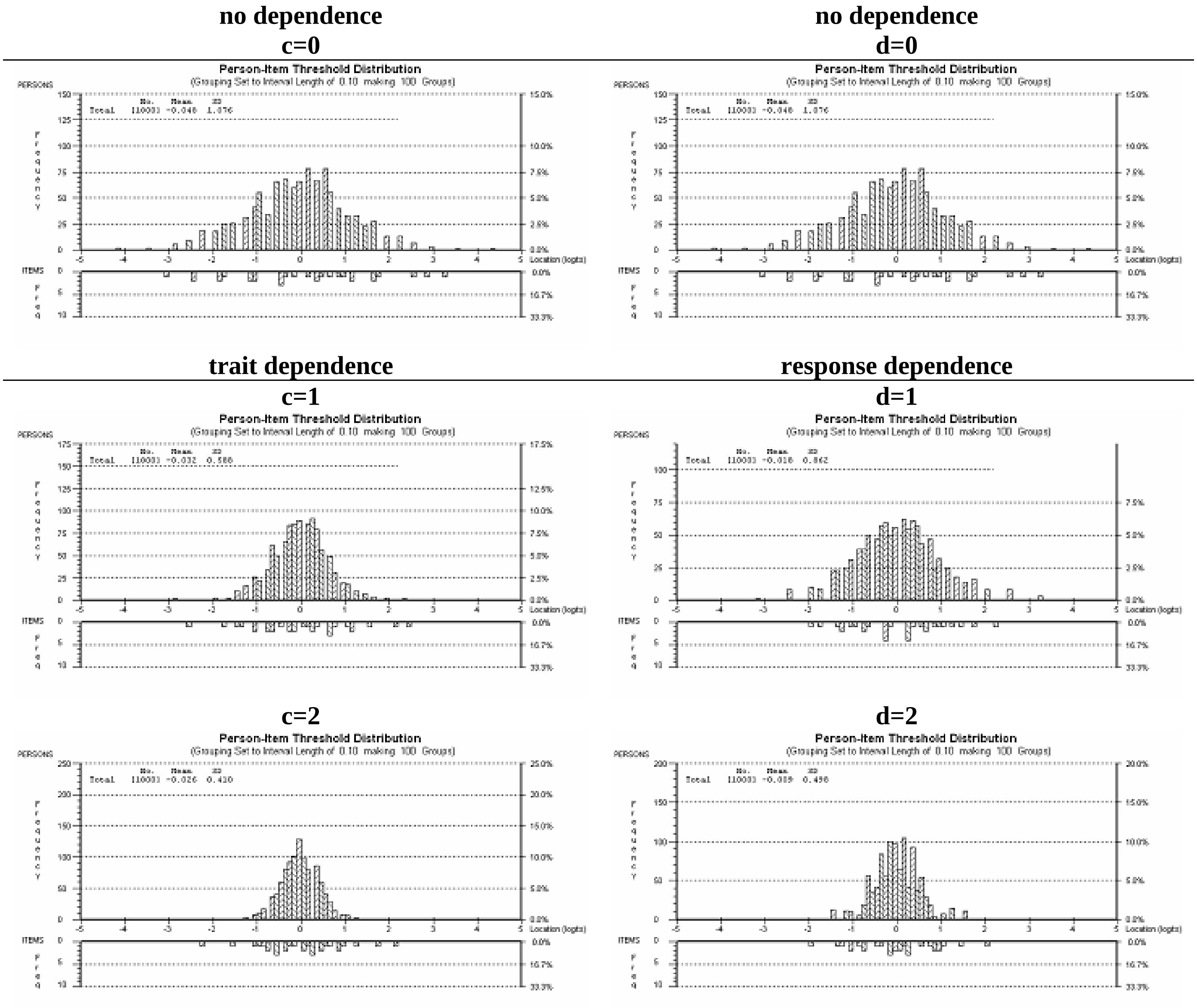

Figure 3. Second analysis – as polytomous items: Person and threshold distributions for both types of dependence as a function of c and d.

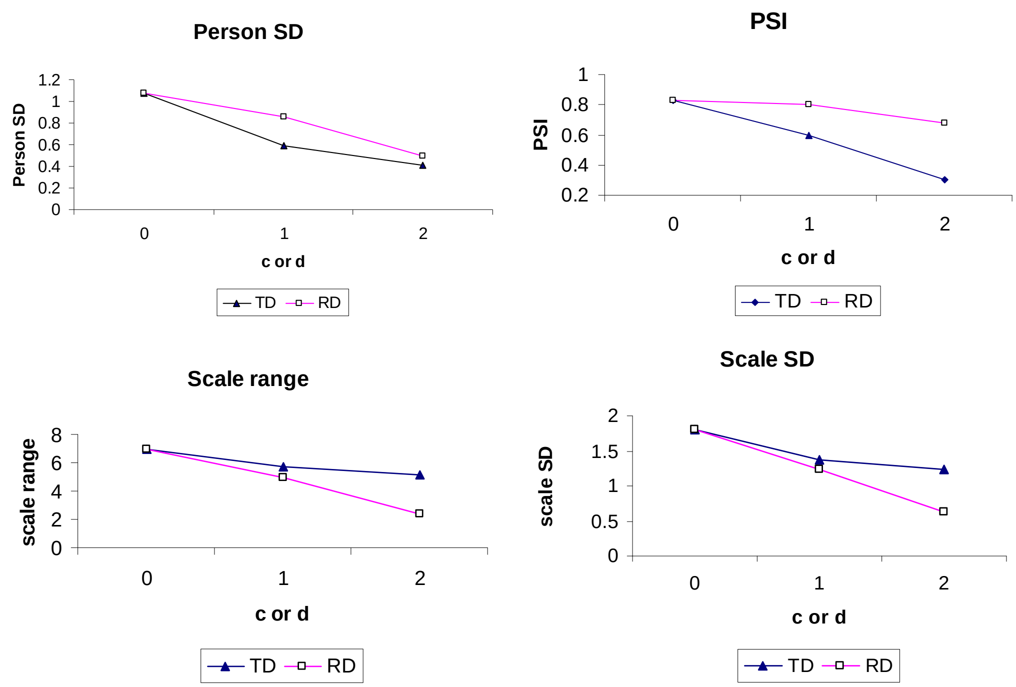

Figure 4. Person SD, PSI, range and SD of the scale as a function of c and d for the subtest analysis. TD = Trait dependence condition, RD = Response dependence condition.

It is evident from Table 4 and Figures 3 and 4, that when there is no dependence, the effects on the parameters are negligible, irrespective of the method of analysis. However, *in the presence of both trait and response dependence*, analysis using polytomous items *reduced* each of the statistics reported as a function of the magnitude of dependence.

In particular, when there was either response dependence or trait dependence in the data, the *PSI* decreased compared with the no dependence condition. The decrease becomes more pronounced with higher magnitudes of dependence. This decreased PSI in the presence of *both* trait and response dependence contributes to the two types of dependence being difficult to distinguish empirically.

4. Conclusions

Trait and response dependence are known to be difficult to distinguish empirically and are generally not distinguished clearly in the literature. In this paper trait and response dependence are formalised algebraically, data are simulated with varying

degrees of dependence according to these formulations, and then analysed according to the Rasch model assuming no violations.

When the simulated data in this paper were analysed assuming no violation the two types of dependence resulted in *opposite effects* on person estimates and an estimate of reliability, the PSI. Trait dependence resulted in a reduced scale; in contrast, response dependence resulted in an increased scale, relative to the case of no violation. Similarly, reliability decreased for data with trait dependence, and it increased for data with response dependence relative to the case of no violation. This confirms the finding that response dependence results in inflated person estimates and reliability (e.g. Smith, 2005), and shows that trait dependence results in understated reliability and reduced variance in the person estimates (Marais & Andrich, 2005).

These opposite effects can be appreciated by reflecting on the nature of the respective violations. With trait dependence responses regress to the middle and do not reinforce each other as a function of the same value of β – the greater the value of c in Eq. (4), the more random the responses, with the total scores regressing to the middle and narrowing the distribution of total scores with lower relative frequencies of low and high total scores. In contrast, with response dependence, a high score on the item on which items depend increases the chances of high scores on these items, and vice versa, thus reinforcing the effect of a high or low score, and spreading the distribution of total scores with greater relative frequencies of low and high total scores.

A standard way of accounting for dependence of either kind is to combine dependent items by the summing of their scores into a polytomous item. The paper confirms that such an analysis typically results in a decreased scale and decreased traditional reliability compared with analysing items assuming no violation. This implies that reliability is inflated when there is dependence *of either kind* in data and these data are analysed assuming no dependence. This is consistent with the understanding that a traditional reliability index, such as Cronbach's α, is not a measure of unidimensionality. The decrease in reliability when taking dependence, of either kind, into account contributes to the two violations being difficult to distinguish empirically.

Figure 5 summarises how response and trait dependence in data, but not accounting for them in the analysis, have *opposite effects* on the PSI - it increases for response dependence and decreases for trait dependence compared with the data with no dependence. However, when analysed taking dependence into account, it results in the *same effect* on the PSI - it decreases for both trait and response dependence, confirming that it was inflated in both cases in the analysis that does not take dependence into account.

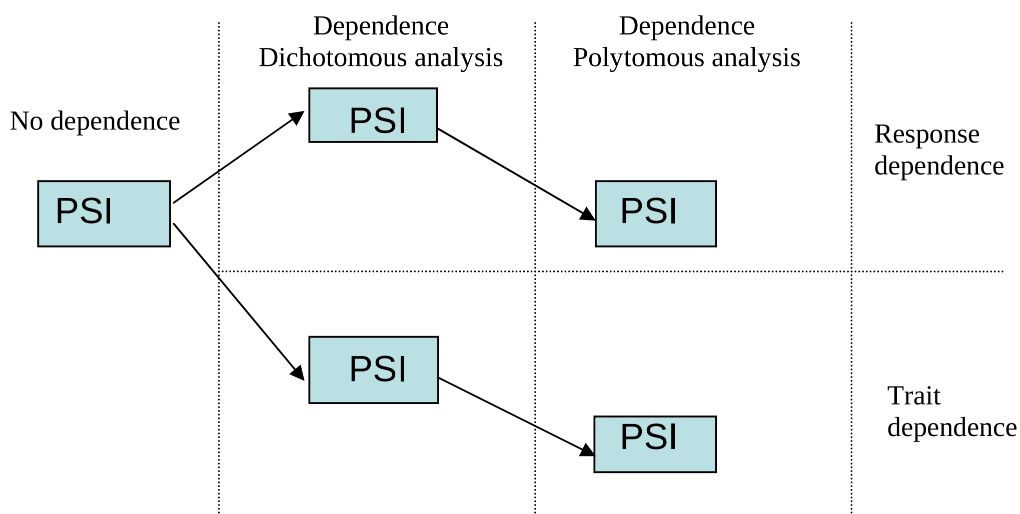

Figure 5. Flowchart of changes in PSI when there is no dependence and when there is either trait or response dependence and data are analysed as dichotomous items and as polytomous items.

Because in real data the effect from forming polytomous items and analysing them is similar, that is, decreased reliability for both trait and response dependence, it is not possible at this stage to distinguish between the two types of dependence from simple statistical analyses. Therefore, in the presence of such evidence of some dependence, the test or questionnaire format is the main source of information as to which of trait or response dependence is operating. For example, in the case of ASAT, composed of different aspects of scholastic achievement, it could be interpreted as trait dependence. On the other hand, in the case of the SF-36, it would be interpreted as response dependence.

Even though decreased reliability when accounting for dependence in an analysis does not help in determining which of trait or response dependence is present, it can give an indication of the *magnitude* or *degree* of dependence present. In the case of multidimensionality for instance, if the magnitude or degree of dependence is large, it may be better to analyse subsets separately as subscales. However, this is a relative issue and depends on the purpose of the scale or assessment.

This paper demonstrates the differences and opposite statistical effects of two conceptually distinct violations of the Rasch model. This may be a starting point in attempting to distinguish between them when real data indicate the presence of some form of dependence. In real data, the challenge is even greater in distinguishing between them because both may be present and may compensate for each other.

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. (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.

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

Bell, R.C., Pattison, P.E. & Withers, G.P. (1988). Conditional Independence in a Clustered Item Test. Applied Psychological Measurement, 12(1), 15-26.

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.

Keith, R.A., Granger, C.V., Hamilton, B.B. & Sherwin, F.S. (1987). The functional independence measure: A new tool for rehabilitation. In M.G. Eisenberg, & R.C. Grzesiak (eds), Advances in Clinical Rehabilitation (Volume 1. pp. 6-18). New York, Springer Publishing Co.

Kreiner, S. & Christensen, K.B. (2007) Validity and Objectivity in Health_Related Scales: Analysis by Graphical Loglinear Rasch Models. In M von Davier & C.H. Carstensen (eds). Multivariate and Mixture Distribution Rasch Models: Extensions and Applications (pp. 329-346). New York, Springer Publishing Co.

Lazarsfeld, P.F. & Henry, N.W. (1968). Latent structure analysis. Boston: Houghton Mifflin.

Marais, I. and Andrich, D. (2005). Studies on the effect of violations of local independence on scale in Rasch models. ARC Linkage Research Report No. 1, Murdoch University.

Rosenbaum, P. R. (1988). Item bundles. Psychometrika, 53, 349-359.

Smith, E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. Journal of Applied Measurement, 3, 205-231.

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

Smith, R.M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. Structural Equation Modeling, 3, 25-40.

Smith, R.M., and Miao, C.Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (ed.), Objective Measurement: Theory into Practice (Volume 2, pp. 316-327). Norwood, NJ:Ablex.

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

Ware, J.E.,Jr., Snow, K.K., Kosinski, M. & Gandek, B. (1993). SF-36 health survey. Manual and interpretation guide. Boston: The Health institute, New England Medical Center

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

Wilson, K. L, Lizzio, A. & Ramsden, P. (1997). The development, validation and application of the Course Experience Questionnaire. Studies in Higher Education, 22(1), 33 - 53.

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.

Appendix

The latent correlation among two items in different subsets and between two subsets of items

From Eq. (4)

$\beta_{ns} = \beta_n + c_s\beta'_{ns}$ for subset s and therefore $\beta_{nt} = \beta_n + c_t\beta'_{nt}$

for subset *t*.

Therefore

$V[\beta_s] = V[\beta + c_s\beta'_s] = V[\beta] + c_s^2 V[\beta'_s]$;

$V[\beta_t] = V[\beta + c_t\beta'_t] = V[\beta] + c_t^2 V[\beta'_t]$.

Further, because

$COV[\beta, \beta'_t] = COV[\beta, \beta'_s] = COV[\beta'_s, \beta'_t] = 0$.

$COV[\beta_s, \beta_t] = COV[\beta + c_s\beta'_s, \beta + c_t\beta'_t] = COV[\beta, \beta] = V[\beta]$.

Therefore the correlation ρ_{st} between two items from different subsets is given by

$$\rho_{st} = \frac{COV[\beta_s, \beta_t]}{\sqrt{V[\beta_s]}\sqrt{V[\beta_t]}} = \frac{COV[\beta_s, \beta_t]}{\sqrt{V[\beta + c_s\beta'_s]}\sqrt{V[\beta + c_t\beta'_t]}}$$

$$= \frac{V[\beta]}{\sqrt{V[\beta] + c_s^2 V[\beta'_s]}\sqrt{V[\beta] + c_t^2 V[\beta'_t]}}.$$

Given $V[\beta] = V[\beta'_s] = V[\beta'_t]$,

$$\rho_{st} = \frac{1}{\sqrt{1 + c_s^2}\sqrt{1 + c_t^2}}.$$

For $c_s = c_t = c$, $\rho_{st} = \dfrac{1}{1 + c^2}$.

For K_s, K_t items in subsets *s* and *t* respectively, the total latent scores, independent of error, for person *n* are given respectively by

$$K_s\beta_{ns} = K_s\beta_n + K_s c_s\beta'_{ns};\quad K_t\beta_{nt} = K_t\beta_n + K_t c_t\beta'_{nt}.$$

Therefore,

$$V[K_s\beta_s] = K_s^2V[\beta] + K_s^2c_s^2V[\beta_s^{'}]\,;\;\; V[K_t\beta_t] = K_t^2V[\beta] + K_t^2c_t^2V[\beta_t^{'}]$$

and

$$COV[K_s\beta_s, K_t\beta_t]] = K_sK_tV[\beta].$$

Therefore the correlation $\rho_{st}^{'}$ between two different subsets is given by

$$\begin{aligned}\rho_{st}^{'} &= \frac{K_sK_tV[\beta]}{\sqrt{K_s^2V[\beta] + K_s^2c_s^2V[\beta_s^{'}]}\sqrt{K_t^2V[\beta] + K_t^2c_t^2V[\beta_t^{'}]}} \\ &= \frac{K_sK_tV[\beta]}{K_s\,K_t\,\sqrt{V[\beta] + c_s^2V[\beta_s^{'}]}\sqrt{V[\beta] + c_t^2V[\beta_t^{'}]}}\end{aligned}$$

and again with $V\beta] = V[\beta_s^{'}] = V[\beta_t^{'}]$,

$$\rho_{st}^{'} = \frac{1}{\sqrt{1 + c_s^2}\sqrt{1 + c_t^2}} = \rho_{st}$$

which is identical to the correlation between two items from different subsets.