Detection of Local Dependence across Latent Traits When Common Passages Are Used

Wen-Chung Wang

National Chung Cheng University

Mark Wilson

University of California at Berkeley

Ying-Yao Cheng

National Sun Yat-Sen University

Correspondence:

Department of Psychology

National Chung Cheng University

Chia-Yi, Taiwan

Telephone: 886-5-2720411 ext. 6430

Fax: 886-5-2720857

E-mail: psywcw@ccunix.ccu.edu.tw

Abstract

It is quite often that a common passage is used to measure more than one latent trait. For instance, we might ask subjects to judge how important is a property (e.g., plentiful experiences, abundant knowledge, open-minded) to creativity development and how much do they posses this property. In order to save space, the common passages are listed in the first column, the ratings on "importance" in the second, and the rating on degrees of "possession" in the third. With this editing, the subjects read the common passage first, judge its importance, and then judge the degrees they have. After responding to the first common passage, they move on to the second common passage, judge the importance and then the degrees, etc. It is suspicious that a response procedure like this would result in interference between the two latent traits because people might overestimate (or underestimate) their degrees of possess if they consider the property (un)important. In this paper, we propose a procedure to detect local dependence between traits when common passages are used. To do so, a virtual item (may be called *item bundle*) should be formed by crossing the response categories in the first trait and those to the second trait. Some parameters are added to model the local dependence. Results of simulation studies show the parameters could be recovered very well. A real data set was analyzed to show implications and applications of the proposed detection procedures.

Keywords: Local dependence, parallel design, item bundle, multidimensional item response model, Rasch model.

Local independence is a critical assumption for item response models and many other statistical models. If the assumption is violated, analysis with item response models based on it is misleading. When a test contains common stimulus for a subset of items, called item bundle or testlet, or some item sets require correct solution of earlier items to solve later ones, local independence may not hold. Rosenbaum (1984, 1988) introduced the idea of *bundle independence* as a way to address these issues. He proposed that instead of expressing the likelihood of a response vector as a product of the probability of each individual item, the likelihood is expressed as a product of the probabilities of the response to the bundles among which one might expect local dependence. In other words, local independence is assumed between bundles not within bundles.

The concept of bundle independence was applied by Wilson and Adams (1995) to the partial credit model and related models, resulting in a type of model they called the item bundle model. A similar approach was taken by Hoskens and De Boeck (1997) to study other types of local dependence. These works focus on local dependence within tests rather than between tests. Is it possible that local dependence exhibit between tests (dimensions or latent traits)? It is, for some particular testing situations. If local dependence is expected between dimensions, the usual unidimensional item response models are not applicable, rather, multidimensional item response models are needed. In the following, we describe a particular inventory design that may cause local dependence between dimensions. A multidimensional Rasch-type model that can be used to model the local dependence is addressed. Several kinds of modeling for the local dependence are proposed. A simulation study was conducted to assess parameter recovery and the results are summarized. A real data set was analyzed to illustrate applications and implications.

The Parallel Design of the Inventory

Cheng and Wang (1999) designed an inventory that contains 15 personal characters. Subjects were asked to judge how important are the characters to creativity development and how much they possess them, both in a three-point Likert-type scale: very little, not much, and very much. The inventory is shown in the Appendix. For the sake of subjects' convenience and space reduction, these characters were printed on the left, and the two scales were parallel on the right. This design made the subjects judge the first character in terms of importance, followed immediately by the degrees of possession, then move to the second character, and so on. This test design is called "parallel design" as the first scale is printed parallel with the second scale.

With this particular response pattern, it is possible that the responses to the second scale are affected by those to the first scale. For example, the subjects might overestimate (or underestimate) their degrees of possession, when they consider the character important to creativity development. This interference is similar to local dependence within item bundles. The difference is that we are dealing local dependence between tests rather than within tests.

If the 15 characters with the first scale (i.e., the first test) had printed on one page and the characters with the second rating (i.e., the second test) on another, the subjects would have responded to all the characters in terms of importance, and responded to them again in terms of possession. This test design is called "sequential design." In such a case, this response pattern is identical to what is usually done when two tests are administered. There should have no interference between the two test responses.

Multidimensional Bundle Modeling

Since in the inventory local dependence is expected between dimensions, the usual unidimensional item bundle models are not applicable. However, the idea of bundle independence is still useful. We could form a bundle that contains a character with two ratings. Given two three-point scales were used, the bundle has nine response categories, one for each score vector. Instead of modeling the probabilities of items in a test, we analyze the bundles. With the Rasch-type models, at most eight item parameters can be freely estimated for a nine-response-category bundle because one category is constrained for model identification. These eight item parameters for a bundle could be modeled as in Table 1. It is referred to as the *saturated* model because all the possible parameters are estimated.

The modeling in Table 1 is analogous to the structural model of two-way analysis of variance (ANOVA). For example, consider there are two factors, *A* and B, each with three levels. There are nine cells altogether. The means of the nine cells can be partitioned into one grand mean, one main effect for Factor *A*, one main effect for Factor *B*, and one interaction effect for the two factors. Each main effect has two degrees of freedom. The interaction effect has four degrees of freedom. Conceptually, the two step difficulties for dimension 1, d_1 and d_2 in Table 1, are analogous to the main effect of Factor *A*. The two step difficulties for dimension 2, e_1 and e_2, are analogous to the main effect of Factor *B*. The other four parameters, t_1, t_2, t_3, t_4, are analogous to the interaction effect and are thus referred to as *interaction parameters*. However, the modeling in Table 1 is different from that in two-way ANOVA in that the main effects and the interaction effects can be perfectly partitioned only with ANOVA.

Table 1

Probabilities for the saturated model for a nine-category bundle

Response category	Score		Probability
	Dim. 1	Dim. 2	
0	0	0	$1/g$
1	1	0	$\exp(q_1 - d_1 + t_1)/ g$
2	2	0	$\exp(2q_1 - d_1 - d_2 + t_2) / g$
3	0	1	$\exp(q_2 - e_1) / g$
4	1	1	$\exp(q_1 - d_1 + q_2 - e_1 + t_3) / g$
5	2	1	$\exp(2q_1 - d_1 - d_2 + q_2 - e_1 + t_4) / g$
6	0	2	$\exp(2q_2 - e_1 - e_2) / g$
7	1	2	$\exp(q_1 - d_1 + 2q_2 - e_1 - e_2) / g$
8	2	2	$\exp(2q_1 - d_1 - d_2 + 2q_2 - e_1 - e_2) / g$

g : Sum of the numerators

To see why d_1 and d_2 can be viewed as the first and the second step difficulties on dimension 1, consider the following log-odds. Let $P(i, j)$ denote the joint probability of a score of *i* in dimension 1 and a score of *j* in dimension 2. Then, form the probabilities shown in Table 1, we have

$$\log\left(\frac{P(1,0)}{P(0,0)}\right) = \theta_1 - \delta_1 + \tau_1$$

$$\log\left(\frac{P(2,0)}{P(1,0)}\right) = \theta_1 - \delta_2 + \tau_2 - \tau_1$$

$$\log\left(\frac{P(1,1)}{P(0,1)}\right) = \theta_1 - \delta_1 + \tau_3$$

$$\log\left(\frac{P(2,1)}{P(1,1)}\right) = \theta_1 - \delta_2 + \tau_4 - \tau_3$$

$$\log\left(\frac{P(1,2)}{P(0,2)}\right) = \theta_1 - \delta_1$$

$$\log\left(\frac{P(2,2)}{P(1,2)}\right) = \theta_1 - \delta_2$$

The parameters pertain to dimension 2, q_2, e_1, e_1, disappear. Moreover, d_1 always attaches to the log-odds of the probability of scoring 1 over the probability of scoring 0, on the fist dimension; d_2 always attaches to the log-odds of the probability of scoring 2 over the probability of scoring 1, on the fist dimension. t_1, t_2, t_3, t_4, represent adjustments to these two step difficulties.

Likewise, e_1 and e_2 can be viewed as the first and the second step difficulties on dimension 2, because:

$$\log\left(\frac{P(0,1)}{P(0,0)}\right) = \theta_2 - \varepsilon_1$$

$$\log\left(\frac{P(0,2)}{P(0,1)}\right) = \theta_2 - \varepsilon_2$$

$$\log\left(\frac{P(1,1)}{P(1,0)}\right) = \theta_2 - \varepsilon_1 + \tau_3 - \tau_1$$

$$\log\left(\frac{P(1,2)}{P(1,1)}\right) = \theta_2 - \varepsilon_2 - \tau_3$$

$$\log\left(\frac{P(2,1)}{P(2,0)}\right) = \theta_2 - \varepsilon_1 + \tau_4 - \tau_2$$

$$\log\left(\frac{P(2,2)}{P(2,1)}\right) = \theta_2 - \varepsilon_2 - \tau_4$$

The parameters pertain to dimension 1, q_2, d_1, d_1, disappear. In addition, e_1 always attaches to the log-odds of the probability of scoring 1 over the probability of scoring 0, on dimension 2; e_2 always attaches to the log-odds of the probability of scoring 2 over the probability of scoring 1, on dimension 2. t_1, t_2, t_3, t_4, are adjustments to the two step difficulties (in fact, to all item difficulties, d_1, d_1, e_1, and e_1).

If t_1, t_2, t_3, and t_4 are zero, the joint probability of any score vector, $P(i, j)$, will be equal to the product of the individual probabilities. In such a case, local independence holds within bundles. To test if they are zero, a reduced model where all the interaction parameters are set to zero is formed, as shown in Table 2. In Table 2, only four item parameters are estimated for the nine-response-category bundle. These four item parameters correspond to the step difficulties for dimensions 1 and 2. As the reduced model is a special case of the saturated model, the usual likelihood ratio test can be applied to compare the two models.

Table 2

Probabilities for the independence model for a nine-category bundle

Response category	Score		Probability
	Dim. 1	Dim. 2	
0	0	0	$1/g$
1	1	0	$\exp(q_1 - d_1)/ g$
2	2	0	$\exp(2q_1 - d_1 - d_2) / g$
3	0	1	$\exp(q_2 - e_1) / g$
4	1	1	$\exp(q_1 - d_1 + q_2 - e_1) / g$
5	2	1	$\exp(2q_1 - d_1 - d_2 + q_2 - e_1) / g$
6	0	2	$\exp(2q_2 - e_1 - e_2) / g$
7	1	2	$\exp(q_1 - d_1 + 2q_2 - e_1 - e_2) / g$
8	2	2	$\exp(2q_1 - d_1 - d_2 + 2q_2 - e_1 - e_2) / g$

g : Sum of the numerators

This reduced model is referred to as the *independence* model because all the interaction parameters are discarded and local independence holds within bundles. There are several models between the saturated model and the independence model. The most interesting one is the model where the interaction is identical across bundles. Specifically, t_1, t_2, t_3, and t_4 are identical for all the 15 bundles. We expect that the local dependence within the bundles is identical across all the bundles, because Likert-type scales were used and the subjects might show the same pattern of interference within the bundles across the whole tests. This model is referred to as the *uniform dependence* model.

The Multidimensional Model

We have just described three kinds of multidimensional bundle modeling. Multidimensional item response models are thus needed to estimate these parameters. A recently proposed multidimensional item response model, the multidimensional random coefficients multinomial logit model (MRCML, Adams, Wilson, & Wang, 1997), can meet this demand. Suppose a set of *D* latent traits underlies the persons' test performances and the persons' positions in the *D*-dimensional space are denoted $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_D)$. Let there be *I* items indexed *i* = 1,..., *I*, and *Ki* response categories in item *i* indexed *k* = 0, 1, ..., *Ki*. Use the vector valued random variable, $\mathbf{X}_i = (X_{i1}, \cdots, X_{iK_i})'$, where

$$X_{ik} = \begin{cases} 1 & \text{if response to item } i \text{ is in category } k, \\ 0 & \text{otherwise.} \end{cases}$$

to indicate the $K_i + 1$ possible responses to item *i*. A response in category 0 is denoted by a vector of 0s. This makes the 0 category a reference and is necessary for model identification. The choice of the 0 category as the reference is arbitrary and does not affect the generality of the model.

Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)'$ denote a vector of *p* free parameters that describe the items. Linear combinations of these are used to describe the empirical characteristics of the response categories of each item. These combinations are defined by design vectors $\mathbf{a}_{ik}$ (*i* = 1, ..., *I*; *k* = 1, ..., K_i), each of length *p*. These vectors are collected to form a design matrix $\mathbf{A} = (\mathbf{a}_{11}, \mathbf{a}_{12}, \cdots, \mathbf{a}_{1K_1}, \mathbf{a}_{21}, \cdots, \mathbf{a}_{2K_2}, \cdots, \mathbf{a}_{IK_I})$ for the whole test. An additional feature of MRCML is the introduction of a scoring vector that allows the specification of the score or performance level that is assigned to each possible item response. A response in category *k* of item *i* is scored *bikd* on dimension *d* (the scoring schema is known a priori). The scores across *D* dimensions can be collected into a column vector **b***ik* = (*bik*1,... , *bikD*)', then be collected into a scoring sub-matrix for item *i*, $\mathbf{B}_i = (\mathbf{b}_{i1}, \ldots, \mathbf{b}_{iK_i})'$, and again into a scoring matrix $\mathbf{B} = (\mathbf{B}_1', \ldots \mathbf{B}_I')'$ for the whole test. Under MRCML, the probability of a response in category *k* of item *i* is expressed as

$$f(X_{ik} = 1; \boldsymbol{\xi} \mid \boldsymbol{\theta}) = \frac{\exp(\mathbf{b}_{ik}'\boldsymbol{\theta} + \mathbf{a}_{ik}'\boldsymbol{\xi})}{\sum_{u=0}^{K_i} \exp(\mathbf{b}_{iu}'\boldsymbol{\theta} + \mathbf{a}_{iu}'\boldsymbol{\xi})}. \quad (1)$$

For a response vector **x** across all items,

$$f(\mathbf{x}, \boldsymbol{\xi} \mid \boldsymbol{\theta}) = \frac{\exp\left[\mathbf{x}'(\mathbf{B}\boldsymbol{\theta} + \mathbf{A}\boldsymbol{\xi})\right]}{\sum_{\mathbf{z}\in\Omega} \exp\left[\mathbf{z}'(\mathbf{B}\boldsymbol{\theta} + \mathbf{A}\boldsymbol{\xi})\right]}, \quad (2)$$

where **W** is the set of all possible response vectors.

This approach to the definitions of items allows a general model to be written that includes most of existing unidimensional Rasch models, such as the simple logistic model (Rasch, 1960), the linear logistic latent trait model (Fischer, 1973), the rating scale model (Andrich, 1978), the partial credit model (Masters, 1982), the ordered partitioned model (Wilson, 1992), FACETS (Linacre, 1989), and the linear partial credit model (Fischer & Pononcy, 1994). In addition, this allows the specification of a range of multidimensional models by imposing linear constraints on the item parameters, such as multidimensional forms of the rating scale model and the partial credit model, and the multidimensional bundle modeling proposed in this study.

The accompanying software *ConQuest* for MRCML was implemented with the marginal maximum likelihood estimation with Bock and Aitkin's (1981) formulation of the EM algorithm (Dempster, Laird, & Rubin, 1977). The estimation procedure has been described in Adams et al. (1997). Assume that persons with **q** vectors are sampled from a population with a multivariate density function $g(\boldsymbol{\theta};\boldsymbol{\gamma})$ and a distribution $G(\boldsymbol{\theta};\boldsymbol{\gamma})$, where **g** indicates a vector of parameters that characterize the distribution. Two alternative forms of *g* could be considered (1) a multivariate normal distribution *N*(**m**, **S**) where **m** and **S** are the population mean vector and the variance-covariance matrix, respectively, and (2) a step distribution defined on a pre-specified set of nodes. A multivariate normal distribution is assumed throughout this study.

A Simulation Study

The simulation condition mimicked the structure of the saturated model. The generating values are the parameter estimates of the saturated model. The sample size is 1210. Two hundred replications were made. The data were generated as follows:

1. A set of item parameters was specified and remained fixed for all data generation.
2. A random sample of **q** vector of appropriate size was drawn from an assumed multivariate normal population.
3. The known fixed item parameters and the random latent vectors **q** were then combined to calculate the probability of each response category in each item for each simulated examinee. These probabilities were then compared to a uniform (0, 1) random number to allocate response to specific categories.

The bias of each parameter is assessed by examining the difference between the mean values across the 200 replications and the generating value:

Bias of $$\hat{\zeta} = \sum_{1}^{200} \hat{\zeta}_i / 200 - \zeta,$$

where z is the generating value and $\hat{\zeta}_i$ is its estimate in the *i*th replication. Recovery is also assessed by examining the mean square error (*MSE*) for each parameter across the 1000 replications:

$$MSE \text{ of } \hat{\zeta} = \sum_{1}^{200} \left(\hat{\zeta}_i - \zeta\right)^2 / 200 \ .$$

and the sampling variance of the estimates:

$$SV \text{ of } \hat{\zeta} = \left(MSE \text{ of } \hat{\zeta}\right)^2 - \left(Bias \text{ of } \hat{\zeta}\right)^2$$

Table 3 lists the generating values for the parameters in the model, the mean of the recovered values across 200 replications, the difference between the mean of the recovered values and the generating value, the sampling variance in the parameter estimates, and mean square errors. The Hotelling's T^2 test was used to test the hypothesis that the expected differences between the generating and the estimated values for the parameters are all zero. The transformed *F* statistic is 1.38 with 123 and 77 degrees of freedom. The *p* value is .06, which is not significant at the .05 level. We accept a hypothesis of no bias in the parameter estimates. The range of the differences between the mean of the recovered values and the generating values for the 123 parameters is between -.0803 and .083. It is not very small because the range of the some generating values is very large, between -4.43 and 2.90. In sum, the software *ConQuest* produces satisfactory parameter recovery.

Table 3

Parameter recovery for the saturated model

Bundle	Parameter	Generating	Average	Difference	*SV*	*MSE*
	σ_1^2	1.08	1.0848	0.0048	0.0042	0.0042
	σ_{12}	-0.06	-0.0609	-0.0009	0.0026	0.0026
	σ_2^2	1.57	1.5691	-0.0009	0.0065	0.0065
1	d_1	-2.44	-2.4509	-0.0109	0.0598	0.0599
	d_2	-0.69	-0.7019	-0.0119	0.0196	0.0197
	e_1	-1.32	-1.3082	0.0118	0.1048	0.1049
	e_2	1.10	1.0893	-0.0107	0.0907	0.0909
	t_1	-0.80	-0.8155	-0.0155	0.1652	0.1655
	t_2	-1.74	-1.7563	-0.0163	0.1567	0.1570

	t_3	0.56	0.5896	0.0296	0.1015	0.1023
	t_4	0.11	0.1219	0.0119	0.1035	0.1037
2	d_1	-2.98	-3.0347	-0.0547	0.1460	0.1490
	d_2	-2.06	-2.0571	0.0029	0.0296	0.0296
	e_1	-0.66	-0.7183	-0.0583	0.1248	0.1282
	e_2	1.23	1.2973	0.0673	0.1635	0.1680
	t_1	-0.84	-0.8586	-0.0186	0.2739	0.2742
	t_2	-2.06	-2.0623	-0.0023	0.2318	0.2318
	t_3	0.39	0.3137	-0.0763	0.1967	0.2025
	t_4	-0.33	-0.3938	-0.0638	0.1623	0.1664
3	d_1	-2.81	-2.7979	0.0121	0.2013	0.2014
	d_2	-1.93	-1.9414	-0.0114	0.0460	0.0461
	e_1	-0.04	0.0430	0.0830	0.1591	0.1660
	e_2	1.54	1.4964	-0.0436	0.2240	0.2259
	t_1	-0.41	-0.4341	-0.0241	0.2514	0.2520
	t_2	-1.32	-1.3631	-0.0431	0.2219	0.2237
	t_3	1.00	1.0564	0.0564	0.2646	0.2678
	t_4	0.32	0.3621	0.0421	0.2207	0.2224
4	d_1	-4.43	-4.5386	-0.0586	0.2499	0.2617
	d_2	-1.92	-1.9160	0.0040	0.0302	0.0303
	e_1	0.42	0.4896	0.0696	0.1903	0.1951
	e_2	2.55	2.5507	0.0007	0.2015	0.2015
	t_1	-1.85	-1.9282	-0.0782	0.3419	0.3481
	t_2	-2.63	-2.6917	-0.0617	0.2994	0.3032
	t_3	-0.29	-0.3026	-0.0126	0.2061	0.2062
	t_4	-1.07	-1.0648	0.0052	0.1876	0.1876

5	d_1	-3.13	-3.1527	-0.0227	0.1917	0.1922
	d_2	-1.38	-1.4004	-0.0204	0.0333	0.0337
	e_1	-0.59	-0.5551	0.0349	0.0701	0.0713
	e_2	2.84	2.8584	0.0184	0.1819	0.1822
	t_1	-1.03	-1.0524	-0.0224	0.2442	0.2447
	t_2	-2.22	-2.2661	-0.0461	0.2280	0.2302
	t_3	-0.24	-0.2347	0.0053	0.2058	0.2058
	t_4	-1.29	-1.3092	-0.0192	0.1975	0.1979
6	d_1	-2.17	-2.2061	-0.0361	0.1121	0.1134
	d_2	-1.66	-1.6612	-0.0012	0.0365	0.0365
	e_1	-0.27	-0.3122	-0.0422	0.1402	0.1420
	e_2	0.98	1.0243	0.0443	0.1525	0.1545
	t_1	0.45	0.4609	0.0109	0.1984	0.1986
	t_2	-0.37	-0.3708	-0.0008	0.1750	0.1750
	t_3	1.41	1.3832	-0.0268	0.1648	0.1655
	t_4	0.85	0.8112	-0.0388	0.1505	0.1521
7	d_1	-1.83	-1.9008	-0.0708	0.1148	0.1198
	d_2	-0.84	-0.8292	0.0108	0.0369	0.0370
	e_1	-0.79	-0.7826	0.0074	0.0228	0.0229
	e_2	2.90	2.9577	0.0577	0.1068	0.1101
	t_1	-0.84	-0.9070	-0.0670	0.1408	0.1453
	t_2	-2.38	-2.4556	-0.0756	0.1371	0.1428
	t_3	-0.32	-0.3775	-0.0575	0.1265	0.1298
	t_4	-1.68	-1.7243	-0.0443	0.1182	0.1201
8	d_1	-2.78	-2.7900	-0.0100	0.1512	0.1513
	d_2	-1.30	-1.3215	-0.0215	0.0323	0.0327

	e_1	-0.92	-0.9714	-0.0514	0.1041	0.1067
	e_2	2.32	2.3410	0.0210	0.1679	0.1683
	t_1	-0.60	-0.5587	0.0413	0.2465	0.2482
	t_2	-1.93	-1.9136	0.0164	0.2384	0.2386
	t_3	0.41	0.4056	-0.0044	0.1758	0.1759
	t_4	-0.66	-0.6923	-0.0323	0.1717	0.1727
9	d_1	-3.15	-3.2138	-0.0638	0.1929	0.1970
	d_2	-2.39	-2.4000	-0.0100	0.0262	0.0263
	e_1	-1.06	-1.0665	-0.0065	0.1728	0.1728
	e_2	1.54	1.5799	0.0399	0.2497	0.2513
	t_1	-0.88	-0.9106	-0.0306	0.3793	0.3803
	t_2	-2.07	-2.1133	-0.0433	0.3378	0.3397
	t_3	0.04	0.0097	-0.0303	0.2675	0.2684
	t_4	-0.63	-0.6677	-0.0377	0.2506	0.2521
10	d_1	-3.10	-3.1516	-0.0516	0.1331	0.1357
	d_2	-2.22	-2.2090	0.0110	0.0257	0.0258
	e_1	-0.79	-0.8106	-0.0206	0.1441	0.1445
	e_2	1.17	1.1973	0.0273	0.1662	0.1669
	t_1	-1.08	-1.1117	-0.0317	0.2408	0.2418
	t_2	-2.69	-2.7125	-0.0225	0.2143	0.2148
	t_3	0.27	0.2376	-0.0324	0.1775	0.1785
	t_4	-0.95	-0.9721	-0.0221	0.1594	0.1598
11	d_1	-2.31	-2.3297	-0.0197	0.1008	0.1012
	d_2	-1.49	-1.4936	-0.0036	0.0222	0.0223
	e_1	-0.33	-0.3481	-0.0181	0.0510	0.0513
	e_2	1.78	1.8244	0.0444	0.0977	0.0997

	t_1	-0.49	-0.5139	-0.0239	0.1363	0.1369
	t_2	-2.12	-2.1468	-0.0268	0.1353	0.1360
	t_3	0.31	0.2832	-0.0268	0.1197	0.1205
	t_4	-0.90	-0.9369	-0.0369	0.1132	0.1146
12	d_1	-2.76	-2.8399	-0.0799	0.1690	0.1754
	d_2	-2.10	-2.1226	-0.0226	0.0257	0.0262
	e_1	-0.74	-0.7228	0.0172	0.1124	0.1127
	e_2	1.48	1.5483	0.0683	0.1684	0.1730
	t_1	-0.81	-0.8616	-0.0516	0.2062	0.2089
	t_2	-2.20	-2.2779	-0.0779	0.2010	0.2071
	t_3	0.21	0.1619	-0.0481	0.1909	0.1932
	t_4	-0.63	-0.6940	-0.0640	0.1701	0.1742
13	d_1	-2.36	-2.3829	-0.0229	0.0917	0.0923
	d_2	-1.11	-1.1185	-0.0085	0.0226	0.0227
	e_1	-0.50	-0.5069	-0.0069	0.0317	0.0318
	e_2	2.32	2.3520	0.0320	0.0916	0.0926
	t_1	-1.01	-1.0206	-0.0106	0.1201	0.1202
	t_2	-2.87	-2.8855	-0.0155	0.1242	0.1244
	t_3	-0.08	-0.0978	-0.0178	0.1061	0.1064
	t_4	-1.54	-1.5728	-0.0328	0.0943	0.0954
14	d_1	-1.05	-1.0515	-0.0015	0.0391	0.0391
	d_2	-0.43	-0.4278	0.0022	0.0209	0.0209
	e_1	-0.56	-0.5568	0.0032	0.0194	0.0194
	e_2	2.01	2.0068	-0.0032	0.0414	0.0415
	t_1	-0.84	-0.8408	-0.0008	0.0522	0.0522
	t_2	-2.32	-2.3288	-0.0088	0.0605	0.0606

	t_3	0.22	0.2214	0.0014	0.0471	0.0471
	t_4	-1.04	-1.0305	0.0095	0.0520	0.0521
15	d_1	-2.82	-2.8913	-0.0713	0.1251	0.1302
	d_2	-1.38	-1.3764	0.0036	0.0257	0.0257
	e_1	-0.84	-0.8139	0.0261	0.0694	0.0701
	e_2	1.97	2.0093	0.0393	0.1368	0.1384
	t_1	-1.11	-1.1850	-0.0750	0.1723	0.1779
	t_2	-2.00	-2.0615	-0.0615	0.1757	0.1795
	t_3	-0.19	-0.2193	-0.0293	0.1347	0.1356
	t_4	-0.79	-0.8326	-0.0426	0.1457	0.1475

An Empirical Example

Twelve hundreds and ten adults in Taiwan were administrated the inventory. It contains 15 personal characters and two scales: importance and possession, as stated in the above section. The original responses to the two scales were reorganized into bundles. Each of the 15 bundles contains nine response categories. The new data set was analyzed with the saturated model. Altogether, 123 parameters were estimated, including 3 parameters for the variance-covariance matrix of the person distribution, and 15 sets of 8 item parameters for the 15 bundles. The means of the person distribution were set to zero for model identification. The parameter estimates are shown in Table 3 as the generating values. The correlation between the two dimensions is -.05, meaning that these two dimensions are almost unrelated. The saturated model has a likelihood deviance (-2×loglikelihood, G^2) of 55135.59.

The uniform dependence model, where the adjustment parameters across the bundles are set to be identical, was formed. Sixty-seven parameters were estimated, including 3 parameters for the variance-variance matrix of the person distribution, and 15 sets of 4 item parameters for the 15 bundles, and 4 interaction parameters. The variances for the two dimensions are 1.08 and 1.56, respectively. The covariance is -.05. The correlation is -.04. They are almost identical to those in the saturated model. The uniform dependence model has a likelihood deviance of 55240.23.

When the independence model, where all the adjustment parameters are zero, was applied, only 63 parameters were estimated, including 3 parameters for the variance-variance matrix of the person distribution, and 15 sets of 4 item parameters for the 15 bundles. The variances for the two dimensions are 1.11 and 1.57, respectively. The covariance is .41. The correlation is .31, meaning that these two dimensions are moderately correlated. The two variances are quite close to those in the saturated model and the uniform dependence model. However, the covariance and the correlation are very different from those in the two other models. The independence model has a likelihood deviance of 55788.74. Table 4 lists the parameter estimates from the independence model. They are very different from those derived from the saturated model, shown in Table 3 as the generating values.

Table 4

Item parameter estimates from the independence model

Bundle	d_1	d_2	e_1	e_2
1	-2.68	-0.30	-2.77	1.38
2	-2.99	-1.45	-2.17	1.05
3	-3.22	-1.29	-1.56	2.01
4	-3.48	-1.26	-1.09	1.67
5	-2.66	-0.45	-1.39	1.93
6	-3.05	-1.10	-1.36	1.94
7	-1.38	0.34	-1.25	1.96
8	-2.92	-0.34	-1.99	2.05
9	-3.00	-1.82	-2.32	1.02
10	-3.02	-1.34	-2.29	0.49
11	-2.31	-0.43	-1.24	1.30
12	-2.68	-1.38	-2.05	1.03
13	-2.00	0.11	-1.42	1.45
14	-0.94	0.62	-1.34	1.70
15	-2.43	-0.86	-1.85	1.39

The two tests were also analyzed with the partial credit model, one test at a time. For the importance test, 31 parameters are estimated, including 1 for the variance of the person population distribution (the mean is constrained for model identification), and 15 sets of 2 step difficulties. The variance is 1.11 and the likelihood deviance is 26216.67. Likewise, for the possession test, 31 parameters were also estimated. The variance and the likelihood deviance are 1.57 and 29648.95, respectively. These two likelihood deviances are summed. The total deviance is 55865.62, with 62 parameters.

According to the likelihood ratio test, the difference of the two likelihood deviances for the summed model and the independence model is 76.88, with 1 degree of freedom and a *p* value of 1.82E-10. The independence model yields a better fit than the summed model. The increment of fit is due to the correlation between the two tests. Likewise, the uniform dependence model fits the data better than the independence mode (p = 1.46E-118), meaning that adding the four interaction parameters improves the fit very significantly. The saturated model fits the data better than the uniform dependence model (p = 8.80E-5), indicating that the interaction is not identical across the 15 bundles. In sum, the saturated fit the data best. Figure 1 summarizes the likelihood ratio tests.

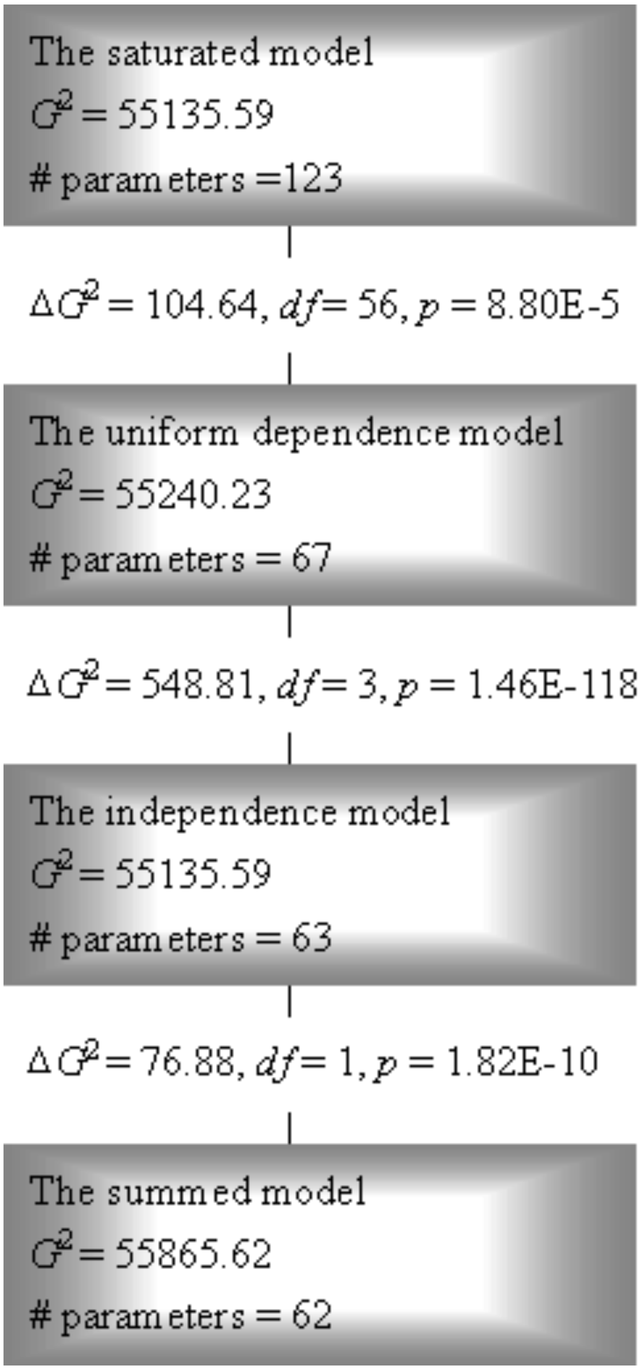

Figure 1.

The likelihood ratio tests for the four models

The results from the independence model leads us to claim that the correlation between the two dimensions is moderately positive (r = .31), meaning that the more people possess the more they consider it important. On the contrary, the results from the saturated model (or the uniform dependence model) show that the two dimensions are unrelated (r = -.05), in other words, the degrees of possession have nothing to do with importance judgment. The conclusions are so different because once the local dependence within the bundles is taken into account as in the saturated model, the true relationship emerges.

One might argue that this is because of some sort of identification problem. The identification problem was solved according to the theoretical guidelines proposed by Volodin and Adams (1995). They showed that if D is the number of latent dimensions, p is the length of the parameter vector, $\mathbf{x}$, K_i +1 is the number of response categories for item i,

and $$K = \sum_{i=1}^{I} K_i$$, then Equation (2) can be identified if and only if rank ([**B**|**A**]) = p + D £ K. In our saturated model, D is equal to 2, P is 118, K is 120, and the rank ([**B**|**A**]) is 120. Therefore, the saturated model can be identified. In addition, both the uniform dependence model and the independence model can also be identified.

To further convince the reader, we performed two data analyses. A data set generated according to the saturated model, selected from the above simulation study, was analyzed with the independence model. The two variances are 1.01 and 1.41 for the dimensions, respectively. The covariance is .39 and the correlation is .32. This variance-covariance matrix is quite close to that found in the read data with the independence model. In other words, we find a positive correlation with the independence model when in fact it is around zero.

We generated another data set according to the independence model. The generating values are those found in the real data set with the independence model. This data set was analyzed both with the saturated model and the true model, the independence model. The variances are 1.08 and 1.47 for the two dimensions, respectively, for the saturated model. The covariance is .41 and the correlation is .33. It has a likelihood deviance of 56499.43. With the independence model, the variances are 1.09 and 1.47 for the two dimensions, respectively. The covariance is .43 and the correlation is .34. It has a likelihood deviance of 56559.41. Both the variance-covariance matrices for the two models are very close to the generating matrix, within sampling fluctuation. According to the likelihood ratio test, these two models are not significantly different, with a *p*value of .48. In sum, the person distribution can be accurately recovered even when the data set has no local dependence within bundles and analyzed as if it has.

Conclusion

This paper addresses multidimensional item response modeling for local dependence between dimensions that arises from the parallel design where a common stimulus was judged with two scales. Each scale is treated as a single dimension. These two responses to the stimulus are reorganized into a bundle. The bundle becomes multidimensional because it combines two kinds of ratings. The bundle is modeled with two sets of item parameters pertaining to the two dimensions, respectively, and a set of interaction parameters representing the local dependence within the bundle. If these interaction parameters are statistically equal to zero, no local dependence within the bundles is found. Otherwise, local dependence exists within the bundles.

Through the simulation study, we find that the parameters of the proposed modeling can be unbiasedly and accurately recovered. The real data analysis shows that local dependence exhibits within the bundles. If it is not analyzed as such, not only the item parameter estimates are very different from those when the local dependence is modeled, but also the person population distribution is misleading. More specifically, when the local dependence is ignored, we find the subjects tend to consider the personal characters important if they possess them more. However, when it is taken into account, the relationship disappears, that is, how people judge have nothing to do with how much they have.

The local dependence detected in the real data set may be a general phenomenon when the parallel design is adopted. We do not suggest using the parallel design, even it saves papers and is convenient for subjects to respond. If one insists using it, local dependence detection procedures should be carried out. If the local dependence is found, the conventional analyses that do not take the dependence into account will give misleading results. The proposed modeling for local dependence between dimensions is not limited to the parallel design. It can be applied to whatever situation that the local dependence is suspected among dimensions.

Reference

Adams, R. J., & Wilson, M. R., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21,* 1-23.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561-573.

Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm. *Psychometrika*, *46*, 443-459.

Cheng, Y.-Y., & Wang, W.-C. (1999). *Creative thinking and its related factors for teachers with science competition awards*. Research report of the National Science Council.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39*, (Series B), 1-38.

Fischer, G. H. (1973). The linear logistic model as an instrument in educational research. *Acta Psychologica, 37*, 359-374.

Fischer, G. H., & Pononcy, I. (1994). An extension of the partial credit model with an application to the measurement of change. *Psychometrika, 59*, 177-192.

Hoskens, M., & De Boeck, P. (1997). A parameteric model for local dependence among test items. *Psychological methods, 2*, 261-277.

Linacre, J. M. (1989). *Many-facetd Rasch measurement.* Chicago: MESA Press.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Rasch, G. (1960). *Probabilistic models for some intelligent and attainment tests.* Copenhagen: Institute of Educational Research. (Expanded edition, 1980. Chicago: The University of Chicago Press.)

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49,* 425-435.

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

Volodin, N, & Adams, R. J. (1995). *Identifying and estimating a D-dimensional Rasch mode*l. Paper presented the international Objective Measurement Workshop. University of California at Berkeley, California.

Wilson, M. R. (1992). The partial order model: An extension of the partial credit model. *Applied Psychological Measurement, 16*, 309-325.

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

Wu, M., Adams, R. J., & Wilson. M. R. (1998). *ConQuest*. Camberwell, Victoria: Australian Council for Educational Research.

Appendix

Judge the following personal character in terms of (a) how important to creativity development, and (b) how much you possess.

Importance Possession

very not very very not very

little much much little much much

1. Multiple positive personalities □ □ □ □ □ □
2. High self-motivation □ □ □ □ □ □
3. Cognitive skills that are related to creativity □ □ □ □ □ □
4. Tendency to take risk □ □ □ □ □ □
5. Abundant professional experiences □ □ □ □ □ □
6. Rich experiences □ □ □ □ □ □
7. Good social skills □ □ □ □ □ □
8. Brilliant □ □ □ □ □ □
9. Aptitudes that are not restricted

   by bias and old methods □ □ □ □ □ □

10. Abundant interests □ □ □ □ □ □
11. Easily attracted by complexity □ □ □ □ □ □
12. Sensitive instinct □ □ □ □ □ □
13. Highly aesthetical □ □ □ □ □ □
14. Tolerance to ambiguous conditions □ □ □ □ □ □
15. Strong confidence □ □ □ □ □ □