EXPLORING LOCAL ITEM DEPENDENCY FOR ITEMS CLUSTERED AROUND COMMON READING PASSAGE IN PIRLS DATA

Valérie Quittre, University of Liege, Belgium, v.quittre@ulg.ac.be
Christian Monseur, University of Liege, Belgium, cmonseur@ulg.ac.be

Abstract

Since the IEA 1991 Reading Literacy Study (Elley, 1994; Wolf, 1995), cognitive data of international assessments are usually scaled according to IRT models. One of the cornerstones of IRT models is the assumption of Local Item Independence (LII).

In PIRLS assessment, as well as in other international assessments in reading literacy such as PISA (Organisation for Economic Cooperation and Development - OECD), testing material is hierarchically structured, i.e. several items are related to a common passage. Monseur & al (submitted) noted that "*this test format may be viewed as the most appropriated to assess a complex process such as reading comprehension. In real life situation, students have to use different cognitive processes to understand various components of a same text*". This embedded structure might therefore violate the assumption of LII. As stated by Embretson and Reise (2000, p. 188), "*Practically, local independence is violated when item responses are linked*".

As noted by Balazs and De Boeck (2006, p. 2) "*ignoring Local Item Dependence can have serious consequences for the goodness of fit of a model, for the parameter estimates and for confidence intervals.*"

This study was aimed at detecting passage-related local item dependencies in PIRLS 2006 cognitive data and estimating its relationship with the country performance.

Keywords: Local Item Dependency, IRT, PIRLS

Introduction

Local Item Independence is a strong assumption in IRT models (Lord and Norwick, 1968, Embretson and Reise, 2000). In such models, the probabilities that one person will provide a specific response to an item rely on his/her ability θ, and one or more parameters (depending on the IRT model) describing the relationship of the item to θ.

For instance, according to the one parameter Rasch model (Rasch, 1960), the probability that a person *i* succeed an item *j*, given the person ability θ_i and the item difficulty δ_j, is equal to:

$$P(X_{ij} = 1|\theta_i, \delta_j) = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)}$$

The likelihood for an item response pattern is then computed by multiplying the probabilities of the individual events:

$$P(X_{ij1} = x_1 \text{ and } X_{ij2} = x_2|\theta_i) = P(X_{ij1} = x_1|\theta_i)P(X_{ij2} = x_2|\theta_i)$$

As shown by the formulae, the joint probability is equal to the product of individual probabilities. It therefore requires the independence between individual events.

Precisely, local item independency means that "*the response to any item is unrelated to any other item when trait level is controlled. The items may be highly intercorrelated in the whole sample; however, if trait level is controlled, local independence implies that no relationships remain between the items*" (Embretson and Reise, 2000, p. 188). In other words, the success or the failure to any item does not affect the probability of succeeding to the other items.

The violation of LII assumption can have substantial consequences on test parameter estimates and on proficiency estimates. Balazs & De Boeck (2006) for instance noted that the IRT models are "*not robust to the violation of LII*".

Tests organised in units around a common stimulus can produce two types of stimulus-related LID (Yen, 1993): a local dependence resulting of an unusual level of interest or prior knowledge about the stimulus and a local dependence produced by the fact that information used to answer different items in the unit is interrelated in the stimulus.

Monseur & al investigated stimulus-related LID in PISA assessments. They first observed an apparent LID in reading and in Mathematics. A thorough analysis showed that in Reading, this LID was essentially the result of speediness. Indeed, non-reached items can artificially create a LID because a student not reaching item j will surely not reach item j+1. It therefore created a correlation between

residuals. Once not reached items were treated as missing, PISA reading passages, except one, did not present substantial LID. The qualitative analysis of this unit brought to light obvious interrelated information into play in different items of the unit. In summary, PISA 2000 reading passages do not seem to influence the success or the failure to the different items within the unit.

On the contrary, in Mathematics, several 2003 units present a substantial LID even after the not-reached items were considered as missing. This dependence "*could result from similarities in the cognitive processes involved in several items or from the specificity of the context. In this particular domain, prior knowledge on the stimulus or the interrelation of the information required to answer different items are certainly major sources of LID.*" (Monseur & al., submitted)

Further, these authors showed a relationship between the importance of LID and the country performance. In reading like in mathematics, low achieving countries showed on average a higher LID than countries with higher performances, leading to potential bias on the estimates of some equity indicators such as the performance variability.

This study investigated passage related LID on PIRLS 2006 assessment. It is of particular importance as any particular rotated booklet consists of only 2 passages (compared to PISA 2000 where any booklet contains from 10 to 13 reading units) and each of these 2 passages are followed by a large number of items. Table 1 presents the number of items by unit for these two assessments.

[Take in Table 1 about here]

Methodology

In the present study, the Local Item Dependence was detecting by using the Yen's Q_3 statistic based on the residuals correlations analysis. Analysing the residuals is a way to control for the student proficiency, since the residuals are the differences between the individuals observed scores and their respective predicted scores. If some sets of items do present a significant level of residuals correlations, those items can be considered as locally dependent, since a link between them remains after controlling for the students' abilities (Yen, 1993).

Residual item correlations were computed for each pair of items within a unit and a Q_3 matrix was obtained for each unit (for one k-item unit, there are k(k-1)/2 Q_3 values in the matrix) . Yen (1993) summarizes the unit Q_3 values by two key values: the median and maximum values in the unit's matrix. Monseur & al suggest a LID coefficient derived from a principal component analysis on all residuals of a particular unit. The variance of the first factor is transformed to obtain a coefficient that ranges from 0 to 1 according to the following formula, where σ_1^2 represents the variance of the first factor:

$$Coef_{dep} = \frac{\frac{\sigma_1^2}{nb_{items}} - \frac{1}{nb_{items}}}{1 - \frac{1}{nb_{items}}} = \frac{\sigma_1^2 - 1}{nb_{items} - 1}$$

This LID coefficient is mathematically equal to the percentage of variance explained by the first factor, minus the percentage of variance that would be explained by the first factor if all the input variables used in the principal component analysis were uncorrelated. Let us assume that 4 variables are submitted to a principal component analysis and the eigenvalue of the first factor is 2.4. The first factor explains 60 % of the total variance. However, the percentage of explained variance with uncorrelated input variables would be equal to 25 %. The coefficient of dependency is therefore equal to 0.60-0.25=0.35 (Monseur & al, submitted).

The analyses were performed on the 23 European countries (or subnational entities) participating in PIRLS. 2006.[1] The data were scaled with the Partial Credit Model as described by Wright and Masters (1982). The scaling was implemented using the ConQuest software (Wu & al., 1997).

The item parameters were estimated on a dataset that consists of 500 students randomly selected in the 23 PIRLS national samples.

The two units from the Reader book (R021U and R021S) were not included in the analyses as they are not linked with the testing material of the others booklets.

Non-reached items were considered as missing answers to make certain that the results are not affected by speediness.

Three student proficiency estimates were computed: (Maximum Likelihood Estimate - MLE). The first estimate (denoted θ_1) is based on the test material of the two texts while the second and the third estimates are respectively based on the first and the second texts. Indeed, preliminary analyses on θ_1 reveal no substantial LID in the PIRLS 2006 testing material. As the student response pattern to one unit contributes to about 50 percent of the student proficiency estimate, it was hypothesized that this redundancy might hide potential LID. The student proficiency estimate based on the first reading passage was therefore used to detect LID among items related to the second reading passage and the reverse.

Findings and Discussion

[1] Austria, Belgium (Flemish), Belgium (French), Bulgaria, Denmark, England, France, Germany, Hungary, Iceland, Italy, Lithuania, Luxembourg, Latvia, Nederland, Norway, Poland, Romania, Scotland, Slovak Republic, Slovenia, Spain and Sweden.

The LID coefficients and the Median Q_3 values are indicators of the dependence at the unit level. They reveal the global context dependency due to the common passage. The data presented in Table 2 show that this type of passage-related LID is on average minor for all PIRLS 2006 units, the most important LID coefficient being only 0.13 (on average for the 23 European participating countries). These results are consistent with the results of Monseur & al. on PISA reading material. Like in PISA reading, PIRLS texts do not generate any context-related LID.

The variations between countries are weak too as shown by the standard deviations of the LID coefficients.

[Take in Table 2 about here]

However, those small LID coefficients do not imply that local item independence is assumed between any pair of items For reminder, as the average number of items per unit units is about 13, Q_3 correlations matrix are quite large with (on average 73 pairs of items).

Substantial correlations are observed for some pairs of residuals within the units. The Maximum Q_3 values[2] given in Table 2 are on average superior to 0.25 for seven out of the eight units analyzed.

Table 3 shows the Maximum Q_3 values at the country level. R011Y and R021N provide Maximum Q_3 values superior or equal to 0.30 for respectively 15 and 13 countries. On the contrary, R021E and R021K produce lower correlations between couples of items.

[Take in Table 3 about here]

The Maximum Q_3 values vary from country to country as shown by the standard deviation of this indicator as well as their minima and their maxima (table 3). Countries with low performances developed on average higher maximum residuals correlations than higher achieving countries (a moderate correlation of -0.33 which rises to -0.66 if Bulgaria is excluded[3]). The figure 1 illustrates this relationship between pair-items dependence and performance.

[Take in Figure 1 about here]

The analysis of the whole Q_3 matrix by unit and by country highlights the pairs of items with correlated residuals. Table 4 summarises for the eight units of PIRLS 2006, the pairs of items affected for a majority of the countries.

[Take in Table 4 about here]

[2] For one k-item unit, the Maximum Q_3 value is the maximum residual correlation out of the k(k-1)/2 values of the Q_3 matrix.
[3] Bulgaria presents particular high Q_3 values despite his high performance.

Table 5 below illustrates for Spain the whole Q_3 matrix of R021N, an example of unit which gives a substantial Maximum Q_3 value for most of the countries.

[Take in Table 5 about here]

For this unit, in Spain, the Maximum Q_3 value is between item 5 and item 6 (0.31). The maximum of correlation between residual 5 and residual 6 is also observed for 16 other countries. Iceland is the only country where this correlation is inferior to 0.20. Local dependence is also revealed between item 11 and item 12 (Q_3 value = 0.25 and superior than 0.20 for 17 countries).

The dependence existing between pairs of items within a unit can be illustrated in R011C, "The little lump of clay", which is a released unit[4]. Table 6 present the Q_3 matrix for England.

[Take in Table 6 about here]

For England, the correlation is maximum between residual 1 and residual 6 (Q_3 value = 0.34). This correlation is also maximum in 8 others countries and is superior to 0.20 for 17 countries. Moderate correlations are also present between residuals 1 and 3 (maximum for 4 countries and superior to 0.20 for 9 countries), between residuals 1 and 8 (superior to 0.20 for 12 countries) and between residuals 10 and 11 (maximum for 6 countries and superior to 0.20 for 20 countries). Figure 2 shows the items concerned by this dependency.

[Take in Figure 2 about here]

Item 1 assesses the comprehension of the general chronology of the story, dividing the passage in five events to be reordered. Then, questions 2 to 9 follow the chronology of the story, each of them focusing on a deep understanding of specifics features of the story. Some of those items require a good comprehension of a part of the story chronology and of the general structure of a fiction. Therefore we could reasonably consider that the relatively high correlation between item 1 and other items is due to a similarity in the kind reading competence needed to answer those items (general comprehension of the structure of a fiction).

The reason of the dependence between item 10 and item 11 appears quite obvious: both items are concerned about the transformation of the clay into a cup. Item 10 focuses on the change of feelings of the lump of clay subsequent to this conversion while item 11 is centered on the little girl's role in this positive change. Information required to correctly answer those two items is clearly interrelated, since it is related to the comprehension of a classical narrative structure (the passage from an initial situation to a denouement mediated by a positive character). Furthermore, independently of the passage, the

[4] Due to its length, the R011C passage is not presented.

similar form of item 10 and item 11 (“Explain why…”), and the same format of answer (open-ended) could additionally generate local dependency between those items.

Conclusion and Implications

The results of this research first highlight minor context-related LID in PIRLS 2006 material and confirm the outcomes of the analyses carried out on PISA 2000 reading data. The general low level of LID indicates that the selected passages for these two international surveys in reading literacy avoid the influence of unusual interest and/or prior knowledge on the passages on the probability of succeeding the different items within the units. Nevertheless, within most of the units, several pairs of items present substantial Q_3 statistic. This type of passage-related LID is probably due to the interrelation of the information required to answer the different items.

Moreover, this study shows a relationship between the Q_3 statistic and the country performance, countries with low performances develop on average higher maximum residuals correlations than higher achieving countries that could not be attributed to speediness.

It might be worth extending the analyses on all the countries that participated in PIRLS 2006. The possible bias on the estimates of country performance could therefore be more investigated.

Finally, test developers should carefully review pairs of items that present moderate or significant LID to better understand its sources so that it could be anticipated during the test development process.

References

Balazs, K., & De Boeck, P. (2006). Detecting local item dependence stemming for minor dimensions. Interuniversity Attraction Pole statistics network (technical report). from www.stat.ucl.ac.be/IAP

Elley, W. B. (1994). *The IEA Study of Reading Literacy: Achievement and instruction in thirty-two school systems*. Oxford: Pergamon Press.

Embretson, S.E. and Reise, S.P. (2000). *Item Response Theory for Psychologists*. Multivariate Applications Books Series. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 371.

Lord, F.M., & Novick, M. (1968). *Statistical theories of mental test scores*. Reading, M.A.: Addison-Wesley.

Monseur, C. (2009). Effect of Local Item Dependence of country performance estimates. PISA Research Conference. Kiel. Germany from www.pisaresconf09.org

Monseur, C., Baye, A., Lafontaine, D. & Quittre, V. (submitted). Item bundle assessment and Rasch Local Item Independence Assumption

Rasch, G. (1960). *Probalistic Models for Some Intelligence and Attainment Tests*. Copenhaguen: Denmark's Paedagogistic Institut.

Yen, W. M. (1993). Scaling performance assessments: strategies for managing local item dependence. *Journal of Educational Measurement 30*, 187-213

Wolf, R. M. (1995). *The IEA Reading Literacy Study: Technical Report*. The Hague, The Netherlands.

Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest Version 2.0. Generalised Item Response Modelling Software*. Camberwell: ACER Press.

Tables and Figures :

EXPLORING LOCAL ITEM DEPENDENCY FOR ITEMS CLUSTERED AROUND COMMON READING PASSAGE IN PIRLS DATA

Table 1: Number of items by unit in PIRLS and in PISA reading

Assessment	# of units including several items	# of units including a single item	# of items	Average # Items/Unit
PIRLS 2006	10	-	126	12.6
PISA reading 2000	34	3	129	3.5

Table 2: Local Item Dependency exprimed by LID coefficient, Median and Maximum Q_3 values for the 8 units of PIRLS 2006 (on average for the 23 European countries)

	LID coefficient				MEDIAN Q_3 Values		MAXIMUM Q_3 Values	
UNIT	MIN	MAX	MEAN	STD	MEAN	STD	MEAN	STD
R011Y	0,12	0,16	0,13	0,01	0,12	0,01	0,33	0,05
R011L	0,1	0,15	0,12	0,02	0,10	0,01	0,29	0,04
R011A	0,09	0,19	0,12	0,02	0,11	0,02	0,26	0,06
R011C	0,09	0,14	0,12	0,02	0,11	0,01	0,29	0,05
R021K	0,1	0,15	0,12	0,01	0,11	0,01	0,26	0,04
R021N	0,08	0,14	0,11	0,01	0,11	0,02	0,30	0,04
R011F	0,08	0,15	0,11	0,02	0,10	0,02	0,29	0,05
R021E	0,08	0,16	0,11	0,02	0,10	0,01	0,24	0,04

Table 3: Maximum Q_3 value by passage and by unit – PIRLS 2006[1]

	R011C	R011F	R011Y	R021E	R011A	R011L	R021N	R021K
AUT	0,25	0,23	**0,31**	0,21	0,20	**0,33**	**0,34**	0,26
BGR	**0,38**	**0,35**	**0,42**	**0,39**	0,23	**0,34**	**0,34**	**0,38**
DNK	0,25	0,29	**0,39**	0,28	0,23	**0,32**	0,29	0,24
FRA	0,21	**0,32**	0,28	0,23	0,23	**0,30**	0,25	0,26
DEU	**0,31**	0,22	**0,30**	0,18	0,20	0,23	0,29	0,23
HUN	**0,33**	0,22	0,27	0,21	0,25	0,27	0,27	0,28
ISL	**0,33**	**0,31**	**0,39**	0,24	0,19	**0,30**	**0,30**	0,23
ITA	0,26	0,26	**0,30**	0,24	0,29	0,29	**0,32**	**0,30**
LVA	0,26	0,24	**0,40**	0,25	0,29	0,25	0,24	0,27
LTU	**0,30**	0,24	**0,30**	0,25	**0,30**	0,25	0,29	0,24
LUX	**0,32**	0,22	**0,30**	0,22	**0,41**	**0,30**	0,24	0,24
NLD	**0,30**	0,29	0,27	0,27	0,25	0,24	0,23	0,21
NOR	0,24	**0,32**	**0,37**	0,23	0,28	**0,37**	**0,32**	0,26
POL	**0,30**	**0,40**	**0,33**	0,22	0,23	0,28	**0,31**	0,26
ROM	**0,37**	**0,37**	**0,38**	0,27	0,25	**0,40**	**0,38**	**0,30**
SVK	**0,32**	**0,31**	0,26	0,24	0,23	0,28	**0,32**	0,23
SVN	0,29	0,27	**0,31**	0,20	0,23	0,29	**0,31**	0,24
ESP	**0,36**	**0,33**	**0,36**	0,24	0,34	0,27	**0,31**	0,24

[1] Maximum Q_3 values are indicated in bold.

SWE	0,25	**0,30**	0,29	0,23	0,25	**0,30**	**0,33**	0,28
ENG	**0,34**	0,26	0,29	0,24	0,26	**0,30**	0,24	0,23
SCO	0,27	**0,31**	**0,42**	0,26	0,25	**0,30**	0,29	0,22
BFL	0,29	0,23	0,29	0,27	0,22	0,26	**0,32**	0,27
BFR	0,24	0,29	0,28	0,22	**0,42**	0,28	**0,30**	0,27
MIN	*0,21*	*0,22*	*0,26*	*0,18*	*0,19*	*0,23*	*0,23*	*0,21*
MAX	*0,38*	*0,40*	*0,42*	*0,39*	*0,42*	*0,40*	*0,38*	*0,38*
MEAN	*0,29*	*0,29*	**0,33**	*0,24*	*0,26*	*0,29*	**0,30**	*0,26*
STD	*0,05*	*0,05*	*0,05*	*0,04*	*0,06*	*0,04*	*0,04*	*0,04*

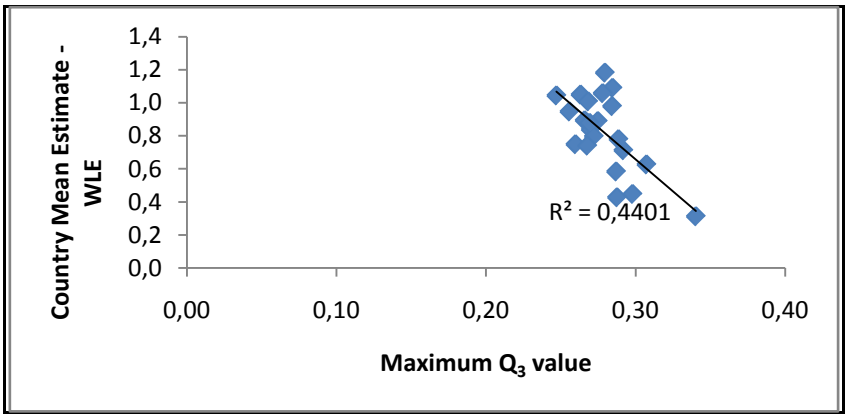

Figure 1: PIRLS 2006 country mean estimates and Maximum Q_3 values – 22 European countries

Table 4: Pairs of dependent items and number of countries concerned – PIRLS 2006

UNIT	Pair of items	Number of countries[2] for which Q_3 statistic ≥ 0.20
R011C	Item 1 – item 3	12
	Item 1 – item 8	12
	Item 1 – item 6	17
	Item 10 – item 11	20
R011F	Item 9 – item 10	12
	Item 1 – item 9	13
	Item 8- item 9	19
R011Y	Item 9- item 13	13
	Item 5 – item 6	14
	Item 13 – item 14	18
	Item 12 – item 14	20
	Item 12 – item 13	21
	Item 8 – item 9	21
R021E	/	
R011A	Item 7 – item 8	12
	Item 8 – item 9	12
R011L	Item 4 – item 10	17
	Item 4 – item 8	18
	Item 10 – item 12	18
	Item 8 – item 10	20
R021N	Item 11 – item 12	17
	Item 5 – item 6	22

[2] Except Bulgaria.

R021K	Item 10 – item 12	13
	Item 7 – item 10	15
	Item 7 – item 12	17

Table 5: Q_3 matrix of unit R021N for Spain[3]

	res1	res2	res3	res4	res5	res6	res7	res8	res9	res10	res11	res12
res1	1											
res2	0,05	1										
res3	0,05	0,17	1									
res4	0,12	0,14	*0,25*	1								
res5	0,17	0,05	0,08	0,09	1							
res6	0,07	0,01	0,10	0,17	***0,31***	1						
res7	0,10	0,09	0,04	0,10	0,02	0,02	1					
res8	0,11	0,05	0,18	0,19	0,09	0,18	0,07	1				
res9	0,08	0,06	0,10	0,10	0,11	0,03	-0,02	0,11	1			
res10	0,08	0,05	-0,01	0,09	0,03	-0,05	0,12	-0,02	0,15	1		
res11	0,10	0,19	*0,26*	0,19	0,10	0,12	0,11	0,13	0,14	0,09	1	
res12	0,08	0,14	0,17	0,08	0,05	0,10	0,05	0,09	*0,22*	0,02	***0,25***	1

Table 6: Q_3 matrix of unit R011C for England

	res1	res2	res3	res4	res5	res6	res7	res8	res9	res10	res11	res12	res13
res1	1												
res2	0,19	1											
res3	***0,29***	0,18	1										
res4	0,15	0,15	0,12	1									
res5	0,09	0,10	0,14	-0,01	1								
res6	***0,34***	0,15	0,14	0,09	0,13	1							
res7	*0,22*	0,12	0,19	0,08	0,08	0,19	1						
res8	***0,23***	0,08	0,14	0,12	0,06	0,15	0,19	1					
res9	0,11	0,06	0,09	0,09	0,07	0,05	0,07	0,07	1				
res10	0,13	*0,24*	*0,20*	0,10	0,08	*0,20*	0,14	0,16	0,15	1			
res11	0,06	0,09	0,16	0,07	0,04	0,17	0,06	0,13	0,10	***0,22***	1		
res12	0,06	0,11	0,06	0,08	0,12	0,13	0,05	0,08	0,09	0,09	0,09	1	
res13	0,14	0,15	0,11	0,18	0,00	0,14	0,12	0,20	0,07	0,12	0,12	0,05	1

[3] The numbers in italic present significant residuals correlations in Spain, the bolded numbers are significant residuals correlations for the most of the countries.

Figure 2: Some of the items of the released R011C unit, *"The Little Lump of Clay"*

1. Number the sentences below in the order the events happened in the story. Number 1 has been done for you.

___ The rain made the lump of clay moist and soft.

___ A boy tried to make the lump of clay into a bowl.

___ A girl made the lump of clay into a cup.

___ The lump of clay dried out.

1 ___ The lump of clay was in the bin.

3. At the beginning of the story, what did the lump of clay wish for?

1 _____

_____

6. The boy left the lump of clay in danger. What was the danger?

1 _____

8. What wonderful thing happened after the lump of clay had been lying by the window for a long time? Why was this so wonderful for the lump of clay?

2 _____

_____

10. Describe the different feelings the clay had at the beginning and the end of the story. Explain why his feelings changed.

3 _____

_____

11. The little girl is an important person in this story. Explain why she was important to what happened.

2 _____

_____