Chapter 16: SCALING PROCEDURES AND CONSTRUCT VALIDATION OF YOUTH INTERVIEW AND PMK CONTEXT QUESTIONNAIRE DATA

INTRODUCTION

The PISA-D Strand C Youth Interview and Person Most Knowledgeable (PMK) Questionnaire are based on the questionnaire framework, described in Chapter 3 of this report. Many of the interview and questionnaire items were designed to be combined in some way to measure latent constructs that cannot be observed directly, such as youth levels of depression or family resources. For these items, scaling procedures were applied to construct meaningful indices.

Three types of scales were developed for PISA-D Strand C:

- **Scales identical to PISA 2015.** These scales used identical sets of items as those used in PISA 2015. They enable countries to make direct comparisons of their results to those of the countries that participated in PISA 2015.
- **Scales that extended those of PISA 2015**. These scales used a subset of items from scales used in PISA 2015 as well as new items relevant to PISA-D countries.
- **Scales identical to PISA-D Strand B.** These scales used identical sets of items or a subset of those used in PISA-D Strand B. They enable countries to make direct comparisons of their results across Strand A,B and C.

This chapter describes the methodology used for scale and construct validation of the derived variables that underlies these types of scales. The measurement of the PISA index of economic, social and cultural status (ESCS), which is a scale that extends that of PISA 2015 and links to PISA-D Strand B, is described separately.

SCALING METHODOLOGY AND CONSTRUCT VALIDATION

Scaling procedures

As in PISA 2015 and in previous cycles of PISA, the derived variables for PISA-D were constructed using IRT *(item response theory)*. The IRT models used in PISA-D are subsets of the generalised partial credit model (Masters and Wright, 1997).

The responses for each item are modelled as a function of the latent construct, θ_j. With a one-parameter model, called the Rasch model, (Rasch, 1960) for dichotomous items, the probability of person *j* selecting category 1 instead of 0 is modelled as:

$$P\left(X_{ji} = 1 | \theta_j, \beta_i\right) = \frac{\exp(\theta_j - \beta_i)}{1 + \exp(\theta_j - \beta_i)}$$

where $P(X_{ji} = 1)$ is the probability of person *j* to score 1 on item *i*, θ_j is the estimated latent trait of person *j* and β_i is the estimated location or difficulty of item *i* on this dimension. In the case of items with more than two (*m*) categories (e.g. Likert-type items), this model can be generalised to the *Partial Credit Model,* which takes the form of:

$$P\left(X_{ji} = k | \theta_j, \beta_i, d_i\right) = \frac{\exp(\sum_{r=0}^{k} \theta_j - (\beta_i + d_{ir}))}{\sum_{u=0}^{m_j} exp(\sum_{r=0}^{u} \theta_j - (\beta_i + d_{ir}))}$$

where $P\left(X_{ji} = k\right)$ denotes the probability of person *j* to score *k* on item *i* out of the m_i possible scores on the item. θ_j denotes the person's latent trait, the item parameter β_i gives the general location or difficulty of the item on the latent continuum and d_{ir} denotes additional step parameters. This model has been used throughout previous cycles of PISA for scaling derived variables of the context questionnaires. However, research literature (especially, Glass and Jehangir, 2014) suggests that a generalisation of this model, the *generalised partial credit model* (GPCM) (Muraki, 1992), is more appropriate in the context of PISA since it allows for the item discrimination to vary between items within any given scale. This model takes the form of:

$$P\left(X_{ji} = k | \theta_j, \beta_i, d_i\right) = \frac{\exp(\sum_{r=0}^{k} \alpha_i(\theta_j - (\beta_i + d_{ir})))}{\sum_{u=0}^{m_j} exp(\sum_{r=0}^{u} \alpha_i(\theta_j - (\beta_i + d_{ir})))}$$

in which the additional discrimination parameter α_i allows for the items of a scale to contribute with different weights to the measurement of the latent construct. This later model was used for the scaling of the PISA-D Strand C scales.

Values for the β_i, α_i, and d_{ir} parameters for each scale are presented throughout this chapter.

Construct Validation

We assessed the cross-country validity of measures of student background, practices, attitudes and perceptions in PISA-D following two approaches implemented for context questionnaires in PISA 2015.

Internal consistency

Cronbach's alpha assesses the internal consistency of each scale within countries and compares it between countries. The alpha coefficient ranges between 0 and 1, with higher values indicating higher internal consistency. Commonly accepted cut-off values are 0.9 to signify excellent, 0.8 for good and 0.7 for acceptable internal consistency. Following the approach used in PISA 2015, in PISA-D Cronbach's alpha assessed internal consistency by scale and by country.

Cross-country comparability

Cross-country validity of the constructs assumes that the same constructs can be measured consistently in different national and cultural contexts. All of the scales and indicators in PISA-D

are based on persons' self-reports. Such measures can suffer from various measurement errors, stemming from retrospective reports of behaviour and cultural differences in attitudes towards self-enhancement, impacting examinees' self-reported beliefs, behaviours and attitudes (Bempechat, Jimenez and Boulay, 2002). The literature consistently shows that response biases, such as social desirability, acquiescence and extreme response choice, are more common in contexts with lower socio-economic development and socio-economic status, and that response styles also differ between genders (Buckley, 2009).

Invariance of item parameters

Following the approach pioneered in PISA 2015, we estimated international item and person parameters based on all examinees across all seven countries. This estimation produced the root mean square deviance (RMSD) item-fit statistic for each group and item as:

$$RMSD = \sqrt{\left(P_o(e) - P_e(e)\right)^2 f(e)\, de}$$

quantifying the difference between the *observed item characteristic curve* (*ICC*, $P_o(q)$) with the *model-based ICC* ($P_e(q)$)) (OECD, 2017, p. 269).

This statistic allows quantifying the extent of the discrepancy between the observed item characteristic curve (ICC) and the model-based ICC. The RMSD is sensitive to the group-specific deviations of both the item difficulty parameters and item slope parameters from the international parameters. Values close to zero indicate good item fit, meaning that the model with international item parameters describes the responses in this group very well. A value of RMSD=0.3 was set as a cut-off criterion, with larger values indicating that the international item parameters are not appropriate for this group. In cases of large RMSD statistics, we allowed a group to receive unique parameters. In this document, we report the final distribution of RMSD values across groups for each item.

Table 16.1   RMSD values by Country

Construct	GTM	HND	PAN	PRY	SEN	WRMSD
DEPRESSION	0.072	0.048	0.058	0.061	0.055	0.061
FAMRES	0.124	0.172	0.147	0.135	0.141	0.156

Evaluating cross-country comparability of latent constructs

PISA 2015 adopted a new approach to evaluating the invariance of latent constructs across groups. The RMSD statistic quantifies how well the international parameters describe a group's observed data and its distribution across groups indicates the international item parameters' fit, i.e. how well the international item parameters function across groups. In the following pages, we report in tabular for the distribution of RMSD country scores, scale by scale. The theoretical minimum (RMSD=0.0) indicates perfect fit of the international item parameters for this group.

A value of RMSD=0.3 was chosen as the cut-off criterion for assigning group-specific parameters, thus indicating the possible maximum of RMSD.

SCALES IDENTICAL TO PISA 2015

For the scales identical to PISA 2015, the international item and person parameters were originally obtained from a calibration process based on a GPCM (see PISA 2015 technical report) for a single analysis based on data from all persons in all countries.

As in PISA 2015, also in PISA-D, for each scale, only persons with a minimum number of three valid responses were included. Youth from the seven countries of PISA-D were unweighted. The calibration of the PISA-D item and person parameters anchored the values of the 2015 parameters using the test analysis modules (TAM) package (Wu et al, 1998) in the R program. The TAM package produced weighted likelihood estimates (WLE; Warm, 1989) as individual participant scores. Additional analyses on the invariance of item parameters across the seven countries assessed the necessity of defining unique parameters for some of the countries (see section "Cross-country comparability" in this chapter). The WLEs obtained with the TAM package were rescaled with a linear transformation to link them to the PISA 2015 scale by subtracting from the PISA-D original WLE scores the PISA 2015 unweighted WLE mean and dividing them by the PISA 2015 unweighted WLE standard deviation.

Table 16.2   Scales identical to PISA 2015

DV Name	Description	Source	Question numbers
MISCED	Mother's education (ISCED)	Youth Interview	Y055cQ01TA; Y056cQ01TA to Y056cQ05TA
HISCED	Highest education of parents (ISCED)	Youth Interview	
FISCED	Father's education (ISCED)	Youth Interview	Y062cQ01TA; Y063cQ01TA to Y063cQ05TA
PARED	Highest education of parents in years	Youth Interview	
BFMJ2	ISEI of father	Youth Interview	Y059CQ01TA and Y259CQ01TA
BMMJ1	ISEI of mother	Youth Interview	Y052CQ01TA and Y252CQ01TA
HISEI	Index highest parental occupational status	Youth Interview	
BYMJ3	ISEI of youth	Youth Interview	Y075BQ01NA and Y275BQ01NA
PMKSEI	ISEI of PMK	PMK questionnaire	HH003 and HH004
PMKISCED	PMK's education (ISCED)	PMK questionnaire	HH005
PMKEDU	Education of PMK in years	PMK questionnaire	

Educational level of parents (MISCED and FISCED; PARED, PMKEDU)

Youth responses on questions Y055cQ01TA, Y056cQ01TA to Y056cQ05TA, Y062cQ01TA and Y063cQ01TA to Y063cQ05TA regarding parental education were classified using ISCED 1997 (OECD, 1999). Three indices on parental education resulted from the recoding of educational qualifications into the following categories: (0) None; (1) ISCED 1 (primary education); (2) ISCED 2 (lower secondary); (3) ISCED Level 3B or 3C (vocational/pre-vocational upper secondary); (4) ISCED 3A (general upper secondary) and/or ISCED 4 (non-tertiary post-secondary); (5) ISCED 5B (vocational tertiary); and (6) ISCED 5A and/or ISCED 6 (theoretically oriented tertiary and post-graduate). The index MISCED indicates the educational level of the mother; FISCED indicates that of the father; and HISCED indicates the highest educational level of parents, corresponding to the higher ISCED level of either parent. PARED is the index of the estimated number of years of education generated from HISCED. In PISA-D the mapping of ISCED levels to years of schooling (PARED) was done in consultation with the participating countries, taking into account changes in countries' educational systems. Similarly, PMKEDU was computed based on the response of the PMK to question HH005.

Highest occupational status of parents

In PISA-D, youth were asked open ended questions on the occupations of their mothers and fathers. The responses were coded to four-digit ISCO codes and then mapped to the international socio-economic index of occupational status (ISEI) (Ganzeboom and Treiman, 2003). In PISA-D we adopted the new ISCO and ISEI in their 2008 version. Based on this information, we computed three indexes: father's occupational status (BFMJ2); mother's occupational status (BMMJ1); and the highest occupational status of parents (HISEI) which corresponds to the higher ISEI score of either parent or to the only available parent's ISEI score. For all three indices, higher ISEI scores indicate higher levels of occupational status.

SCALES THAT EXTENDED THOSE OF PISA 2015

For those scales administered in PISA 2015 and then extended with new additional items in PISA-D, the scale scores in PISA-D allow linking with those reported in PISA 2015 because of a common calibration linking procedure consisting of two phases: joint calibration and linking transformation.

The joint calibration phase produced international item and person parameters using a generalised partial credit model (see equation 2) in a single analysis and based on joint data, composed by a 5% simple random sample of unweighted persons in all countries from PISA 2015 and all unweighted persons in all countries from PISA-D. We fixed the values of the common items to the PISA 2015 parameters and let the TAM package generate the parameters of the new items. For each scale, only persons with a minimum number of three valid responses were included. We conducted additional analyses on the invariance of item parameters across PISA-D countries and considered assigning unique parameters if necessary (see section "Cross-country comparability" in this chapter). From this concurrent calibration we derived WLEs for the 5% sample from PISA 2015 and for PISA-D.

In the linking phase, the PISA-D WLEs obtained in the calibration phase were initially standardised and then linked to the 2015 metric by a linear transformation, subtracting the PISA 2015 mean and dividing by the PISA 2015 standard deviation, both obtained from data from all persons in 2015.

Table 16.3   Scales Extended from PISA 2015

DV Name	Description	Variables	IRT scaling
HOMEPOS15	Home possessions		YES
ESCS	Index of economic, social and cultural status	HOMEPOS15, PARED, HISEI	

Household possessions

In PISA 2015, youth reported the availability of household items at home (ST011) including three country-specific household items that were seen as appropriate measures of family wealth within the country's context. In addition, youth reported the amount of possessions and books at home (ST012 and ST013). Five indices were derived from these items: i) family wealth possessions (WEALTH); ii) cultural possessions (CULTPOSS); iii) home educational resources (HEDRES); iv) ICT resources (ICTRES); and v) home possessions (HOMEPOS). Table 16.3 below gives an overview of the indicator items for each of these indices.

Table 16.4   Indicators of household possessions and home background indices

Item	Description	Item is used to measure index		
		Strand C HOMEPOS15	HOMEPOS15	HOMEPOS
ST062Q01TA	A desk to study at		X	X
ST062Q02TA	A room of your own		X	X
ST062Q03TA	A quiet place to study		X	X
ST062Q04TA	A computer you can use for school work		X	X
ST062Q05TA	Educational software		X	X
ST062Q06TA	A link to the Internet		X	X
ST062Q10TA	Books of to help you with your school work		X	X
ST062Q12TA	A dictionary		X	X
ST064Q01NA	A table to have meals	X	X	
ST064Q03NA	A washer	X	X	
ST064Q04NA	A refrigerator or freezer	X	X	
ST064Q06NA	A stove or burner for cooking	X	X	
ST063Q01TA	Televisions	X	X	X
ST063Q02TA	Cars	X	X	X
ST063Q03TA	Rooms with a bath or shower	X	X	X
ST063Q04TA	Cellphones with internet access (smartphones)	X	X	X
ST063Q05TA	Computers	X	X	X
ST063Q06TA	Musical instruments	X	X	X

Item	Description	Strand C HOMEPOS15	Item is used to measure index	
			HOMEPOS15	HOMEPOS
ST066Q01NA	Number of books in the home	X	X	X*

Note: *the question "Number of books in the home" in PISA-D has different response categories than in 2015.

The HOMEPOS15 scale for Strand C was constructed in two steps. In the first step, international item parameters for PISA 2015 items administered also in PISA-D were obtained from a concurrent calibration of the 2015 and PISA-D Strand B and C data. This step is identical with the regular scaling of HOMEPOS in PISA 2015. In the second step, items from PISA-D Strand C were scaled whereas parameters were fixed for all items administered also in 2015 and for which no unique (i.e. country-specific) item parameters became necessary. Item parameters for all other items (except national items) were freely estimated but constrained to be equal across countries. We conducted additional analyses on the invariance of item parameters across countries to consider assigning unique parameters if necessary. Once this process was finished, we estimated WLEs for all youth from PISA-D. By restricting the largest subset of items to be equal across PISA-D and PISA 2015, the HOMEPOS15 scores can be regarded to be on a joint scale, allowing for comparisons of countries across PISA-D and PISA 2015, and thus allowing to be used in the calculation of trend ESCS.

Table 16.5   Scale reliabilities for Household possessions indices in PISA-D countries

Country	Crombach's alpha
Guatemala	0.749
Honduras	0.708
Panama	0.811
Paraguay	0.694
Senegal	0.715

Table 16.6   Item parameters for Household Possessions (HOMEPOS15)

Item	beta	d_1	d_2	d_3	alpha
Y040cQ01TA	-0.70843	1.80899	-0.64006	-1.16892	1.088631
Y040cQ02TA	0.86955	0.93075	-0.03666	-0.89409	1.232472
Y040cQ03TA	0.30502	1.87318	-0.58952	-1.28365	1.127248
Y040cQ04NA	-0.74917	0.39305	-0.73634	0.34330	1.126953
Y040cQ06NA	0.78970	1.65749	-0.31055	-1.34694	2.509960
Y040cQ07NA	1.08752	0.76733	-0.45244	-0.31489	1.155935
Y041cQ01NA	-2.66710				1.680299
Y041cQ03NA	-0.61033				1.667431
Y041cQ04NA	-2.59320				2.450752
Y041cQ06NA	-2.73295				1.782885
Y042cQ01NA	-0.13249	1.06106	-0.43941	-0.62166	0.475449

ECONOMIC, SOCIAL AND CULTURAL STATUS

In PISA-D the PISA index of economic, social and cultural status (ESCS) was computed with the same procedure applied to the other PISA cycles.

The ESCS is a composite score built by the indicators parental education (PARED), highest parental occupation (HISEI) and home possessions (HOMEPOS15) including books in the home via principal component analysis (PCA) (see the description of these variables above). These three variables represent the traditional components of socio-economic status: education, occupational status and income. In absence of a direct measure of household income, the presence of material possessions and assets in the household functions as a proxy for family wealth.

**Computation of ESCS in PISA-D**

For youth with missing data on one out of the three components, a regression imputation based on the other two variables predicted the third (missing) variable, adding a random component to the predicted value. Missing data on more than one component resulted in a missing value for the ESCS score. As the goal of the computation of ESCS was to link it to the index in PISA 2015, after imputation, all three components were standardised for PISA-D countries based on the OECD means and standard deviations from the PISA 2015 public dataset.

In PISA 2015, standardised variables, including imputed values, entered in the PCA to obtain ESCS values. As in previous cycles, ESCS was the component score for the first principal component. The PCA ran across equally weighted countries, including OECD members as well as partner countries/economies. Thus, all countries and economies contribute equally to the estimation of ESCS scores. However, for the purpose of reporting the ESCS scale has been transformed with zero being the score of an average OECD student and one being the standard deviation across equally weighted OECD countries.

In PISA-D we utilised the factor loadings and first eigenvalue from the PCA analysis of PISA 2015 and computed ESCS as the sum of the standardised variables multiplied by their factor loadings, divided by the first eigenvalue.

Figure 16.1   Computation of ESCS in PISA-D
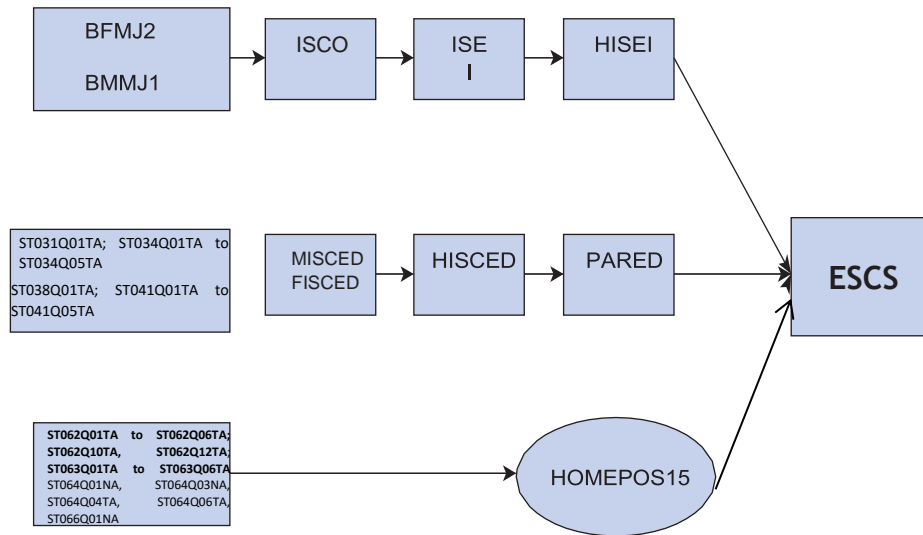
SCALES UNIQUE TO PISA-D

For the scales unique to PISA-D, international item and person parameters were obtained from a GPCM in a single analysis based on data from all persons in all seven countries using the TAM package in R. For each scale, only persons with a minimum number of three valid responses were included. Respondents were left unweighted, and all countries contributed equally to the estimation. We conducted additional analyses on the invariance of item parameters across countries and evaluated the possibility to assign unique parameters in case of severe misfit (see section "Cross-country comparability" in this chapter). The weighted likelihood estimates (WLE; Warm, 1989) obtained from this process represented individual participant scores. We transformed the WLEs into an international PISA-D metric ranging from 0 to 10, where 0 indicated that all the item responses were negative (or corresponding to the least endorsing option) and 10 indicated that all the item responses were positive (or corresponding to the most endorsing option). In some cases, we produced categorical variables based on critical cut-offs on the WLE continuum to identify vulnerable groups.

Table 16.7   Scales unique to PISA-D

DV Name	Description	Source	Variables	IRT scaling
DEPRESSION	Depression	Youth Interview	Y021cQ05 to Y021cQ10	YES
DEPRECAT	Levels of depression	Youth Interview	Y021cQ05 to Y021cQ10	
FAMRES	Family resources	Youth Interview	Y040cQ01 to Y040cQ07; Y041cQ01 to Y040cQ06; Y042cQ01; Y045cQ01; Y044cQ01; Y243cQ01; Y036cQ01	YES

DV Name	Description	Source	Variables	IRT scaling
POVERTY	Household poverty index	Youth Interview	Y040cQ01 to Y040cQ07; Y041cQ01 to Y040cQ06; Y042cQ01; Y045cQ01; Y044cQ01; Y243cQ01; Y036cQ01	

Depression

In the PISA-D student questionnaire, we asked youth about their mental health using six items originally developed by TLB in OurSchool Suvey. We asked youth to talk about their feelings at home or at school, and asked how often they occurred. The items touch upon crying without a reason, feeling lonely, not being able to enjoy life as others, feeling depressed or sad, struggling with sleeping and feeling bothered by a lot of things.

Higher WLEs and higher difficulty correspond to higher level of depression on all items. Once transformed to a 0-10 metric, the variable DEPRESSION indicated that youth who scored 0 did not display any sign of depression, and those who scored 10 reported intense occurrence of negative feelings.

The continuous variable DEPRESSION was recoded into an ordinal variable with three categories: i) 0, Not depressed, if DEPRESSION is lower than 4.31; ii) 1, Moderately depressed, if DEPRESSION ranges from 4.31 to 5.8; and iii) 2, Depressed, if DEPRESSION is greater than 5.8.

Table 16.8   Scale reliabilities for depression indices in PISA-D countries

Country	Crombach's alpha
Guatemala	0.775
Honduras	0.727
Panama	0.831
Paraguay	0.800
Senegal	0.782

Table 16.9   Item parameters for Depression

Item	beta	d_1	d_2	d_3	alpha
Y021cQ05NA	1.62467	-0.67298	0.19481	0.47817	0.83759
Y021cQ06NA	1.58161	0.11499	-0.14868	0.03369	1.31759
Y021cQ07NA	0.92158	-0.37023	-0.23997	0.61020	0.65987
Y021cQ08NA	2.02280	0.84802	-0.14036	-0.70765	1.69834
Y021cQ09NA	0.74208	-0.53743	0.04770	0.48973	0.48094
Y021cQ10NA	0.55137	0.04502	-0.28125	0.23623	0.68198

FAMRES and Poverty

The PISA-D student questionnaire contains several items related to material possessions at home as well as infrastructural and personal resources in the household. Other than the trend items

and the new items that constitute homeposs15, we also asked youth whether they shared a toilet facility with other people who are not members of their household and whether they had a flush toilet; what was the material of the floor in their home; whether any of the household members had a bank account; and finally whether the student experienced hunger in the last month. The WLEs of the index of family resources were coded on a 0 – 10 metric were 0 corresponded to a complete lack of resources and basic services in the home and 10 indicated the presence of all the listed items. A measure of household poverty based on FAMRES defined four categories: extremely poor, severely poor, poor and not poor. The four categories were based on three cut off points on the WLE scale, after ordering the items by their difficulty score, or probability greater than .5 to endorse the item. The WLE score of -3.5 corresponded to a response pattern indicating lack of very basic resources and experience of hunger. The WLE score of -1.8 corresponds to a pattern of limited resources and in particular lack of flush toilet in the household; finally, WLEs up to -.75 corresponds to patters of resources where youth report to have basic household possessions but do not have a washer, which is an item that requires running water and electricity in the home.

Table 16.10   Scale reliabilities for family resources indices in PISA-D countries

Country	Crombach's alpha
Guatemala	0.752
Honduras	0.544
Panama	0.515
Paraguay	0.754
Senegal	0.836

Table 16.11   Item parameters for Family Resources

Item	beta	d_1	d_2	d_3	alpha
Y041cQ01NA	-3.04817				1.33153
Y041cQ03NA	-0.73868				1.22466
Y041cQ04NA	-3.14500				2.00151
Y041cQ06NA	-2.93396				1.28216
Y040cQ01TA	-0.88866	1.950902	-0.70163	-1.24927	0.91666
Y040cQ02TA	0.73938	0.53967	-0.20545	-0.33422	0.72477
Y040cQ03TA	0.01226	1.769387	-0.68908	-1.08031	0.90159
Y040cQ04NA	-0.72873	0.311918	-0.77645	0.464534	0.71095
Y040cQ06NA	0.76155	1.202482	-0.34374	-0.85874	1.35597
Y040cQ07NA	0.79568	-0.02775	-0.26271	0.290454	0.52419
Y042cQ01NA	-0.15848	1.047106	-0.44006	-0.60705	0.31572
Y045cQ01NA	-1.12832				0.31081
Y044cQ01NA	-2.56594				1.51018
Y243cQ01NA	-1.36991	0.96702	-0.96702		0.70874
Y059cQ01NA	-1.56820				0.52368