Update of Study 2 (Form II exam data)

(AGM November 2009)

Progress update:

1) So far, 2007 and 2008 examination data have been combined and cleaned for analysis. Preliminary results using multilevel modelling of the data for these two years were disseminated at three 2009 events (UKFIET, BAAL, CREOLE).
2) We are still combining 2003, 2006 data.
3) Although not anticipated in the original design, the 2009 examination data will be included in our analysis when it is available (likely to be February 2010).
4) From both Studies 2 and 5.1, we have found considerable evidence that English language ability is a significant predictor of students' performance in other examination subjects (See below the analyses).
5) To support the interpretation of findings from the analysis of the examination data, we are also collecting additional (i.e. not in the original design) information using two questionnaires: 1) Student Questionnaire - focusing on, for example, opportunities to learn English in and outside school, exam preparation and SES data; 2) Head of School Questionnaire - focusing on, for example, school factors such as availability of textbooks, teacher qualifications for Form II teachers, and examination preparation camps.
6) We are also collecting further data on students English language ability, in particular, their receptive English vocabulary knowledge. We were not happy with the 'off the shelf' vocabulary measure we used during the learner workshops in January 09; we developed a "special purpose" vocabulary measure based on the corpus of textbooks and past exam papers and piloted this measure in September 09 (data entry for this pilot was undertaken in Zanzibar) and the revised version was administered in this round of data collection in our targeted schools.

Preliminary findings of 2007 and 2008 data

Below I present some preliminary findings about the effects of language abilities on examination performance in Maths, Biology and Chemistry. Multilevel modelling was used; the following two figures present an overview of the 2007 and 2008 Form II examination dataset. There are 10 districts in the data, 161 schools (156 in 2007, 158 in 2008, some new schools joined in 2008, a couple of schools that had data in 2007 but did not have data in 2008), and altogether 44455 students.

A summary of the key findings is presented in the Powerpoint slides (see Appendix 1, p.13)

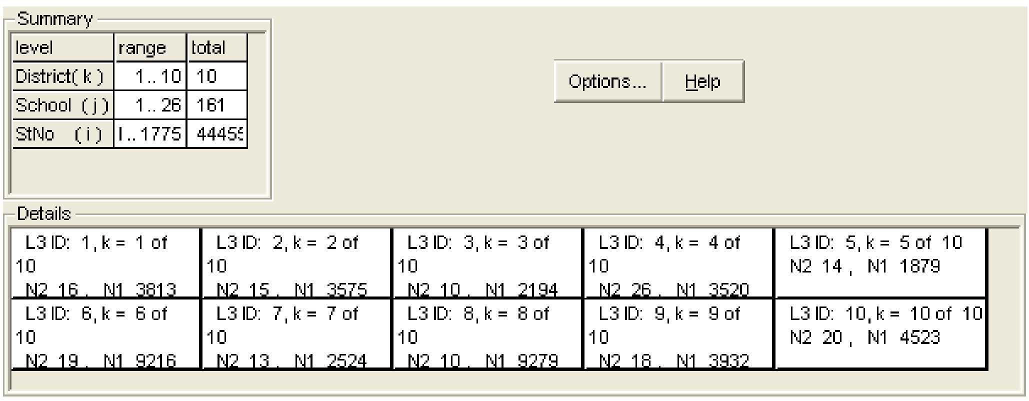

Figure 1: Data hierarchy view

Names

Edit name | Data | Toggle Categorical | Categories | Description | Copy | Paste

Name	Cn	n	missing	min	max
Year	1	44455	0	2007	2008
StNo	2	44455	0	1	1573
District	3	44455	0	1	10
School	4	44455	0	1	26
ENGLISH	5	44455	0	-4.07	264.79
KISW	6	44455	0	-7.42	153.1
DINI	7	44455	0	-10.64	142.07
KIARAB	8	44455	0	-8.01	144.45
GEOG	9	44455	0	-9.38	122.38
CIVICS	10	44455	0	-10.53	165.44
HISTOR	11	44455	1	-3.1	129.2
CHEMIST	12	44455	0	-2.02	146.82
PHYSICS	13	44455	0	-2.72	184.04
BIOLOG	14	44455	0	-1.99	159.24
MATHS	15	44455	0	-15.67	128.97
cons	16	44455	0	1	1

Figure 2: Variables and minimum and maximum scores
(**Note:** Due to the application of A.F., some students had scores below zero, and some well above 100. This issue was raised in the last AGM meeting)

Rationale for using 2-level modelling

The variances of each subject (English, Mathematics, Biology and Chemistry) at three different models (at pupil level only, at pupil and school levels, at pupil, school and district levels) are reported below, followed by the rationale for using 2-level modelling (i.e. pupil and school).

(1) ENGLISH

English: variance at pupil level (only)

$$\text{ENGLISH}_i \sim \text{N}(XB, \Omega)$$

$$\text{ENGLISH}_i = \beta_{0i}\text{cons}$$

$$\beta_{0i} = 28.294(0.062) + e_{0i}$$

$$\left[e_{0i}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[170.609(1.144)\right]$$

$$\textit{-2*loglikelihood(IGLS Deviance)} = 354628.800(44455 \text{ of } 44455 \text{ cases in use})$$

English variance at pupil, school levels

$$\text{ENGLISH}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{ENGLISH}_{ij} = \beta_{0ij}\text{cons}$$

$$\beta_{0ij} = 28.277(0.476) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[35.238(4.052)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[148.571(0.998)\right]$$

$$\textit{-2*loglikelihood(IGLS Deviance)} = 349091.500(44455 \text{ of } 44455 \text{ cases in use})$$

English variance at pupil, school and district levels

$$\text{ENGLISH}_{ijk} \sim \text{N}(XB, \Omega)$$
$$\text{ENGLISH}_{ijk} = \beta_{0ijk}\text{cons}$$
$$\beta_{0ijk} = 28.321(0.989) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim \text{N}(0,\ \Omega_v)\ :\ \Omega_v = \begin{bmatrix} 7.893(4.377) \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 27.485(3.289) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 148.574(0.998) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 349070.700(44455 of 44455 cases in use)

(2) MATHS

Math: pupil level only

$$\text{MATHS}_{i} \sim \text{N}(XB, \Omega)$$
$$\text{MATHS}_{i} = \beta_{0i}\text{cons}$$
$$\beta_{0i} = 13.314(0.043) + e_{0i}$$

$$\begin{bmatrix} e_{0i} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 82.477(0.553) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 322316.400(44455 of 44455 cases in use)

Math: pupil and school levels

$$\text{MATHS}_{ij} \sim \text{N}(XB, \Omega)$$
$$\text{MATHS}_{ij} = \beta_{0ij}\text{cons}$$
$$\beta_{0ij} = 14.274(0.419) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 27.800(3.154) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 62.600(0.421) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 310767.600(44455 of 44455 cases in use)

Total=90.4, School=30.75%, Pupil=69.25%

Maths: pupil, school and district levels

$$\text{MATHS}_{ijk} \sim \text{N}(XB,\ \Omega)$$

$$\text{MATHS}_{ijk} = \beta_{0ijk}\text{cons}$$

$$\beta_{0ijk} = 14.287(1.017) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\left[v_{0k}\right] \sim \text{N}(0,\ \Omega_v) \ : \ \Omega_v = \left[9.016(4.622)\right]$$

$$\left[u_{0jk}\right] \sim \text{N}(0,\ \Omega_u) \ : \ \Omega_u = \left[19.305(2.278)\right]$$

$$\left[e_{0ijk}\right] \sim \text{N}(0,\ \Omega_e) \ : \ \Omega_e = \left[62.601(0.421)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 310731.600(44455 of 44455 cases in use)

Note: As in ENGLSIH, district level variance is not significant.

(3) BIOLOGY

Biology: pupil level only

$$\text{BIOLOG}_{i} \sim \text{N}(XB,\ \Omega)$$

$$\text{BIOLOG}_{i} = \beta_{0i}\text{cons}$$

$$\beta_{0i} = 22.910(0.050) + e_{0i}$$

$$\left[e_{0i}\right] \sim \text{N}(0,\ \Omega_e) \ : \ \Omega_e = \left[109.767(0.736)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 335023.200(44455 of 44455 cases in use)

Biology: pupil and school levels

$$\text{BIOLOG}_{ij} \sim \text{N}(XB,\ \Omega)$$

$$\text{BIOLOG}_{ij} = \beta_{0ij}\text{cons}$$

$$\beta_{0ij} = 23.458(0.373) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u) \ : \ \Omega_u = \left[21.601(2.489)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e) \ : \ \Omega_e = \left[93.367(0.627)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 328437.100(44455 of 44455 cases in use)

Total: 114.968, school=18.79%, pupil=81.21%

Biology: pupil, school and district levels

$$\text{BIOLOG}_{ijk} \sim \text{N}(XB,\ \Omega)$$

$$\text{BIOLOG}_{ijk} = \beta_{0ijk}\text{cons}$$

$$\beta_{0ijk} = 23.472(0.740) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\left[v_{0k}\right] \sim \text{N}(0,\ \Omega_v) \ : \ \Omega_v = \left[4.267(2.450)\right]$$

$$\left[u_{0jk}\right] \sim \text{N}(0,\ \Omega_u) \ : \ \Omega_u = \left[17.655(2.115)\right]$$

$$\left[e_{0ijk}\right] \sim \text{N}(0,\ \Omega_e) \ : \ \Omega_e = \left[93.368(0.627)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 328421.600(44455 of 44455 cases in use)

Note: District level difference is not significant either.

(4) CHEMISTRY

CHEMISTRY: PUPIL LEVEL ONLY

$$\text{CHEMIST}_i \sim \text{N}(XB,\ \Omega)$$
$$\text{CHEMIST}_i = \beta_{0i}\text{cons}$$
$$\beta_{0i} = 33.185(0.076) + e_{0i}$$

$$\begin{bmatrix} e_{0i} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 255.891(1.716) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 372649.800(44455 of 44455 cases in use)

CHEMISTRY: PUPIL AND SCHOOL LEVELS

$$\text{CHEMIST}_{ij} \sim \text{N}(XB,\ \Omega)$$
$$\text{CHEMIST}_{ij} = \beta_{0ij}\text{cons}$$
$$\beta_{0ij} = 34.536(0.623) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 60.852(6.968) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 215.141(1.446) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 365577.800(44455 of 44455 cases in use)

Total=275.993, school=22.05%, pupil=77.95%

CHEMISTRY: PUPIL, SCHOOL AND DISTRICT LEVELS

$$\text{CHEMIST}_{ijk} \sim \text{N}(XB,\ \Omega)$$
$$\text{CHEMIST}_{ijk} = \beta_{0ijk}\text{cons}$$
$$\beta_{0ijk} = 34.480(1.035) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim \text{N}(0,\ \Omega_v)\ :\ \Omega_v = \begin{bmatrix} 7.048(4.802) \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 54.379(6.445) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 215.143(1.446) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 365571.500(44455 of 44455 cases in use)

Note: As in other subjects, the difference between districts is not significant for chemistry either.

The differences in total variances, taking ENGLISH as an example, between Model 1 (170.609) and Model 2 above (35.238+148.571=183.809) demonstrate the importance of using multilevel analyses and the preciseness of multilevel modelling than simple regression analyses without considering the situations that students cluster with school and schools within district. If we take into consideration only 1 level (pupil-level) as in traditional regression analyses, the variance in ENGLISH is somewhat underestimated (170.609 compared to 183.809). Taking into account school level (i.e. school effectiveness factors), the total variance increased slightly. We notice that the variance attributable to school level factors explains nearly 19.17% of the total variance (35.238/183.809); the pupil level variance is 80.83% of the total variance. The deviance test also indicated a large decrease from 354628 to 349091; even without adding an explanatory variable in the model. Similar findings are observed in other subjects.

As district-level variances are not significant for all the four subjects (English, maths, biology and chemistry), we decide to use only two level analyses (i.e. pupil and school) in the subsequent analyses on the relationship between ENGLISH and other core subjects.

Another advantage of using multilevel analyses is that we are able to partition the variance attributable to different levels, in this case, we will be able to identify what percentage of the variance is attributable to school-level and pupil-level factors.

The next section illustrates to what extent students' English language ability, compared to Kiswahili and Arabic (2 other language factors) may affect their performance in other core subjects (maths, biology and chemistry). Because significant variances in KISW and KIARAB are observed, we are also interested in finding out the extent to which these language-related variables, compared to ENGLISH, can account for the variances in other core subjects (maths, chemistry and biology).

Kiswahili

$$\mathrm{KISW}_{ij} \sim \mathrm{N}(XB, \Omega)$$
$$\mathrm{KISW}_{ij} = \beta_{0ij}\mathrm{cons}$$
$$\beta_{0ij} = 45.555(0.663) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \left[68.693(7.892)\right]$$

$$\left[e_{0ij}\right] \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \left[273.116(1.835)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 376166.400(44455 of 44455 cases in use)

Arabic

$$\mathrm{KIARAB}_{ij} \sim \mathrm{N}(XB, \Omega)$$
$$\mathrm{KIARAB}_{ij} = \beta_{0ij}\mathrm{cons}$$
$$\beta_{0ij} = 33.936(0.680) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \left[72.483(8.286)\right]$$

$$\left[e_{0ij}\right] \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \left[236.714(1.591)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 369838.300(44455 of 44455 cases in use)

Maths and English

$$\text{MATHS}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{MATHS}_{ij} = \beta_{0ij}\text{cons} + 0.336(0.003)\text{ENGLISH}_{ij}$$

$$\beta_{0ij} = 4.766(0.354) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u) \ : \ \Omega_u = \left[18.882(2.135)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e) \ : \ \Omega_e = \left[45.835(0.308)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 296898.100(44455 of 44455 cases in use)

Total: 64.717, school=29.18%, pupil=70.82%

Compared with the cons model of maths: Total=90.4, School=30.75%, Pupil=69.25%):

$$\text{MATHS}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{MATHS}_{ij} = \beta_{0ij}\text{cons}$$

$$\beta_{0ij} = 14.274(0.419) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u) \ : \ \Omega_u = \left[27.800(3.154)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e) \ : \ \Omega_e = \left[62.600(0.421)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 310767.600(44455 of 44455 cases in use)

Finding: Adding ENGLISH explains (90.4-64.717)/90.4=28.41% of the maths total variance; and (27.8-18.882)/27.8=32.08% of the school level total variance and (62.6-45.835)/62.6=26.78% of the pupil level total variance. In other words, ENGLISH can explain nearly 1/3 of the total variance of maths, and 1/3 of school-level and 1/3 of pupil-level variances in maths too.

Compared with Kiswahili and Arabic as predictors

$$\text{MATHS}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{MATHS}_{ij} = \beta_{0ij}\text{cons} + 0.176(0.002)\text{KISW}_{ij}$$

$$\beta_{0ij} = 6.277(0.381) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u) \ : \ \Omega_u = \left[21.456(2.439)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e) \ : \ \Omega_e = \left[54.222(0.364)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 304361.900(44455 of 44455 cases in use)

Total: 75.678, school=28.35%, pupil=71.65%

KISW explains (90.4-75.678)/90.4=16.29% of the maths total variance.

$$\text{MATHS}_{ij} \sim \text{N}(XB, \Omega)$$
$$\text{MATHS}_{ij} = \beta_{0ij}\text{cons} + 0.212(0.002)\text{KIARAB}_{ij}$$
$$\beta_{0ij} = 7.069(0.363) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[19.907(2.264)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[51.978(0.349)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 302477.900(44455 of 44455 cases in use)

Total: 71.885, school=27.69%, pupil=72.31%
KIARAB explains (90.4-71.885)/90.4=20.48% of the maths total variance.

KISW and KIARAB are less capable of explaining the MATHS total variance than ENGLISH.

BIOLOGY AND ENGLISH

$$\mathrm{BIOLOG}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{BIOLOG}_{ij} = \beta_{0ij}\mathrm{cons} + 0.517(0.003)\mathrm{ENGLISH}_{ij}$$
$$\beta_{0ij} = 8.848(0.289) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[11.928(1.372)\right]$$

$$\left[e_{0ij}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[53.718(0.361)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 303856.400(44455 of 44455 cases in use)

Total=65.646, school=18.17%, pupil=81.83%

Compared with the biology cons model (Total: 114.968, school=18.79%, pupil=81.21%):

$$\mathrm{BIOLOG}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{BIOLOG}_{ij} = \beta_{0ij}\mathrm{cons}$$
$$\beta_{0ij} = 23.458(0.373) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[21.601(2.489)\right]$$

$$\left[e_{0ij}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[93.367(0.627)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 328437.100(44455 of 44455 cases in use)

Finding: Adding ENGLISH alone explains (114.968-65.646)/114.968=42.90% of the total variance in BIOLOGY (see the cons model), and (21.601-11.928)/21.601=44.78% of school-level variance in BIOLOGY, and (93.367-53.718)/93.367=42.47% of the pupil-level variance in BIOLOGY.

Compared with Kiswahili and Arabic as predictors

Kiswahili

$$\mathrm{BIOLOG}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{BIOLOG}_{ij} = \beta_{0ij}\mathrm{cons} + 0.330(0.002)\mathrm{KISW}_{ij}$$
$$\beta_{0ij} = 8.444(0.306) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[12.855(1.479)\right]$$

$$\left[e_{0ij}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[63.730(0.428)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 311439.300(44455 of 44455 cases in use)

Total=76.585, school=16.79%, pupil=83.21%
KISW explains (114.968-76.585)/114.968=33.39% of the BIO total variance.

Arabic

$$\text{BIOLOG}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{BIOLOG}_{ij} = \beta_{0ij}\text{cons} + 0.289(0.003)\text{KIARAB}_{ij}$$

$$\beta_{0ij} = 13.660(0.323) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[14.925(1.727)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[73.644(0.495)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 317868.000(44455 of 44455 cases in use)

Total=88.569, school=16.85%, pupil=83.15%

KIARAB explains (114.968-88.569)/114.968=22.96% of the BIO total variance.

Finding: KISW and KIARAB are less capable of explaining the BIOLOGY total variance than ENGLISH.

CHEMISTRY AND ENGLISH

$$\text{CHEMIST}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{CHEMIST}_{ij} = \beta_{0ij}\text{cons} + 0.793(0.004)\text{ENGLISH}_{ij}$$

$$\beta_{0ij} = 12.122(0.497) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0, \Omega_u) : \Omega_u = \left[36.497(4.174)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0, \Omega_e) : \Omega_e = \left[121.784(0.818)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 340290.000(44455 of 44455 cases in use)

Total=158.281, school=23.06%, pupil=76.94%

Compared with the chemistry consl model (Total=275.993, school=22.05%, pupil=77.95%)

$$\text{CHEMIST}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{CHEMIST}_{ij} = \beta_{0ij}\text{cons}$$

$$\beta_{0ij} = 34.536(0.623) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0, \Omega_u) : \Omega_u = \left[60.852(6.968)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0, \Omega_e) : \Omega_e = \left[215.141(1.446)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 365577.800(44455 of 44455 cases in use)

Finding: Adding ENGLISH alone explains (275.993-158.281)/275.993=42.65% of the total chemistry variance, and (60.852-36.497)/60.852=24.36% of school level chemistry variance, and (215.141-121.784)/215.141=43.39% of pupil level chemistry variance.

Compared with Kiswahili and Arabic as predictors

Kiswahili

$$\text{CHEMIST}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{CHEMIST}_{ij} = \beta_{0ij}\text{cons} + 0.487(0.004)\text{KISW}_{ij}$$

$$\beta_{0ij} = 12.365(0.521) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0, \Omega_u) : \Omega_u = \left[38.417(4.412)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0, \Omega_e) : \Omega_e = \left[150.537(1.012)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 349687.400(44455 of 44455 cases in use)

TOTAL=188.954, school=20.33%, pupil=79.67%
KISW explains (275.993-188.954)/275.993=31.54% of the total chemistry variance.

Arabic

$$\text{CHEMIST}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{CHEMIST}_{ij} = \beta_{0ij}\text{cons} + 0.479(0.004)\text{KIARAB}_{ij}$$

$$\beta_{0ij} = 18.295(0.538) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[42.508(4.874)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[160.944(1.081)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 352664.500(44455 of 44455 cases in use)

Total=203.452, school=20.89%, pupil=79.11%
KIARAB explains (275.993-203.452)/275.993=26.28%

Finding: KISW and KIARAB are less capable of explaining the CHEMISTRY total variance than ENGLISH.

Summary of key preliminary findings (see also Appendix 1)

1. It is very clear that ENGLISH is a significant and substantial predictor of the students' performance in other core subjects: maths, biology and chemistry.
2. Although two other language factors (Kiswahili and Arabic) are also significant predictors of the students' performance in other core subjects, it is noted that they are less capable of explaining the variances than ENGLISH.
3. The statistics, in particular the school-level variances explained in the cons models (i.e. without using explanatory variables) as well as in the models including ENGLISH as the single explanatory variable, demonstrated a substantial proportion of the variance is attributable to school level factors (29.18% for maths, 18.17% for biology, and 23.06% for chemistry in the models using ENGLISH as the explanatory variable; 30.75% for maths, 18.79% for biology and 22.05% for chemistry in the cons models), although apparently pupil level factors play the major role in determining academic achievements As ENGLISH itself is affected as other core subjects by various school and pupil factors, this probably explains why there is not much improvement in terms of fitness of the models (measured by the change of % of school-level variance in the total variance). We therefore recommend that the Ministry collects and monitors both school- and pupil-level data, on a regular and systematic basis, to examine what factors (in particular, in relation to English language learning opportunities at home and at school) and how much these factors can account for the variances (in the tradition of school effectiveness studies), as well as the role of ENGLISH language ability compared to the other factors. Pupil and head teacher questionnaires in our targeted schools were administered during the current round of data collection (November 2009). In addition, a new English vocabulary test was administered in our targeted schools. These instruments will inform on these issues but, at a later stage, it is recommended that the MOEVT explore these issues further. Alternatively, SUZA in collaboration with the Ministry, could explore these issues through a further funded research study.
4. Further analyses on the Form II exam results are being conducted.

Appendix 1: Powerpoint slides about the Form II 2007 and 2008 analyses presented at UKFIET Oxford Conference

40

ENG and MATH

$$\text{MATHS}_{ij} \sim N(XB, \Omega)$$
$$\text{MATHS}_{ij} = \beta_{0ij}\text{cons} + 0.336(0.003)\text{ENGLISH}_{ij}$$
$$\beta_{0ij} = 4.766(0.354) + u_{0j} + e_{0ij}$$
$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 18.882(2.135) \end{bmatrix}$$
$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 45.835(0.308) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 296898.100(44455 of 44455 cases in use)

Total: 64.717, school=29.18%, pupil=70.82%

ENGLISH explains (90.4-64.717)/90.4=28.41% of the maths total variance

$$\text{MATHS}_{ij} \sim N(XB, \Omega)$$
$$\text{MATHS}_{ij} = \beta_{0ij}\text{cons}$$
$$\beta_{0ij} = 14.274(0.419) + u_{0j} + e_{0ij}$$
$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 27.800(3.154) \end{bmatrix}$$
$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 62.600(0.421) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 310767.600(44455 of 44455 cases in use)

Total=90.4, School=30.75%, Pupil=69.25%

University of BRISTOL

41

ENG and BIO

$$\text{BIOLOG}_{ij} \sim N(XB, \Omega)$$
$$\text{BIOLOG}_{ij} = \beta_{0ij}\text{cons} + 0.517(0.003)\text{ENGLISH}_{ij}$$
$$\beta_{0ij} = 8.848(0.289) + u_{0j} + e_{0ij}$$
$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 11.928(1.372) \end{bmatrix}$$
$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 53.718(0.361) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 303856.400(44455 of 44455 cases in use)

Total=65.646, school=18.17%, pupil=81.83%

ENGLISH alone explains (114.968-65.646)/114.968=42.90% of the total variance in BIOLOGY

$$\text{BIOLOG}_{ij} \sim N(XB, \Omega)$$
$$\text{BIOLOG}_{ij} = \beta_{0ij}\text{cons}$$
$$\beta_{0ij} = 23.458(0.373) + u_{0j} + e_{0ij}$$
$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 21.601(2.489) \end{bmatrix}$$
$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 93.367(0.627) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 328437.100(44455 of 44455 cases in use)

Total=114.968, school=18.79%, pupil=81.21%

University of BRISTOL

42

English and CHEM

$\text{CHEMIST}_{ij} \sim \text{N}(XB, \Omega)$

$\text{CHEMIST}_{ij} = \beta_{0ij}\text{cons} + 0.793(0.004)\text{ENGLISH}_{ij}$

$\beta_{0ij} = 12.122(0.497) + u_{0j} + e_{0ij}$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[36.497(4.174)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[121.784(0.818)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 340290.000(44455 of 44455 cases in use)

Total=158.281, school=23.06%, pupil=76.94%

ENGLISH explains (275.993-158.281)/275.993=42.65% of the total CHEM variance

$\text{CHEMIST}_{ij} \sim \text{N}(XB, \Omega)$

$\text{CHEMIST}_{ij} = \beta_{0ij}\text{cons}$

$\beta_{0ij} = 34.536(0.623) + u_{0j} + e_{0ij}$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[60.852(6.968)\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[215.141(1.446)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 365577.800(44455 of 44455 cases in use)

Total=275.993, school=22.05%, pupil=77.95%

University of BRISTOL

43

How about KISWAHILI & ARABIC?

- Although other two languages (Kiswahili and Arabic) are also significant predictors of the students' performance in maths, biology and chemistry, it is noted that they are less capable of explaining the variances than ENGLISH.
- KISWAHILI explains (275.993-88.954)/275.993=31.54% of CHEM total variance, (114.968-76.585)/ 114.968=33.39% of BIO total variance, (90.4-75.678)/90.4=16.29% of MATH total variance.
- ARABIC explains (275.993-203.452)/275.993=26.28% of the CHEM total variance, (114.968-88.569)/114.968=22.96% of the BIO total variance, (90.4-71.885)/90.4=20.48% of the maths total variance.

S P I N E Student Performance in National Examinations

University of BRISTOL

44

Summary of the multilevel models (a)

- It is very clear that ENGLISH is a significant and substantial predictor of the students' performance in MATH, BIO & CHEM.
- The school-level variances explained in the cons models as well as in the models including ENGLISH as the single explanatory variable demonstrated a **substantial proportion of the variance is attributable to school factors**

S Student P Performance I in N National E Examinations

University of BRISTOL

45

Summary of the multilevel models (b)

- Not much improvement in terms of **fitness of the models** (measured by the change of % of school-level variance in the total variance).
- Therefore, essential to collect **further school- and pupil-level data** to examine what factors (e.g. English language learning opportunities at home and at school, academic English proficiency) and how much they account for the variances (in the tradition of school effectiveness studies)

S Student P Performance I in N National E Examinations

University of BRISTOL