January 2014

Outliers in educational achievement data: Their potential for the improvement of performance

Tooba Farooqui
Aga Khan University, Examination Board

Isba Mustafa
Aga Khan University, Examination Board

Thomas Christie
Aga Khan University, Examination Board

Pak. J. Statist.
2014 Vol. 30(1), 71-82

OUTLIERS IN EDUCATIONAL ACHIEVEMENT DATA: THEIR POTENTIAL FOR THE IMPROVEMENT OF PERFORMANCE*

Farooqui, T.,[1] Mustafa, I.[2] and **Christie, T.[3]**
The Aga Khan University, Karachi, Pakistan
Email: [1]tooba.farooqui@aku.edu
[2]isbahzaidi@gmail.com
[3]thomas.christie@aku.edu

ABSTRACT

Statistical outliers (Barnett and Lewis, 1994) have customarily been identified as either potential threats to data integrity (Cook, 1977) or as potential distortions of estimates of central tendency (Strutz, 2010). In both circumstances outliers are treated as having nuisance value. On the other hand, Gladwell (2008) views outliers as success stories. The study adopts the latter approach to reach to classrooms and teachers which add value in student's academic capabilities. Scores of 86207 students from 423 government schools in 2010 middle school (Grade VI to VIII) promotion examinations are analysed to explore the possibility of identifying gifted teachers.

The study reports successful use of non-parametric methods of box plot and quartile formula to identify outstanding student performance. This performance is independent of overall school achievement, school size, student gender and is attributable to either individual or combined teaching effort.

KEYWORDS

Outlier; value added; middle school promotion exams; government schools.

1. INTRODUCTION

The use to which academic achievement data is put distinguishes formative from summative assessment. Formative assessment provides feedback during the academic year to be incorporated in teaching and learning, thus has the potential to nourish student achievement. Summative assessment provides a statement of achievement at the end of the academic year which cannot be altered through effort and its input in classroom is titled as backwash effect. Feedback is viewed as a tool to enhance learning while backwash is believed to lead to fragmentary efforts (Biggs, 2006). Aga Khan University-Examination Board (AKU-EB) works for 'formative use of assessment'. This use of summative assessment data is aimed to improve student learning and classroom practices. This quest is achieved by identification of cognitive and content areas for improvement

* The paper was first presented in the *9th International Conference on Statistical Sciences* at Lahore, Pakistan on 5th of July 2012.

and by making meaningful comparisons of school performance available to affiliated schools.

The paper is built on student achievement captured through middle school promotion exams which are summative as they occur at the end of the academic year. However we attempt formative use of this data to identify exemplars of successful teaching efforts. Use of statistical analysis to identify successful efforts gained popularity with the publication of Outliers-Story of Success. This book advocates recognition of distinctiveness of an outlier instead of labeling it as an abnormal statistical observation (Gladwell, 2008).

The study is built on nine mean scores per school, each based on three separate subject promotion examinations at the end of Grade VI, VII and VIII. Outliers are the data points which stand out distinctively from the remainder of the set (Sim, 2005; Barnett, 1993) within each school. They may be deleted because they lead to biased estimation of central tendencies (Lalitha, 2012; Zijlstra, 2007) but when the central tendency of Pakistani education appears to be so unsatisfactory these data points may open a fresh and meaningful perspective (see Gladwell, 2008).

Use of Gladwell's approach to identify effective classroom practices through promotion exams is not a safe game. The paper uses exam data from government schools in which 42% students scored below the 33% marks required to pass the exam leading to a skewness of .3 or greater in seven out of 9 subject-grade results (Table 1). This finding is in confirmation to the reported low student achievement in Pakistan (for example in ASER, 2011; Anderabi, 2008). This low achievement is in harmony with poor input in education in terms of teacher quality, learning material and school infrastructure reported by Aly (2007). However, there are teachers who survive with excellence in this dire situation in the government schools (Rizvi, 2007). Inspired by this finding, the paper uses this highly skewed data to search for teachers who add value to their students' capabilities.

Table 1:
Student Performance in Middle School Promotion Exams

Subject	Grade	N	% Pass	Mean	Std. Deviation	Skewness
English	VI	27377	58.1	39.29	23.41	.33
	VII	23209	58.5	37.56	21.57	.33
	VIII	21437	66.4	42.76	22.48	.18
Math	VI	23136	53.5	34.26	21.01	.54
	VII	23201	55.5	35.60	22.02	.46
	VIII	20875	62.1	39.67	22.28	.37
Science	VI	21953	48.8	32.31	19.31	.40
	VII	20509	43.9	35.30	19.87	.44
	VIII	22837	67.3	42.68	20.53	.16

Source: Students' scores in Middle School Promotion Exams 2010

The search for outlier teachers is aimed to respond to two research questions;

a) What is the frequency of occurrence of within school outlier performance in English, Mathematics and Science promotion examinations in middle school classes (Grade VI to VIII) in rural Sindh (a province in Pakistan)?
b) Are such performances related to outstanding teachers, outstanding schools or simple reporting error?

The unit of observation is subject-grade result within schools.

2. METHODOLOGY

We use student achievement data of middle school promotion exams. These exams are developed by District Education Offices and are conducted and marked by the schools. In 2010, exams in English, Mathematics and Science in 14 districts in Sindh and Balochistan (two out of four provinces in Pakistan) were developed with the support of AKU-EB through the USAID funded ED-LINKS Project and were conducted in 2100 government middle schools along with exams in five other subjects which did not receive project support. AKU-EB facilitated centralized marking guided by question specific mark schemes of 100886 students' scripts from 454 schools. The study uses 423 schools which belong to Sindh.

The study uses subject-grade school results to identify value added teachers. School accountability through student achievement got empirical attention with government initiatives of 'No Child Left Behind' in US, 'Every Child Matters' in UK and National Assessment Programme–Literacy and Numeracy in Australia. Statistical debates are about use of longitudinal or latitudinal data, comparison within or between schools and on adjusted or unadjusted scores (Heck, 2000; Tekwe, 2004; Lisa, 2004). This study uses students' marks in exams standardized for each subject and grade at the district level (each district has its own exams), to compare subject-grade results within each school.

Demographic covariates such as students' gender and socio-economic background and school location affect student achievement (Raudenbush, 2004) thus influence comparisons between schools. Comparison between cohorts within a school has issues of cohort effect, missing data and student mobility (Kelly, 2007). Comparison of subject-grade mean scores in one calendar year within a school controls for these factors.

This within school comparison is made to identify outliers which Grubbs defines as 'one that appears to deviate markedly from other members of the sample'. Box plot, trimmed means, extreme standardized deviate, Dixon test, and outliers in regression are the popular methods to detect outliers. The study uses a box plot quartile formula and Grubbs Test. In the study student scores were scaled to eliminate differences in subject difficulty at each Grade. Each district had its own exam thus T scores were calculated district-wise. Outliers within schools were identified in mean T scores in three subjects in three Grade levels.

Method 1 applies the non-parametric method using the quartiles of standardized subject-grade school means. The quartile scores $Q1$, $Q2$ and $Q3$ of the standardized

means and the inter quartile range $(IQR = Q3 - Q1)$ have been calculated in each school to find (a) $Q1 - 1.5 * IQR$, (b) $Q3 + 1.5 * IQR$ where any observation below a) or above b) is an outlier[1].

Method 2 applies parametric statistics using Grubbs Test *for Detecting* Outliers, also called the E*xtreme Studentized Deviate* (ESD) on the dataset of 223 schools with 9 grade-subject results. Formula of Grubbs Test is $Z = \frac{\left[x_i - \bar{x}\right]}{s}$. The test was applied to the T scores with the outliers defined as $Z < > \pm 2.5$. For using Grubbs Test we test the normality by using One-Sample Kolmogorov-Smirnov Test for the subject-grade results of 223 schools in the sample and found all the p-values between 0.137 and 0.804 which are > 0.05. Hence our data is normally distributed**.**

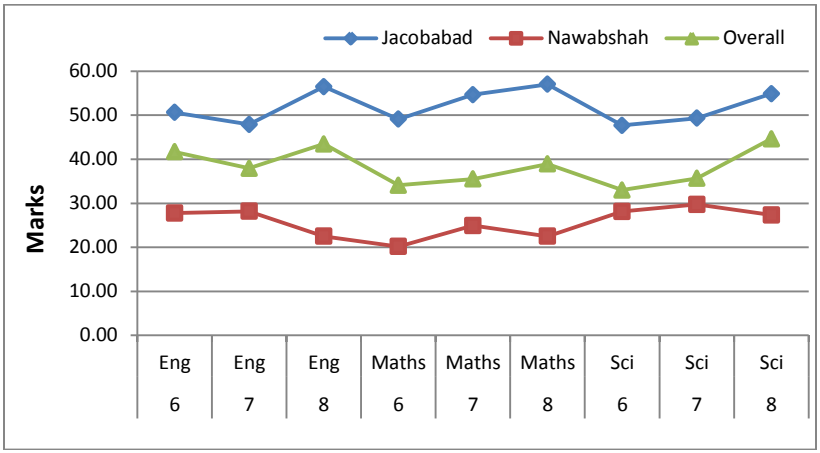
We performed Man Whitney-U Test to check normality of distribution of scores in schools without outliers in comparison to schools with positive outliers and schools with negative outliers. We found significance value in the range of 0.55-0.916 which are > 0.05. Thus, retain the null hypothesis that outliers are randomly distributed in the population.

The average number of observations on which each mean T score is based is 68 which suggests that in many cases we are looking at multiple sections of a Grade in a school. Typically one teacher handles a subject in one section thus several teachers teach the same subject in multiple sections of a Grade. Unfortunately the data was not captured at section level so phone calls were made to the head teacher. The caller gave the context of the enquiry and asked the head teacher to ensure that they had the time table of 2009-2010 academic years available for further discussion to take place by phone the next day. The next day the conversation started with inquiry about number of sections of the outlier grade –subject in 2009-2010. When asked for the possible source of outlier result 92% of the head teachers identified a subject teacher as the primary source. In these cases an attempt was made to identify what made the teaching special. Head teachers were also asked for possibility of contribution of the school environment in the outliers result which 15% of them found as a secondary source.[see Annexure One: Protocol for phone calls].

3. FINDINGS

The data are drawn from districts of Sindh. Each district constructs its own promotion examinations though in every case they are based on the text books of the Sind Textbook Board and are similarly structured. The level of reported result, the mean subject-grade raw scores at district level varies markedly (Figure 1). District Jacobabad and Nawabshah have highest and lowest overall student mean scores. The scores from the other districts have been merged for comparison. These three sets of student performance data are non-overlapping in each of the nine subject-grade results.

[1] Although other approaches have also been used, we have confirmed this formula Q3+1.5*IQR from the following educational links. http://www.amstat.org/publications/jse/v19n2/dawson.pdf; http://www.mathwords.com/o/outlier.htm; http://www.stat.wmich.edu/s160/book/node8.html; http://www.unc.edu/~rls/s151-09/class4.pdf;http://www.miniwebtool.com/outlier-calculator/



Source: Students' scores in Middle School Promotion Exams 2010

Figure 1: District-wise Mean Scores of Students in Exams

The possibility remains that these independently produced examinations vary in their discriminating power. However, Table 2 gives little reason for concern on that account. All data was collected from centralized marketing centers, one per district.

Table 2:
Outliers in Districts

District	Total Schools	Outlier Results			Total Outliers	
		English	Math	Science	Frequency	% of Subject Grades
Dadu	23	2	2	3	7	**3.4**
Jacobabad	32	1	1	3	5	**1.7**
Jamshoro	20	2	2	0	4	**2.2**
Kashmore	33	5	5	0	10	**3.4**
Mirpurkhas	41	6	4	6	16	**4.3**
Tharparkar	37	3	4	7	14	**4.2**
Nawabshah	11	1	3	5	9	**9.1**
Sanghar	12	1	3	0	4	**3.7**
Shikarpur	8	0	1	2	3	**4.2**
Sukkur	6	3	2	0	5	**9.3**
Total	**223**	**24**	**27**	**26**	**77**	

Source: Students' scores in Middle School Promotion Exams 2010

In the first phase of data cleaning we have student scores of 424 schools with 3112 subject-grade mean scores in the sample. There are 118 schools with one, two or three out of nine subject grade results, they have been included in the calculation of district standardized scores but are not included in the search for outliers**.** In the second phase of data cleaning we have 376 schools with more than 3 subject grade results The frequency of remaining schools is reported in Table 3. But we were interested in 223 schools having complete (nine) subject grade results for 2007 data points.

Table 3:
Frequency of Outlier Subject / Grade Means

Types of Schools	# Schools	Subject Grade result	Median	Schools with Outliers	%age
With complete data	223	2007	9	64	29
With incomplete data	153	1005	7	35	23
Total	**376**	**3012**	**9**	**99**	**26**

Table 4 reveals that these schools with complete data register 77 outlier results in 64 schools in 10 districts. The distribution of outliers among the districts is un-even (Table 4).

Table 4:
Schools with Subject Results and Outliers

Count of Subject Grade Results	Schools without Outliers	Schools with one Outlier	Schools with Two Outliers	Total Schools
4	15	2	0	17
5	11	5	0	16
6	34	6	0	40
7	18	5	0	23
8	40	16	1	57
9	159	51	13	223
Grand Total	**277**	**85**	**14**	**376**

Source: Students' scores in Middle School Promotion Exams 2010

The number of schools from the last two districts, Shikarpur and Sukkur is comparatively low. The marking activity in these two districts coincided with the conduct of secondary school certificate examinations by Sukkur Board which called upon the same pool of potential markers and resulted in the output from the centralized promotion exam being reduced to six subject grade means and the exclusion of 30 schools from Table 3.

There is no subject bias as outliers are found in English math and science results in almost the same ratio (Table 5). Also, this percentage of outliers is above expected 1.5 outliers data points on one side of the distribution tail. We have identified more outliers

in Grade VI and VIII. These are the entry and exit points in middle schools. It seems that primary schooling may retain its influence on student's knowledge in Grade VI. Grade VIII demands additional teaching and learning efforts for successful exit. Grade VII appears to be a dip where the enthusiasm at entry to the middle school has dried but the anxiety of school level exit has not yet started.

After exploring the student achievement data we proceed to the first research question;

a) What is the frequency of occurrence of outliers performance in English, Mathematics and Science promotion examinations in middle school classes in rural Sindh?

Table 5:
Outliers in Subject and Grade Results

Subject	Frequency			Total	
	Grade VI	Grade VII	Grade VIII	Frequency	Percent of Occurrence
English	8	8	8	24	3.6
Math	11	7	9	27	4.0
Science	8	5	13	26	3.9
Total	**27**	**20**	**30**	**77**	
Percent of Occurrence	4.0	3.0	4.5		

Source: Students' scores in Middle School Promotion Exams 2010

Now we search for the response to the second question;

a) Are such performances related to outstanding teachers, outstanding schools or simple reporting error?

To identify school factor in identification of outliers results we look at student performance, school size and demographic features of a school. The 64 schools with outliers (Mean score 39.51 SD 11.1) do not differ in achievement from 159 schools without outliers (Mean score 41.33 SD 11.0) in terms of student performance ($t=1.12$ with 221 d.f., n.s.). Thus outliers results are not dependent upon overall level of achievement of students in a school.

We look into cumulative percent of mean scores of schools with and without outliers. The distribution of mean scores of schools with and without outliers is not different. Kolmogorov Simirnov Observed D is 0.08 against a Critical D value of 0.2.

There is a potential statistical biasing effect on the variance of school means of sample size. Running the analysis within school greatly reduces the range of sample sizes within school and there is no significant difference in the number of students in the schools with outliers (mean 216.6 and SD 140.6) and without outliers (mean 201.03, SD 141.2) ($t<1$).

School demographics are for further threats of biasing effects. We investigate school location, school gender and school exit level.

School exit level does not have influence on the probability of having an outliers (Chi-square=4.425, d.f.=2 n.s.) (Table 6).

Table 6:
Outlier Results in Middle, Secondary and Higher Secondary School

School exit level	School with Outliers		Schools without Outliers	
	N	% of valid	N	% of valid
Middle (till Grade VIII)	8	14	38	27
Secondary (till Grade X)	41	69	82	59
Higher Secondary (till Grade XII)	10	17	19	14
Missing	5		20	
Total	**64**	**100**	**159**	**100**

Source: Students' scores in Middle School Promotion Exams 2010

The probably of outliers results is not dependent on school location in urban or rural area (Chi-square < 1.0). Gender difference does not influence outliers result probability and girls, boys and co-education schools have the same probability of outliers result (Chi-square=1.71, with 2 d.f., n.s).

4. FOLLOW-UP

The telephone numbers of 31 out of 66 schools with outliers' results were traced. The selection criterion for making a phone call was access to a phone number through multiple resources. 21 schools had functioning phone numbers. These 21 schools had 25 outliers results. For one outlier result the head teacher refused to accept the result. Another linked it with cheating in exams. The remaining 23 results are attributed to the teacher. Most frequent reasons offered for value-added teaching were punctuality of the teacher and their frank and encouraging attitude to students. On inquiry, four head teachers mentioned the school environment and management contribution to the outliers result. 13 of these 25 outliers subject-results belong to classes taught by one teacher; remaining results are joint efforts of teachers due to multiple sections in a Grade or multiple teachers teaching the outliers subject in one class. Student performance data was not captured at section level so we cannot distinguish teacher contribution in schools with multiple sections.

We used two non-parametric formulas. Quartile formula identified 77 outliers in 223 schools whereas Box Plot showed 72. The school, subject and Grade of these 72 outliers are common. Five outliers identified through quartile are not visible in Box Plots. These five outliers occur in four schools. The outliers subject-grade result in these schools is closer to the rest of the subject-grade results as compared to other outliers results. Using

parametric statistics we found two subject-grade outliers results. These two outliers belong to schools with mean subject grade scores in the range of 34.13 to 36.02 (33% is the cut score to pass the exam).

20 out of 21 head teachers were able to point out the root of outliers student performance. We do not know about team or individual teaching in the rest of 223 schools in the sample. Thus, whether team teaching or individual teaching cause outliers result is not known. This area needs further investigation.

Telephone inquiry reveals attribution of 23 out of 25 outliers results to the teacher. Half of the results spring from classes with more than one teacher. Thus, with this data and methodology we cannot measure individual verses joint teaching as the source of outliers subject-grade school results.

5. CONCLUSION

4% results of math, science and English are outliers. Therefore there is no subject bias and the opportunity to outperform is open to skill based subjects of math and English and to content based subject of science. Grade bias is observed as Grade VII has fewer outliers results. The absence of excitement of middle school entry in Grade VI and of anxiety to enter secondary school after Grade VIII is one possible explanation. There is no effect of student gender, school location and school exit level on the probability of having an outliers result.

The non-parametric method identified outliers results as 4% of the 2007 data points; something over expectation of 1.5%. 23 out of 25 outliers results were attributed to the teacher by the head teachers. On the basis of these two findings we can say that the outliers results are not random error in the achievement data.

The parametric method identified 2 outliers in 2007 data points which makes 0.1% of the data. In a normal distribution we expect 1.5% data points in the critical region with alpha=0.05. Thus, we cannot recommend both methods of identification. There is a clear advantage to the non-parametric approach for obtaining the same piece of information.

ACKNOWLEDGEMENT

The data was collected by Aga Khan University-Examination Board in the USAID funded EDLINKS Project.

REFERENCES

1. Aly, J.H. (2007). Education in Pakistan-A White Paper. Ministry of Education, Government of Pakistan. Available at http://planipolis.iiep.unesco.org/upload/Pakistan/Pakistan%20National%20Education%20Policy%20Review%20WhitePaper.pdf (Accessed on 02/02/ 2009).
2. Andrabi, T., Das, J., Khwaja, A.I., Vishwanath, T. and Zajonc T. (2008). Pakistan: Learning and Educational Achievements in Punjab Schools (LEAPS): Insights

to Inform the Education Policy Debate. The World Bank. Available at http://siteresources.worldbank.org/PAKISTANEXTN/Resources/Publications-and-Reports/367132-1208398596098/2008LEAPS.pdf (Accessed on 06/09/ 2011).

3. Andrejko, A. (2004). Value-Added Assessment: A View from a Practitioner. *Journal of Educational and Behavioural Statistics*, 29(1), 7-9. Available at http://jeb.sagepub.com/content/29/1/7.full.pdf+html (Accessed on 31/05/2012).
4. Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd Edition, John Wiley & Sons, Chichester.
5. Biggs, J. (1998). Assessment and Classroom Learning: a Role for Summative Assessment? *Assessment in Education: Principles, Policy & Practice*, 5(1), 103-110. Available at http://www.tandfonline.com/doi/pdf/10.1080/0969595980050106 (Accessed on 03/04/2012).
6. Cook, R.D. (1997). Detection of Influential Observations in Linear Regression. *Technometircs*, 19(1), 15-18. Available at http://www.ime.usp.br/~abe/lista/pdfWiH1zqnMHo.pdf (Accessed on 25/06/12).
7. Gladwell, M. (2008). *Outliers-The Story of Success*. Penguin Books; London.
8. Grubbs, F.E. (1950). Sample Criteria for Testing Outlying Observations. *Annals of Mathematical Statistics*, 21, 27-58. Available on http://web.ipac.caltech.edu/staff/fmasci/home/statistics_refs/OutlierProc_1969.pdf (Accessed on 19/04/2013).
9. Heck, R.H. (2000).Examining the Impact of School Quality on School Outcomes and Improvement: A Value-Added Approach. *Educational Administration Quarterly*, 36(4), 513-552. Available at http://eaq.sagepub.com/content/36/4/513.full.pdf+html (Accessed on 31/05/2012).
10. Kelly, S. and Monczunski, L. (2004). Overcoming the Volatility in School-Level Gain Scores: A New Approach to Identifying Value Added with Cross-Sectional Data. *Educational Researcher*, 36(5), 279-287. Available at http://www.jstor.org/stable/pdfplus/1393325.pdf?acceptTC=true (Accessed on 31/05/2012).
11. Lalitha, S. and Kumar, N. (2012). Multiple Outlier Test for Upper Outliers in an Exponential Sample. *Journal of Applied Statistics.* 39(6), 1323-1330. Available at http://www.tandfonline.com/doi/pdf/10.1080/02664763.2011.645158. (Accessed on 02/05/2012).
12. Raudenbush, S.W. (2004). What Are Value-Added Models Estimating and What Does This Imply for Statistical Practice? *Journal of Educational and Behavioural Statistics*, 29(1), 121-129. Available at http://jeb.sagepub.com/content/29/1/121.full.pdf+html (Accessed on 31/05/2012).
13. Rizvi, N.F. (2007). Teachers as Leaders in the Pakistani Context. In Ali, S. and Rizvi, M. (Eds) *Quality in Education; Teaching and Leadership in Challenging Times*, 525-534, The Aga Khan University-Institute for Educational Development, Karachi.
14. South Asian Forum for Education Development (2012).*Annual Status of Education Report 2011*, Lahore.
15. Strutz, T. (2010). *Data Fitting and Uncertainty – A Practical Introduction to Weighted Least Squares and Beyond,* Vieweg and Tenbner, Amsterdam.

16. Tekwe, C.D., Carter, R.L., Chang-Xing Ma, Algina, J., Lucas, M.E. and Jeffrey, R., Ariet, M., Fisher, M. and Resnick, M.B. (2004). An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance. *Journal of Educational and Behavioural Statistics.* 29(1), 11-36. Available at http://jeb.sagepub.com/content/29/1/11.full.pdf+html(Accessed on 31/05/2012).
17. Zijlstra, W.P., Ark. V.A. and Sijtsma, K. (2007). Outlier Detection in Test and Questionnaire Data. *Multivariate Behavioural Research,* 42(3), 531-555. Available at http://www.tandfonline.com/doi/pdf/10.1080/00273170701384340. (Accessed on 15/04/2012).

ANNEXURE-I

PROTOCOL FOR TELEPHONE CALL TO THE HEAD TEACHERS

The Aga Khan University Examination Board

Outliers in Educational Achievement Data:
Their Potential for the Improvement of Performance

2012

Interviewer will need: Telephone number of the schools identified with outlier teachers, excel sheet to record the responses
Recording of the interview: Record head-teacher's responses on the excel sheet during the interview.

First Call:

Hello, I am Isbah Mustafa, Assistant Manager, Special Projects at The Aga Khan University Examination Board. I wanted to speak to the head-teacher of XYZ School, have I reached to the right person. I would like to speak about middle school promotion examinations in 2008 in which AKU-EB analyzed student performance and sent back the results to your school in EDLINKS Project. Are you aware of that? Now we are relooking at student data for a research and have identified ABC subject-grade result of your school as exceptionally good. First of all, I must congratulate you for this performance by your student. In the research we are trying to know the possible reasons for this exceptional performance. Can we speak about this? Should I call you tomorrow to talk about the possible reasons? In the meantime I request that you gather information about ABC subject-grade from the teachers teaching middle school in 2008 and find the time-table of 2008 of that Grade? Probably that will help us in discussing student performance.

Thanks and bye.

Second Call:

Hello, I am Isbah Mustafa form The Aga Khan University Examination Board, I called you yesterday about ABC subject-grade result of your school in middle school promotion examinations-2008. Can we speak now?

1. Were you successful in finding the time-table of that class?
2. How many sections did that class have in 2008?
3. Can you think of the reason for ABC subject-grade result of your school being exceptional and not the other 8 subject-grade results of the school?
   (Ask the following set of questions in case the identified reason by the head teacher is The Teacher. Support the investigation with new sub questions, if needed).
4. Can you identify the teacher/s?
5. What is the gender of the teacher/s?
6. Did the identified teacher/s teach other subjects/grades in middle school in 2008?
7. Can you recall the exceptional element in teaching of the identified teacher/s?
8. Can you recall school factors in addition to the teaching of the identified teacher/s?

Thank you and congratulations again on the performance of students of your school.

Bye.