



THE AGA KHAN UNIVERSITY

eCommons@AKU

Examination Board

AKU-Pakistan

November 2012

Variation in standards in Pakistani school examination boards

Thomas Christie

Aga Khan University, Examination Board, Karachi

Follow this and additional works at: http://ecommons.aku.edu/pakistan_eb

 Part of the [Curriculum and Instruction Commons](#), [Curriculum and Social Inquiry Commons](#), [Educational Administration and Supervision Commons](#), and the [Teacher Education and Professional Development Commons](#)

Recommended Citation

Christie, T. (2012). Variation in standards in Pakistani school examination boards. *In Search of Relevance and Sustainability of Educational Change : An International Conference at Aga Khan University Institute for Educational Development, November 1-3, 2012*, 24-31.

Available at: http://ecommons.aku.edu/pakistan_eb/6

VARIATION IN STANDARDS IN PAKISTANI SCHOOL EXAMINATION BOARDS

Thomas Christie, Aga Khan University Examination Board

The Research Opportunity

Some 4000 students apply for entry to AKU's MBBS degree every year although there are only 100 places available. These students are unlikely to consider themselves average. They are applying for the most sought after course in one of the three top ranked universities in Pakistan. Moreover, entry is 'needs blind' so financial considerations need not influence the decision to compete except in relation to the not inconsiderable cost of sitting the Aga Khan University Admissions Test (UAT). Student motivation can be taken to be fairly uniform. The admission process is in two stages. In Stage-1 the top scoring 10 percent of applicants in UAT are identified, 431 in 2011, 320 in 2012. At this stage the selection decision is based purely on test score: no demographic data are available to the constructors and scorers of UAT. The selected candidates then go into a detailed second phase scrutiny which comprises school performance, achievement in external examinations and two independent interviews to identify the 100 candidates who will enter the University. This research explores how the secondary school examination system conditions candidates' chances of accessing a highly desired educational opportunity. Final admission decisions will have taken account of Higher Secondary School Certificate (HSSC) and A level examinations but here we are simply concerned with how preparation for different examinations at the end of secondary school stage, Grade X for SSC and grade XI for O level, relates to performance in the UAT.

The University Admissions Test (UAT)

Table-1 seeks to establish the credentials of the University Admission Test as a measure of educational potential. There are two English essays, one double marked, to yield a total possible score of 18 derived from a common six band scale ranging from Novice with ineffective communication of ideas to Superior with fluent and natural use of English with minimal errors. The English MCQ tap reading skills. The science items derive from the syllabus elements common to the 2006 revision of the National Curriculum for grades XI and XII and the Cambridge A level topics and outcomes. This section alone deploys negative marking.

The third group, the reasoning items, are of the type commonly referred to as 'word problems' in Mathematics examinations. In both 2011 and 2012 science reasoning items have the highest correlation with the total UAT score. Not that the superiority is significant but the repetition in two different tests of the top place is a very strong indication that the UAT calls for more than the mere recall of information. Table-2 displays the first principal component of these two incompletely parallel versions of the UAT. A principal components analysis indicates how many underlying attributes are being measured but not what they are. Between 2011 and 2012 there was one change of UAT structure. While in the short writing task, "Explain the meaning of the given Urdu saying in a paragraph", only the saying differed, in the argumentative essay there was a change of format as well as topic. Both topics were well rehearsed in newspaper op-ed pieces, 2011 from economics and 2012 from medical ethics, but the economics argument offered no ideational support whereas in 2012 candidates were given a range of eight assertions to draw upon. The essays were more

homogeneous as a result and the writing loading in Table-2 reduced in consequence. The fluctuation in physics is also a function of reduced variance. The physics test in 2011 was too demanding and therefore contributed less variance to the UAT as a whole. Nevertheless although the analysis is sensitive to these variations the test does remain stable overall. Clearly one can entertain a range of hypotheses about what is being measured but the test constructors set out to capture readiness for medical education and it seems not unreasonable to suppose that the UAT is measuring the ability to think through problems encountered in short order, some of them for the very first time, just as in an Accident and Emergency Unit. However the problems set have to be sufficiently challenging to identify the top ten percent of a highly (self) selected population. These candidates are part of the intellectual capital of the nation. The creation of intellectual capital is traditionally the guiding purpose of schools and the recognition of these outcomes is the task of school examination boards working within the confines of the social expectations codified in the national curriculum.

Comparison of Academic Standards

There are three secondary school curricula in circulation in Pakistan, the National Curriculum of 2002 and earlier which was characteristic of the SSC while the candidate group under review was being educated, the revisions of 2006 through 2008 which are now recognised in part through revised question paper formats but implemented in full only by the two boards with a national remit, Federal Board of Intermediate and Secondary Education (FBISE) and Aga Khan University Examination Board (AKU-EB), and finally the GCE syllabuses examined by Cambridge International Examinations (CIE). Admission to a university medical school is conditional in all three curricula on a common subject combination, the premedical group, which comprises English, Urdu, Mathematics, Pakistan Studies, Islamiyat, Physics, Chemistry and Biology.

In the Boards of Intermediate and Secondary Education (BISE) which run the national school achievement certificates, great reliance is placed on the syllabus, or to be precise, the single text book approved for the course. Any departure therefrom can render the examination null and void. The FBISE, AKU-EB and GCE examine syllabuses and do not disallow any textbook. The choice of learning material is at the discretion of the schools. The BISE submit to further control of examination format with agreed numbers of MCQ, short response, essay and where appropriate practical marks with rather looser control of question choice while in O levels question choice varies from subject to subject from none at all in O level mathematics to a free for all in the general paper. Within these restrictions the marks of the BISE are treated as fixed and unchallengeable, grades are derived mechanically from the marks and any suggestion that they might be scaled is seen as tampering with a revealed truth. The grade boundaries, 80% of available marks down to 33% are treated as benchmarks, a kind of gold standard “untainted by values, culture or power” (Bloxham & Boyd, 2012, p.617). Certainly the candidates are essentially powerless. They cannot appeal against the marks assigned by the BISE.

There is only one softening of this absolutist stance. The award of an SSC is based on the raw mark total of eight subjects though failure in any one means failure overall. However it is well established that “decision rules which do not allow any compensation at all tend to show the worst results in terms of classification accuracy” (Van Rijn et al, 2012). This is presumably the rationale for ‘grace marks’, 5 marks out of 550 deployed over two subjects at the pass/fail borderline to the candidate’s best advantage. Otherwise standards are absolute and absolutely the standards of the

BISE. The candidates play no obvious part in establishing standards. The cut-offs are determined before a single candidate is examined and will remain fixed until IBCC chooses to change them.

How different are O level standards. They are essentially subject based and as subjects are intrinsically different fall back upon man, or at least boys and girls, as the measure of all things. The O level standard was initially that about 60% of about 15% of the population who stayed on beyond the school leaving age, should be awarded a pass. So O level standards were based on percentages of candidates, not percentages of marks. It is the candidates and of course their schools who set the standard in “a socio-cultural paradigm [which] recognises assessment as a context-dependent, socially-situated, interpretative activity” (Shay, 2004). No wonder that where parents can afford it schools flock to O level. When CIE approached IBCC for recognition of its new A* grade as equivalent to 95% of available marks, CIE recognised the inherent improbability of such a mark in any subject which leaves room for examiners’ judgement with the reassurance that this would be a rare award to a few outstanding individuals. But in the socio-cultural paradigm, the grades are in fact examiners’ judgements of what the marks may mean. It sometimes seems that in Pakistan’s private schools it is grades other than A* and A which are rare awards these days but we do not know. Contrary to the guidelines laid down in *Standards for Educational and Psychological Testing* (1999) which encapsulates the combined wisdom of the American Educational Research Association, the American Psychological Association and the National Council for Measurement in Education, CIE does not publish its Pakistani grade or score distributions. It does not reveal the social impact of its grading decisions. They are a commercial secret.

There are thus two approaches to standard setting in Pakistan, standards defined by the questions and their mark schemes which are determined in advance and standards defined by the examiners who ensure that the best candidates’ performances are deemed worthy of 95% of marks regardless of subject. The Cambridge approach betrays its origins in norm –referencing, whereas the BISE are operating a kind of absolutist criterion referencing: a competence is demonstrated or it is not, and that is that. The BISE standards are embedded in the construction of the paper.

The educational impact of the two examination systems can be evaluated by reference to AKU’s University Admission Test but only for a self-selected group of high flying candidates. Candidates know their SSC results when they make the decision to apply. Most of them will also have an impression that to get into AKU medicine you have to be a pretty high achiever, but the strength of that impression is not known and is certainly not uniform given that applicants are distributed nationwide. The conventional approach to equating results through a common calibrating instrument is to scale each Board’s distribution of scores using the means and sds of the calibrating test, in this case the UAT, but the accuracy of that procedure depends upon the representativeness of the Board samples. If the students attempting UAT are a biased sample of the students from the parent board and the bias is not a constant across boards, the measures of central tendency upon which such comparisons are based will be differently centred, rendering the comparison untenable.

Competitive Edge

In seeking an alternative to scaling, we return to the notion of “being in with a chance”. The various boards have been compared by considering whether they confer any competitive advantage on their students, that is whether students at the same grade level experience different fates. CIE with a top grade equivalent, they claim, to 95% of marks occupies a space for which the BISE do not even have a name. Their A1 grade starts at 80% of marks though the Federal Government of Pakistan makes a cash award to every candidate who scores more than 90% SSC marks. Given the

prevalence of O level students in the applicant samples, the analysis has had to play in O level territory. Candidates recorded as scoring between 90% and 100% and between 80% and 89.9% are located in 'bounded' categories, that is the top and bottom of the grades is known and what lies between the cut-offs shares a common designation in all BISE. While we cannot evaluate the general educational standards of the BISE by reference to UAT we can establish whether any board gives talented candidates receiving the same overall SSC grade a competitive edge in pursuing an academic education in medicine.

Table-3 sets out the odds on being considered for an MBBS place for candidates with SSC scores in the same grade interval. Not included are 52 applicants from the US of whom 4 were shortlisted, nor 43 applicants, none successful, taking school examinations from other countries.

There are four noteworthy features of Table-3 in relation to public examination standards in Pakistan. The first is the consistency in the pattern of outcomes between years. In every board the examination questions are different in consecutive years but the educational value of its top grades in the premedical group remains remarkably stable even though the number of top grade candidates applying for AKU fluctuates quite widely. There is clear evidence that a standard is being maintained in consecutive years, but unfortunately it is not the same standard in every board.

Karachi and the Sindh BISE are at one extreme. Not a single applicant has reached the shortlist in 2011 and only two in 2012. The Sindh boards examine half of the elective subjects in Year IX and the remaining subjects in year X. All other Pakistani Boards examine every subject twice, covering half of the subject syllabus in year IX and the rest of the subject syllabus in year X. Sindh has never offered a rationale for not falling into line with the other provinces but the implicit message that subjects are bodies of knowledge to be mugged up just for the short term is associated with extremely deleterious wash back effects. In each of the three sciences the Karachi mean UAT score is more than half a standard deviation below the overall mean of all applicants and in science reasoning it falls more than a full standard deviation short of the overall mean, a huge effect. It is not that there is no talent in the whole of Sindh. These young people have been misled by the apparent value of their BISE marks.

At the other extreme we have O levels which certainly dominate the pursuit of places. The effect of grading as an interpretative activity (Christie and Forrest,1981; Gipps,1999) is clear to see. O level borderlines are drawn by very senior examiners who have educational values, by no means always venal. They clearly have a view of A* as having much less to do with 95% than with potential to benefit from a high quality university education. It is notoriously difficult to capture that potential in a verbal description but they seem to share much the same assumptions that underlie AKU's UAT. The O level examiners are spotting talent that has flowered into remarkable educational achievement in very privileged schools and they are good at it. Look at the fate in 2012 of those they have passed over from the same privileged schools: it is no different from that of high grade SSC candidates who have had one fewer year of rather more mundane schooling. This is not the educational diet for everyone.

“A deep strategic approach to studying is generally related to high levels of academic achievement, but only where the assessment procedures emphasize and reward personal understanding. Otherwise surface strategic approaches may well prove more adaptive.” (Entwistle, 2000, p4)

That leaves the fourth feature of Table-3, the noise in between which is gradually taking on a more structured appearance. The two BISE with a national remit, the Federal Board and AKU-EB, are at

the moment clearly distinguished from the BISE with a provincial identity. The distinction is at least in part curricular. Examining across provinces the national boards have to examine not text books but syllabuses, the 2006 revision of the National Curriculum in both cases. Its emphasis on understanding and the application of knowledge is clearly paying dividends in the expansion of genuine educational opportunity beyond the deep pockets required for O level schooling. But the distinction is not permanent. The Lahore BISE is just two years behind them in adopting the 2006 syllabuses and the benefit is being further reinforced by the reorganisation of the Punjab BISE. They have adopted a new strategy of giving each Board sole responsibility for a few subjects for the whole province. The emergence of a subject focus within the Punjab Boards is an important counterpoise to the “one strategy fits all” that underlies rote memorisation. If standards are associated with subjects rather than an overall aggregate, distinctive features of subject performance come into play and examiners begin to look for subject appreciation rather than excellent recall. There is some evidence that such a shift is occurring in Lahore. The top tier are getting a sound preparation for university education. There is a qualitative difference between these students and those who fall in the 80 – 89% grade. The hope for education in Pakistan is that as the exam boards get better at identifying the skills which are relevant for higher education the students they identify will be able to sustain their performance levels for the benefit of the nation as a whole.

References

- Bloxham, S and P, Boyd (2012), Accountability in grading student work: securing academic standards on a Twenty-first century quality assurance context, British Education Research Journal, 38, 4, p615-634.
- Christie, T. And Forrest, G.M. (1981), Defining public examination standards, London Macmillan Education
- Gipps, C (1999) Socio-cultural aspects of assessment. Review of Research in Education 24, 355-392
- Shay, SB (2004), The assessment of complex performance: a socially situated interpretative act, Harvard Education Review 74 (3), p307-329.
- Van Rijn, PW, Beguin AA and HHFM Verstralen (2012), Issues and implications of high stakes decision making in final exams in secondary education in the Netherlands, Assessment in Education Principles, Policy and Practice 19(1), p117-136.

Table-1: Internal structure of the AKU University Admission Test

(Correlations above and to the right of the diagonal relate to 2011, below and to the left to 2012.)

Overall (n=4273)	Eng. Essay	Eng. MCQ	Biology	Chemistry	Physics	Science Reasoning	Maths Reasoning	2011 Total
Eng. Essay (18 marks)		0.46	0.41	0.30	0.21	0.46	0.38	0.61
Eng. MCQ (30 Marks)	0.38		0.55	0.32	0.24	0.56	0.44	0.73
Biology (20 marks)	0.31	0.58		0.60	0.51	0.62	0.41	0.82
Chemistry (20 marks)	0.24	0.31	0.55		0.63	0.46	0.24	0.70
Physics (20 marks)	0.26	0.39	0.58	0.62		0.37	0.18	0.60
Science Reasoning (30 marks)	0.32	0.53	0.69	0.62	0.61		0.60	0.84
Maths Reasoning (20 marks)	0.33	0.59	0.56	0.31	0.42	0.54		0.66
2012 Total (158 marks)	0.48	0.75	0.84	0.70	0.75	0.85	0.75	

Table-2: First principal component of UAT in 2011 and in 2012.

Subject	Marks	2011 (n=4879)			2012 (n=3196)		
		Mean	S.D.	Factor 1	Mean	S.D.	Factor 1
English Essay	18	9.23	3.30	.64	9.99	2.57	.50
English Reading	30	14.08	4.68	.73	17.01	5.78	.73
Biology	20	9.66	4.65	.84	9.91	5.22	.84
Chemistry	20	9.05	4.63	.71	7.82	4.79	.72
Physics	20	5.91	3.52	.61	7.86	4.54	.76
Science Reasoning	30	16.43	5.73	.85	17.10	5.47	.85
Maths Reasoning	20	8.67	4.49	.70	12.86	5.82	.70
Total Score	158	63.79	20.56		82.55	25.74	
% Variance explained				54%			53%

Table-3: Probability of being included in the top 10 percent of applicants to AKU's MBBS

of the top two grades of school achievement scores.

BISE	School marks 90 – 100%				School marks 80 – 89.9%			
	2011		2012		2011		2012	
	n	p	n	p	n	p	n	p
AKU-EB	5	0.20	3	0.33	58	0.10	54	0.07
FBISE	67	0.16	73	0.19	213	0.05	142	0.06
BISE Lahore	220	0.11	89	0.06	144	0.01	71	0.00
Rest of Punjab	332	0.06	133	0.11	390	0.02	102	0.02
All KPK,AJK G-B	82	0.05	79	0.08	439	0.01	498	0.01
BISE Karachi	23	0.00	5	0.00	219	0.00	140	0.01
Rest of Sindh	10	0.00	14	0.00	128	0.00	115	0.00
All BISE	739	0.08	396	0.10	1591	0.02	1122	0.02
O Levels	902	0.37	873	0.28	218	0.10	177	0.03