Capacity Review of the Punjab Examination (PEC) and the Punjab Education Assessment System (PEAS)

Gabrielle Matters and Khoo Siek Toon
Australian Council for Educational Research (ACER)

October 2012
Capacity Review of the Punjab Examination Commission (PEC) and the Punjab Education Assessment System (PEAS)

Gabrielle Matters and T. Moo Siek Toon
Australian Council for Educational Research (ACER)

October 2012

The authors acknowledge the input of Juliette Mendelovits and Leila Ismail (ACER)
## Issue and revision record

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Originator</th>
<th>Checker</th>
<th>Approver</th>
<th>Description</th>
</tr>
</thead>
</table>

This document is issued for the party which commissioned it and for specific purposes connected with the above-captioned project only. It should not be relied upon by any other party or used for any other purpose.

We accept no responsibility for the consequences of this document being relied upon by any other party, or being used for any other purpose, or containing any error or omission which is due to an error or omission in data supplied to us by other parties.

This document contains confidential information and proprietary intellectual property. It should not be shown to other parties without consent from us and from the party which commissioned it.
Capacity Review of the Punjab Examination Commission (PEC) and the Punjab Education Assessment System (PEAS)

Content

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive Summary</td>
<td>i</td>
</tr>
<tr>
<td>1.</td>
<td>Abstract</td>
</tr>
<tr>
<td>2.</td>
<td>Main Report</td>
</tr>
<tr>
<td>2.1</td>
<td>Purpose of the consultancy</td>
</tr>
<tr>
<td>2.2</td>
<td>Achievement of the terms of reference</td>
</tr>
<tr>
<td>2.3</td>
<td>Acknowledgements</td>
</tr>
<tr>
<td>2.4</td>
<td>Background</td>
</tr>
<tr>
<td>2.5</td>
<td>Outline of project</td>
</tr>
<tr>
<td>2.6</td>
<td>Pakistan and the Punjab</td>
</tr>
<tr>
<td>2.7</td>
<td>Punjab Education Sector Reform Programme</td>
</tr>
<tr>
<td>2.8</td>
<td>Punjab Examination Commission</td>
</tr>
<tr>
<td>2.9</td>
<td>Situational analysis – PEC</td>
</tr>
<tr>
<td>2.10</td>
<td>Assessment preparation</td>
</tr>
<tr>
<td>2.11</td>
<td>Logistics</td>
</tr>
<tr>
<td>2.12</td>
<td>Scaling and student-level reporting</td>
</tr>
<tr>
<td>2.13</td>
<td>Student reports</td>
</tr>
<tr>
<td>2.14</td>
<td>Sub-population reporting</td>
</tr>
<tr>
<td>2.15</td>
<td>Advice for teachers</td>
</tr>
<tr>
<td>2.16</td>
<td>Punjab Education Assessment System</td>
</tr>
<tr>
<td>2.17</td>
<td>Situational analysis – PEAS</td>
</tr>
<tr>
<td>2.18</td>
<td>Method, analysis and reporting</td>
</tr>
<tr>
<td>2.19</td>
<td>PEC and PEAS compared</td>
</tr>
<tr>
<td>2.20</td>
<td>To merge or not?</td>
</tr>
<tr>
<td>3.</td>
<td>Findings and Issues</td>
</tr>
<tr>
<td>3.1</td>
<td>Nomenclature and technology</td>
</tr>
<tr>
<td>3.2</td>
<td>Findings according to areas of operation</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Issue 1: Item and instrument development</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Issue 2: Communications strategy</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Issue 3: Calibrated items banks for assessments</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Issue 4: Governance</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Issues not directly related to this review</td>
</tr>
<tr>
<td>4.</td>
<td>Conclusions about where technical skills do not meet required capacity</td>
</tr>
<tr>
<td>4.1</td>
<td>Enabling Context</td>
</tr>
<tr>
<td>4.2</td>
<td>Technical Quality</td>
</tr>
<tr>
<td>4.3</td>
<td>System Alignment</td>
</tr>
<tr>
<td>5.</td>
<td>Options and next steps</td>
</tr>
<tr>
<td>6.</td>
<td>Recommendations for the design of a capacity-enhancing programme</td>
</tr>
<tr>
<td>7.</td>
<td>Institutional linkage and training programme</td>
</tr>
</tbody>
</table>
7.1 Institutional linkage

7.2 Training programme

7.2.1 Qualifications of proposed trainees – item writing and test construction

7.2.2 Qualifications of proposed trainees – Psychometrics, data analysis and reporting

7.3 Training modules

Annexes

<table>
<thead>
<tr>
<th>Annex</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Memorandum of Understanding</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>Terms of Reference (TOR)</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>PEC Work Breakdown Structure</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>PEAS Organisational Chart</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>Test Development Cycle</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>Extract from <em>The Punjab Examination Commission Act 2010</em></td>
<td>51</td>
</tr>
</tbody>
</table>
Executive Summary

Cambridge Education (CE) is currently implementing the three-year technical assistance package of the World Bank (WB) Punjab Education Sector Project (PESP) under the Punjab Education Sector Reform Programme I (PESRP I). This package aims to support the Government of Punjab as it embarks on developing and implementing second-generation reforms in the education sector.

CE commissioned the Australian Council for Educational Research (ACER) to undertake a capacity review of the Punjab Examination Commission (PEC) and the Punjab Education Assessment System (PEAS) to support them in presenting a detailed case to donors, seeking support for capacity-building activities for, and institutional strengthening of, PEC and PEAS.

ACER sought evidence of potential technical needs in each organisation through an all-encompassing desk review in Australia, intensive meetings conducted with senior staff of PEC and PEAS in Lahore, and the acquisition of past tests and internal working documents.

What emerged from this review was an illumination of the strengths and gaps in the capacity of PEC and PEAS to deliver the assessment programmes and associated processes for which they are responsible. The strengths of PEC and PEAS are attested to by the fact that they have already delivered examinations and results, surveys and analyses, with a small number of highly committed staff. The candidature for the 2010 Grade 5 examination was higher than 1.25 million students. Sample populations of nearly 15,000 for large-scale assessments were drawn from a student population of 7.7 million.

If it is agreed that the workplace competencies necessary for delivering high-quality assessment programmes in Punjab are deficient, urgent steps must be taken to remedy the situation. It is suggested that a small group from PEC/PEAS with appropriate qualifications and/or aptitudes be provided with training by a recognised provider outside the country.

Recommendations for the design of a capacity-enhancing programme follow.

Based on the findings of our review, we concluded that PEC is in need of capacity building in both the test construction area and the psychometrics, reporting and statistical analysis area. PEAS is better placed in terms of test construction but would still need to continue staff professional development in this area to ensure continuity of good practice. Both organisations are in urgent need of training and capacity building in the psychometrics area, especially in basic principles in psychometrics, equating techniques, and scale construction to move towards item banking.

Capacity building should include training of current staff and a plan to increase staffing level. The training should involve intensive workshops for current staff and potential recruits, as well as on-going mentoring to ensure quality assessment instruments and psychometric procedures for the calibration and scaling of tests for valid and useful reporting of results. Data collected should be used for research to inform policy and to improve practice. PEC and PEAS quantitative staff needs to be trained in research methods and statistical modelling to perform effective data analysis for reporting research findings. Training workshop should not be more than two weeks at a time. We recommend two two-week workshops in each area in the first year and at least one two-week workshop per year in the next two years. We propose and cost three training workshop modules that can be combined to design capacity-enhancing programmes. One workshop module is for the policy makers and senior officers/advisors (Module A, to be conducted in Melbourne), one for item writing and test construction (Module B) and one for psychometrics, research methodology and...
statistical modelling (Module C). Module A is of one week duration. Modules B and C are each of two weeks duration. Modules B and C can be conducted in Australia or in Pakistan.
1. Abstract

One of the fifteen milestones in the Chief Minister’s School Reforms Road Map relates to quality assessment in Punjab. This report analyses the capacity of PEC and PEAS to deliver high-quality assessment programmes and associated processes. A capacity-strengthening programme is recommended, accompanied by three costed training modules that can be used for building programme designs depending on available budget.
2. Main Report

2.1 Purpose of the consultancy

The purpose of this consultancy was to assist CE by conducting an analysis of the capacity of the Punjab Examination Commission (PEC) in delivering high quality systemwide examination and the Punjab Education Assessment System (PEAS) in conducting sample-based assessment for system monitoring and providing CE with a capacity review with recommendations for necessary training programmes to strengthen the capacity of PEC and PEAS to deliver the assessment programmes and associated processes for which they are responsible.

An adjunct purpose of the consultancy was to review the functions and operations of PEC and PEAS in the light of a possible merger between the two organisations, and consider issues that might be associated with such a merger.

In summary, the objectives of the review were:

- To assess the capacities that PEC and PEAS need in order to deliver quality assessment programmes and associated reports;
- To assess the technical needs of PEC and PEAS; and,
- To propose strategies in the form of programmatic options.

The Terms of Reference (TOR) appear as Annex 1.
2.2 Achievement of the terms of reference

<table>
<thead>
<tr>
<th>TOR Tasks</th>
<th>Progress made and agreements reached (with whom)</th>
<th>Proposed/agreed follow up (by whom and when)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desk review</td>
<td>Documentation that was relevant, current and comprehensive was studied and websites were accessed.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>We established a tentative, yet detailed, formal understanding of PEC and PEAS – their structures, mandate,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ethos and staff capacity. This led to asking more questions particularly in the areas of psychometrics and test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>construction.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>This desk review occurred in the period 4–10 May 2012.</td>
<td></td>
</tr>
<tr>
<td>In-country visit</td>
<td>Intensive participatory meetings were conducted in Lahore with the heads/directors and technical staff of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PEC and PEAS.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consultants from ACER who visited Lahore were Dr Khoo Siek Toon, Research Director, Methodology &amp; Psychometrics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and Dr Gabrielle Matters, Principal Research Fellow, Assessment &amp; Psychometric Research in the period 11–18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>May 2012.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>We assessed how formal policies and processes are enacted, institutional and individual capacity to fulfill</td>
<td></td>
</tr>
<tr>
<td></td>
<td>their mission, and constraints that reduce the capacity of the institutions to deliver their mission.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>We tested out our understandings of PEC and PEAS as formed in the desk review and adjusted these as necessary.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Our tasks were to assess the capacities that PEC and PEAS need in order to deliver quality assessment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>programmes and associated reports; assess the technical needs of PEC and PEAS; and propose strategies in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the form of programmatic options. We produced tentative findings from the needs analysis for presentation at</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the wrap-up meeting. It was acknowledged that these findings required further reflection, validation and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>extension.</td>
<td></td>
</tr>
<tr>
<td>Report writing</td>
<td>A 28-page report with five annexes was composed in the period 25 June to 10 July 2012. The report recommends</td>
<td></td>
</tr>
<tr>
<td></td>
<td>three training modules that can be used to design capacity building programs. The module options were</td>
<td></td>
</tr>
<tr>
<td></td>
<td>costed by ACER’s Senior Accountant, Ms Alison Thomas. The report was delivered to CE on 13 July 2012 (after</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CE’s approval of an extension).</td>
<td></td>
</tr>
</tbody>
</table>
2.3 Acknowledgements

We appreciate the time so generously made available to us by:

Captain Naseem Nawaz, Chief Executive Officer, PEC
Mr Malik Muhammad Iqbal Awan, Operations Manager, PEC
Dr Bashir Gondal, Senior Research Officer, PEC
Dr Abdullah Faisal, Director, PEAS
Dr Nasir Mahmood, Research Analyst/Psychometrician & Head of IT Department, PEAS
Mr Muhammad Azeem, Assessment Expert & Coordinator-1 for District-Wide Large-Scale Assessment

We are grateful to Dr Nasir Jalil, Chief Technical Adviser/Team Leader, TA for Punjab Education Sector Project, and Ms Rosalind McGinley, Senior Education Consultant, Cambridge Education, for briefing us, making arrangements for us, and facilitating access to a range of senior personnel and documents important to our work.

We welcomed the opportunity to meet Ms Umbreen Arif, Education Specialist Consultant, World Bank and Dr Farah Masood, Programme Director, Performance Management Unit/School Education Department.

We wish to place on record our deep appreciation of the many indications of friendship and acts of kindness we experienced during our visit to Lahore.

2.4 Background

Cambridge Education (CE) is currently implementing the three-year technical assistance package of the World Bank (WB) Punjab Education Sector Project (PESP) under the Punjab Education Sector Reform Programme I (PESRP I). This package aims to support the Government of Punjab as it embarks on developing and implementing second-generation reforms in the education sector.

Working with the government institutions – Programme Monitoring Implementation Unit/School Education Department, Department of Staff Development, Punjab Education Council, Punjab Education Assessment System; and donors (GIZ, UNICEF, WB and DFID) – CE is providing a diverse range of technical services to: i) enhance fiscal sustainability and the effectiveness of public expenditures including education; ii) enhance the quality of school education; (iii) improve and expand access through improvements in school participation and completion rates and reduction in gender and regional disparities; and, iv) strengthen school management and governance within the sector.

2.5 Outline of project

The project, which was undertaken in the period 4 May to 8 July 2012 (not continuously) by two senior staff in ACER’s Assessment & Psychometric Research programme, comprised three stages:

- The first stage was a desk review of available documentation to establish a detailed, formal understanding of PEC and PEAS – their structures, mandate, ethos, and staff capacity. This required
that both institutions provided key documentation in hard and soft copy to ACER. Stage 1 was undertaken in Australia. Understandings were deepened during Stage 2.

- The second stage was undertaken in Pakistan. Stage 2 consisted of intensive participatory meetings in Lahore with the heads/directors and technical staff of PEC and PEAS. These discussions were essential for assessing how formal policies and processes are enacted, the current level of institutional and individual capacity to fulfil their mission, and the constraints that reduce the capacity of the institutions to deliver their mission.

- The third stage of the review was undertaken in Australia. It involved the preparation of a report based on an analysis of the evidence collected, covering the findings from the needs assessment, and proposing a series of capacity-building strategies and activities, in the form of costed options. Stage 3 required re-visiting of understandings developed in Stages 1 and 2.

Only some of our findings were discussed by stakeholders during our visit to Pakistan. This discussion occurred at the “wrap-up” meeting, when preliminary findings were presented, feedback and advice received, and issues discussed. These conversations provided a basis for the final report. Thus the report represents a worthy working synthesis of stakeholder input, research data, donor opinion, and consultant opinion.

It is important to note that, while the functions of PEC and PEAS are very different, many technical tasks may overlap. It is also necessary to state at the outset that all the staff we worked with in the two organisations demonstrated professionalism, patience and humour, a thirst for knowledge, and an impressive commitment to valid and reliable assessment of students in Punjab schools.

2.6 Pakistan and the Punjab

The Islamic Republic of Pakistan has over 190 million inhabitants and is the sixth most populous nation in the world. The GDP per capita is US$2,800 and, although the unemployment rate is officially 5.6%, substantial underemployment exists and almost a quarter of the population is below the poverty line.\(^1\) National primary school participation rates are 62% for girls and 70% for boys.\(^2\) National secondary school participation rates are 29% for girls and 35% for boys.\(^3\)

The Punjab is a geographically and culturally distinct region that straddles the modern nations of Pakistan and India. The western part of that broader geographical entity is the Pakistan province of Punjab.

2.7 Punjab Education Sector Reform Programme

In 2003 the Punjab Education Sector Reform Programme (PESRP) was set up in response to reduced investment in education. This stemmed from a financial crisis of the Pakistan Government in the 1990s, and led to slower than expected growth in educational participation and a lack of spending on educational infrastructure. PESRP was originally envisaged to operate for only three years but remains active at the time of writing this report (2012).

A variety of initiatives have been implemented under the PESRP. These include:

- Provision of free textbooks for all students across primary and secondary schooling

---

\(^1\) CIA world fact book
\(^2\) UNESCO http://www.unicef.org/infobycountry/pakistan_pakistan_statistics.html
\(^3\) UNESCO http://www.unicef.org/infocountry/pakistan_pakistan_statistics.html

275694//1/A 18 October 2012
C:\Documents and Settings\bar57089\Desktop\MM Colour Techreport.doc
Capacity Review of the Punjab Examination Commission (PEC) and the Punjab Education Assessment System (PEAS)

- Payment of a stipend for girls to attend schools (Rs200 or $2 per month)
- Professional development for teachers
- Capacity development for education managers
- Textbook reform – to adhere to the new national curriculum introduced in 2007
- Upgrade and provision of buildings and facilities
- Upgrade of teacher employment conditions: To address teacher absenteeism, teachers are being employed on five-year contracts at specific schools rather than as permanent Ministry employees. Continuation of employment contract is subject to review. Salary increments are subject to performance reviews (previously automatically incremented annually) where qualifications and experience are more highly weighted than the opinions of the recruitment committee.
- Performance incentives for teachers
- Establishment of school councils
- Restructuring of the Punjab Education Foundation: This initiative includes an unsuccessful attempt to increase private sector investment in education. The Foundation grants financial aid to not-for-profit organisations to establish schools; pays for the tuition of students in "Foundation Assisted Schools" across 18 districts – benefiting some 470,000 students so far. It is also piloting a voucher scheme in 10,000 homes in the slums of Lahore.

The PESRP established a system whereby school inspectors monitor administrative activities such as upkeep of facilities, accuracy of the school roll, delivery of the stipend to girls, distribution of free textbooks, and teacher absenteeism.

Under PESRP the Punjab Examination Commission (PEC) was established in 2006 in order to:

...conduct standardized examinations for Grade 5 and Grade 8 to enable education managers in the province to assess learning outcomes in general as well as to identify institutions which require particular attention. PEC is responsible for the entire cycle of assessment from registration of students to setting question papers, printing and dissemination of the question papers, conduct of examinations, marking, dissemination of results and analysis of results.4

2.8 Punjab Examination Commission

As noted above, PEC is responsible for designing, implementing and reporting on assessments for all Grade 5 (primary school) and Grade 8 (middle school) students in Punjab. This includes students in private institutions offering Grade 5 or 8. The assessment domains for Grade 5 candidates are Maths, Science, English, Urdu, Social Studies, and Islamiat. For candidates in Grade 8 there is also Computer Sciences or Arabic, plus one other subject from a list of electives. The language of the assessments reflects the language of subject instruction – Urdu or English.

Apparently, there is some resistance in schools to the uptake of the PEC examinations. The following warning notice appears on the front page of the PEC website:5

Grade V and VIII examination is mandatory as per government policy for all students of public and private schools functioning within the province of Punjab. A misperception that examinations conducted by Punjab Examination Commission are optional is totally false and baseless.

4 http://www.pesrp.edu.pk/pmiu.php?v=programs
5 http://pec.edu.pk/
2.9 Situational analysis – PEC

PEC has eighteen (18) core staff – a Chief Executive Officer, four IT staff, three researchers, four staff in the operations area, and six staff in administration and finance. The PEC Work Breakdown Structure (WBS) is included as Annex 2.

Item developers, supervisory staff, and marking staff are employed at particular times of the year, and trained in huge numbers. The nature and sequence of steps in the process of acquiring staff is the same for each of those three roles: call for nominations, receipt of nominations, scrutiny of candidates, approval of candidates, submission of approvals to EDO, notification of selected staff, training of marker trainers at provincial level, and training at district.

There appears to be a view that a standard “job application” does not suit the situation for appointing item developers and marking staff. As mentioned elsewhere, it is preferable to have permanent positions for item developers. As for markers, the selection process should be more like an accreditation program.

2.10 Assessment preparation

As noted earlier there is a considerable range of assessment domains for both Grades 5 and 8. It is clear that the assessments are linked to the national curriculum. It is important, however, to note that the curricula are mostly formatted under content and sub-content headings as a set of Student Learning Outcomes (SLOs) that indicate what students are expected to know and be able to do at various grades. The National Curriculum does not contain details of content knowledge and there is only cursory attention to pedagogy. Outcomes-based education has its detractors for this reason.

Thus the assessments produced by PEC simply reflect the outcomes-based nature of the curriculum such as “Define an irrational number” (Mathematics Grade 8 curriculum) or “Explain the role of a judge in a court” (Social Studies Grade 4 curriculum).

<table>
<thead>
<tr>
<th>No.</th>
<th>Content</th>
<th>Sub-Content</th>
<th>SLO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Heredity in organisms</td>
<td>Basis of heredity (chromosomes, DNA and genes in plant and animal cells)</td>
<td>Define heredity and recognise the importance in transferring of characteristics from parents to offspring</td>
</tr>
<tr>
<td>32</td>
<td>Sources and effects of heat and energy</td>
<td>Sources and effects of heat</td>
<td>Describe the sources and effects of heat</td>
</tr>
</tbody>
</table>

The point being made here is that teachers are required to teach “sources of heat and energy”, item writers are required to develop items to bring forth “what students know and can do” about sources and effects of heat, and all of this in the absence of details about the topic. Where extra resources are provided for
teachers they have a better chance of understanding the declarative knowledge required of their students and teaching accordingly. If teachers and item writers (and markers) are not experts in their subject (which may often be the case from the information gathered), it is difficult to believe that the chain from intended curriculum to enacted curriculum to assessed curriculum is not broken.

PEC uses a matrix of curriculum content and SOLO (Structure of Observed Learning Outcomes6) as an assessment blue-print. Each cell in the matrix is given a weight and the associated assessment content follows accordingly. This adds yet another layer to the system for classifying items.

Assessments that are in subjects taught in both English and Urdu have the same content. There is no information, however, on the translation procedure or linguistic quality assurance. In fact, the English versions of the tests that are available contain grammatical and spelling errors (the science papers are particularly problematic), indicating that they are not professionally edited. Presentation is a validity issue not merely an issue of aesthetics.

In any assessment delivered in multiple languages it is very important that high quality translations are accomplished. This is a necessary but not sufficient condition for psychometric equivalence of the tests. The absence of this sort of equivalence is a threat to the reliability of the results. If, for example, the grammar within the assessment is compromised in any language version this places an additional cognitive burden on students – indeed, it may lead to a misunderstanding of the intent of the question through no fault of the student.

2.11 Logistics

Assessing the enormous numbers of students across thousands of schools is a huge undertaking. For assessment security the tests must all be delivered on a single day (per grade and subject). The transport of the assessment materials alone is a considerable logistical feat. Completed assessments are returned to the PEC where responses to the open-ended questions (OEQs) (40% of the paper) are marked and response sheets for multiple-choice questions (MCQS) are scanned. PEC takes only six weeks to mark the assessments, scale the scores and report the results.

While PEC is to be congratulated on the efficiency of their operation, a question arises about the relative intensity of the technical arm of the operation and the potential for strict operational deadlines to compromise the quality of the examination development process.

By comparison, the international PISA7 project allows national centres 12 weeks from the time of testing to scan, mark and verify the data from 5,000 students for submission to the international contractor. Cleaning and scaling of data take yet another month. These timeframes allow for meticulous verification checks including cleaning the data twice independently and reconciling the separate data files.

2.12 Scaling and student-level reporting

Assessment data are scaled by PEC staff using the Rasch model and transformed onto a scale. Although the results for all districts can be obtained from the PEC website, no information is available to assist users

---

6 Developed by John Biggs in 1982, the SOLO taxonomy is a model that describes levels of increasing complexity in students’ understanding of subjects.

7 OECD’s Programme for International Student Achievement
with the interpretation of the results. In fact the technical information that is publicly available is scant. For example, there is no indication of the mean of the reporting scale, or the variance. There is no advice whatsoever on how the “fail” result is applied. It is in fact not derived systematically but set post hoc to ensure that a certain percentage of students pass a particular subject in a particular year. Standards of achievement need to be pre-set and maintained over time so comparability over the years is possible, particularly when statements need to be made about trends in the “pass” rate. There is no evidence that tests are equated over time which means performance of students across years cannot be compared.

The several test forms in each subject administered each year have not been equated onto the same scale. The test forms administered would almost certainly vary in difficulty. The assumption is that the test forms are parallel forms and are equal in difficulty. Such an approach is arbitrary and two students with the same ability could pass or fail the examination depending on the difficulty of the test form sat.

2.13 Student reports

PEC provides individual student reports which grade the student’s performance on each of the expected SLOs by subject. It is not clear, however, how each competency is measured. In particular, given the large number of competencies that are measured in each assessment, it is difficult to determine how reliably each SLO is measured.

2.14 Sub-population reporting

PEC produces “secondary analysis” reports annually, and these are all available online.8 These reports are entirely descriptive, focusing on comparative performance between districts, school type and gender. If the results are used for teacher incentives and rewards, then it would not be fair in comparing schools without taking into consideration factors such as socio-economic status (SES). It is well documented that students with higher SES backgrounds tend to have higher levels of educational achievement. Poorly performing districts or school types might be performing poorly because of the SES of their student intake rather than relative shortcoming of their teachers and pedagogy. Further analyses including such inferences would be useful for policy makers and the general public.

2.15 Advice for teachers

PEC has published two volumes of advice for teachers – one at Grade 5 and one at Grade 8. The main purpose of the advisory document is to demonstrate to teachers the degree to which SLOs (as determined by the curriculum) were attained by their students in the year just passed, the common misconceptions that were revealed in student responses, and the growth in achievement between Grades 5 and 8. This is done largely through examining percentage correct on a released item set. Commentary about each item is included and there are “teacher reflections” to consider and prompt discussion. Items are also mapped to the SLOs, the underlying structure of the curriculum. Section 6, supposedly on links to the curriculum and text books, merely signals that work is in progress. Section 7 on exam technique is useful for teachers (e.g., Tell students: “Don’t tick more than one multiple-choice response”). This information is also dealt with in the exam instructions. While this publication could be improved (say by completing Section 6 and providing information on growth in content areas rather than only individual items), there is a solid foundation on which to build teachers’ knowledge about assessments, and measures should be taken to

8 http://pec.edu.pk/publications.html
ensure that all teachers are using and following this manual so that there is consistency of approach to examinations. PEC is to be commended on the thoughtful design and approach taken.

2.16 Punjab Education Assessment System

The Punjab Education Assessment System is part of the National Education Assessment System (NEAS). PEAS is charged with monitoring the educational levels of the Punjab student population through assessment surveys.

2.17 Situational analysis – PEAS

PEAS has eleven (11) core staff: a director, a deputy director, one assessment expert, five subject specialists, one system analyst, and one finance person.9

Refer to Annex 3 for the PEAS Organisational Chart.

PEAS administered assessment surveys to students in Grades 4 and/or 8 in 2005, 2007, 2008 and 2011. PEAS worked closely with NEAS to produce these system-wide tests up to 2008. The 2011 survey of Grade 4 students’ achievement in mathematics, Social Studies and Urdu was the first independent design and administration by PEAS.

2.18 Method, analysis and reporting

Known as the District-Wide Large Scale Assessment (D-WLSA), the 2011 survey of Grade 4 students’ achievement in mathematics, social studies and Urdu, assessed approximately 14,000 students in 935 schools.

The report for the 2011 D-WLSA represents a significant advance on previous reports in that it contains some useful technical information about the research methods used by PEAS.

Aspects of instrument development are described in the report: there is what appears to be a well considered process of a small-scale “try-out” of the items, followed by revision and a statistical pilot of sufficient size to meet the requirements of the types of psychometric analyses required in the developmental stage (e.g., to determine differential item functioning by gender or language of test).

It is noted in the 2011 report that selection of students within schools is random but there is no information on the school sampling frame, school sampling methods (How was the probability of selection determined? Were replacement schools used?), sampling outcomes (How many schools did not respond?), sample weighting, or non-response adjustments. These are very important aspects of complex survey methods as they assist in determining the extent to which the data are reliable. In fairness there appears to be a technical report planned (there is an empty link on the website) so these issues may be addressed in that document.

Sufficient technical detail is provided to suggest that some methodological improvements might be considered. With a complex survey the use of plausible values and replicate weighting methods are

9 http://www.peas.gop.pk/PEASTeam.htm

275694//1/A 18 October 2012
C:\Documents and Settings\bar57089\Desktop\MM Colour Techreport.doc

10
commonly employed to avoid statistical bias and attenuation. There is no evidence that these methods were used.

There is insufficient technical detail for the reader to understand fully the basis of some claims that are made. First, there are comparisons between performances across the three domains. Statements such as “Overall students performed better in Mathematics than Social Studies and Language [Urdu].” While this is couched as an expert opinion it is clearly based on empirically derived proficiency levels and, as the three subjects appear to be independently scaled, it is difficult to understand how a direct comparison could be made.

Another concern is the comparison of scores over time. Achievements in mathematics, social studies and Urdu were surveyed in 2005, 2006 and 2008, and the report claims that performance improved in 2011. However, there is no indication that the assessments are linked and equated. Comparisons across years are not possible without the scales being equated. No public databases exist so that independent researchers can check the veracity of the analyses reported by PEAS. A set of databases containing de-identified student item scores, background variables and scaled scores would also be helpful to interrogate the test design; for example, to answer questions on linkages across booklets. Another very useful function of a public database is that it allows researchers to extend analysis beyond the basic descriptive statistics reported; for example, by using simple multiple and multilevel regression techniques to model effects on educational outcomes.

A central goal of PEAS is “to identify factors influencing students’ achievement”. In the 2011 report this is addressed by cross-tabulation of single contextual variables such as gender with student achievement. Unfortunately, the sources of information for the analyses presented under the rubric of learning environment are not well described. It is difficult to ascertain whether the data source is teacher self-reports or student reports of teacher behaviour. If the data were from teacher self-reports, with the students randomly selected from the school and only one teacher from each school answering the questionnaire, the link between teacher and student would be difficult to establish.

Nevertheless, the cross-tabulations are of interest because they indicate associations between context and educational performance. These simple cross-tabulations are very limited and it is difficult for the reader to obtain a clear picture of what is taking place. Much effort has been put into the report to aid interpretation of the results and to advise on best practice – this is the strength of the report’s design. More can be made of these contextual variables; for example, multi-level regression analysis controlling for socio-economic status may have been more revealing.

2.19 PEC and PEAS compared

Tables 1 and 2 represent our understandings, in summary form, of the nature and functions of PEAS and PEC, on the basis of documents supplied to us by PEC and PEAS, conversations with staff of PEC and PEAS, briefings by CE, and accessing information of our own accord (e.g., information on other education systems). It is on the basis of these understandings and evidence gathered that we have made our
judgments about the existence of strengths and gaps in the capacities required of PEC and PEAS to deliver quality assessment programmes and associated reports.

Table 2.1: Status of PEAS and PEC

<table>
<thead>
<tr>
<th>PEAS</th>
<th>PEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEACE13 established in 2002 as implementation partner of NEAS14 in conducting large-scale national assessments</td>
<td>PEC established under an Act of Parliament in 2004 for the “improvement of the examination system and for conducting examinations of elementary education and to provide for ancillary matters”15</td>
</tr>
<tr>
<td>Restructured in 2005 and designated as PEAS, with responsibilities for assessment activities at provincial level</td>
<td>Conducted assessment studies in 2005, 2006, 2007, 2008 and 2011 at Grades 4 and 8 (these included student questionnaires)</td>
</tr>
<tr>
<td>Role changed after the 18th amendment16</td>
<td>PEAS now works independently at provincial level</td>
</tr>
<tr>
<td>Remit refers to elementary, secondary and higher-secondary levels of education</td>
<td>Remit refers to elementary level of education</td>
</tr>
<tr>
<td>Any kind of large-scale assessment other than achievement testing for grading and promoting students</td>
<td>Examination for whole population studying that subject for reporting results of individual students</td>
</tr>
<tr>
<td>Sample-based tests providing research data</td>
<td>Examinations only</td>
</tr>
<tr>
<td>Large-scale assessment describes the level of achievement of a whole school system or a clearly identified part of it</td>
<td>Examinations are for reporting and certifying individual achievement.</td>
</tr>
<tr>
<td>Varniously referred to as survey, large-scale assessment, test</td>
<td>Varniously referred to as examination, large-scale testing.</td>
</tr>
<tr>
<td>Not high-stakes assessment</td>
<td>High-stakes17 assessment</td>
</tr>
<tr>
<td>Standardised instruments</td>
<td>Instruments are not standardised</td>
</tr>
</tbody>
</table>

Table 2.2 summarises the differences in assessment nature and delivery of PEAS and PEC.

Table 2.2: Nature and delivery of assessments, PEAS and PEC

<table>
<thead>
<tr>
<th>PEAS</th>
<th>PEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large-scale assessments</td>
<td>Examinations</td>
</tr>
<tr>
<td>Rationale</td>
<td>Assessment of “health” of system</td>
</tr>
<tr>
<td>Purpose</td>
<td>Assessment of achievement of individual students</td>
</tr>
<tr>
<td>Who is assessed?</td>
<td>To provide feedback to stakeholders (policy makers, schools, the wider community)</td>
</tr>
<tr>
<td></td>
<td>To certify and select students (in the case of PEC examinations, to promote students)</td>
</tr>
<tr>
<td></td>
<td>Sample of students at a particular age</td>
</tr>
<tr>
<td></td>
<td>All students who wish to take the particular</td>
</tr>
</tbody>
</table>

13 Provincial Education Assessment Centres

14 National Education and Assessment System

15 The Punjab Examination Commission Act 2010 (Act XI of 2010), passed by the Punjab Assembly 14.7.2010

16 The 18th Amendment to the Constitution of Islamic Republic of Pakistan: Federal and Provincial Issues

17 In the sense of being associated with important consequences for the student – here being promoted or receiving a certificate
### Table 2.1

<table>
<thead>
<tr>
<th>PEAS</th>
<th>PEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>or grade level</td>
<td>examination at the grade level set</td>
</tr>
<tr>
<td>Often referred to as …</td>
<td>Assessment for learning 18</td>
</tr>
<tr>
<td>Here referred to as …</td>
<td>Assessment for research</td>
</tr>
<tr>
<td>Nature</td>
<td>Diagnosis</td>
</tr>
<tr>
<td></td>
<td>Promotion</td>
</tr>
<tr>
<td>What is tested</td>
<td>Proficiency</td>
</tr>
<tr>
<td></td>
<td>Achievement</td>
</tr>
<tr>
<td>Procedure</td>
<td>Standard setting</td>
</tr>
<tr>
<td></td>
<td>Standard testing</td>
</tr>
<tr>
<td>Alignment with …</td>
<td>Curriculum framework</td>
</tr>
<tr>
<td></td>
<td>Test blue-print</td>
</tr>
<tr>
<td>Scoring</td>
<td>Sophisticated techniques</td>
</tr>
<tr>
<td></td>
<td>Marking scheme (criteria)</td>
</tr>
<tr>
<td></td>
<td>Simple process (pass mark)</td>
</tr>
<tr>
<td>Backwash effects (on curriculum, pedagogy etc.)</td>
<td>Unlikely because no information at the level of the individual student</td>
</tr>
<tr>
<td></td>
<td>Potentially negative (prospect of failure which can lead to early drop-out)</td>
</tr>
<tr>
<td></td>
<td>Data not suitable for research purposes: Teacher incentives, for example, are linked to exam results increasing incidence of malpractice</td>
</tr>
<tr>
<td>Monitoring potential</td>
<td>Must be designed with monitoring in mind (e.g. equating methods)</td>
</tr>
<tr>
<td></td>
<td>Exam questions and candidate populations change from year to year</td>
</tr>
<tr>
<td>Cost</td>
<td>Affordable in terms of amount of information that can be fed back into the system</td>
</tr>
<tr>
<td></td>
<td>Expensive (because of high-stakes nature)</td>
</tr>
</tbody>
</table>

#### 2.20 To merge or not?

As well as serving as a collection of summary statements, the tables above emphasise differences between PEC and PEAS and identify commonalities at the same time. The scope of the work that PEC and PEAS undertake is huge and the volume of information about each of them is vast. Judgments about a possible merger between the two organisations need to be made respectfully and on the basis of evidence and systems in other parts of the world. Where there exists a statutory authority and an education department there exist the issues of competing status, potential duplication of duties, communication channels and so on.

On the surface, the overlap between PEAS and PEC seems to be large – mainly due to the use of “fuzzy” nomenclature and terminology. The sets of skills required by both organisations definitely overlap (after all both organisations are carrying out educational assessment and in the same province). Some of the skills are interchangeable between organisations. It is clear to the authors, however, that the purposes of the organisations are different. These differences are best captured by the terms that were coined at our meetings in Lahore and relate to the purposes for which the assessments are used: PEAS does assessment for research and PEC does assessment for reporting (i.e. certifying). The principles for any good assessment are the same regardless of the purpose of the assessment.

Issues that might be associated with a merger of PEC and PEAS are considered later in this report.

---

18 Term coined by the Assessment Reform Group in England in 1999 to highlight the value of formative assessment
3. Findings and Issues

3.1 Nomenclature and technology

Attempts to create a comprehensive and accurate picture of PEC and PEAS were hindered by the nomenclature that is used. PEC is a commission; PEAS is a system. Both are organisations or institutions and PEAS is also a faculty. PEC’s title includes the word examination whereas PEAS’s title includes the word assessment. This is not merely a semantic issue; it gets in the way of examining the implications of a merger between the two organisations.

For PEC and PEAS, as for organisations the world over, the assessment terminology that is used in documents and conversations is not consistent across and within organisations. A distillation of the scholarly literature results in the following definitions, which will be used for the purposes of this review.

Examinations (more specifically, external examinations) are subject-specific, based on a formal document such as a curriculum framework or syllabus, and set by an external group of teachers or examiners. Results are reported for individual students.

External examinations are generally developed and administered by independent curriculum and/or certification authorities that are external to schools. They are developed and administered in a secure environment under controlled conditions. External examinations are typically marked in a centrally managed marking operation. Markers receive training in applying rubrics (or marking schemes) that are provided by the examining body. Aggregation of data and production of overall grades is a statistical operation and final grades are usually published some time after the completion of the examination(s). As well as being trained before they start marking, markers are also monitored during the marking operation.

Large-scale tests/assessments are usually developed and implemented by independent corporate or educational organisations that are remote from school sites. They are usually administered and marked in much the same way as external examinations. Multiple-choice questions are optically scanned and marked according to a key. For open-ended questions, markers are typically drawn from the ranks of the teaching profession although there are often concerns about this aspect of teachers’ assessment skills. Marking operations involve marker training based on the provision of marking schemes and exemplar materials that demonstrate the range of ways in which the targeted area of learning may be demonstrated in the evidence: candidate responses.

Within PEC and PEAS (as is often the case in other contexts as well), these distinct terms are used interchangeably: assessment, testing, tests, examinations, surveys, large-scale assessments, and system-wide assessment, to name a few.

Regardless of interchangeable terminology and in order to reduce confusion in what follows, the reader is best advised to think of the assessments that are carried out by PEC as examinations and the assessments that are carried out by PEAS as surveys.

3.2 Findings according to areas of operation

We examined the following areas of operation that were supplied to us, each to a lesser or greater extent:

a. Curriculum frameworks (assessment objectives) and test item accuracy, validity and reliability
b. Procedures and practices for pre-testing accuracy and reliability and developing item banks
c. Ensuring exam/test security (during, before, and after the exams/tests)
d. Scoring and coding of tests, test calibration, scale construction, data analyses and communicating results to various audiences

e. Communication of information at various stages to (potential) test-takers, looking at relevance, completeness, and timeliness

f. Helping schools and students familiarise themselves with test content and test-taking requirements and procedures in advance of test administration

g. Preparation and training of district education administrations for administering the tests

h. Test administration procedures and practices by district education departments (all stages)

i. Possibilities for, and feasibility of, third-party monitoring of key stages

j. The possibility of a merger between PEC and PEAS and potential attendant issues.

Areas a, b and d can be categorised as test construction and psychometrics; Areas c, g, h and i as operational; Areas e and f as communications; and Area j is a special area.

Table 3.1 lists the areas of operation above and, where sufficient information has been gathered, we provide comments and identify needs.

Table 3.1: Discussion points

<table>
<thead>
<tr>
<th>Discussion points Scope #</th>
<th>Area of operation</th>
<th>Comments and needs identified</th>
</tr>
</thead>
</table>
| 1                          | Curriculum frameworks (assessment objectives) and test item accuracy, validity and reliability | Settle on one type of item classification for the test construction matrix after considering the need for Bloom and SOLO and SLOs
Training required: Developing test blueprints (Test Specifications) with reference to curriculum.
Test construction as a process of design not just the application of a set of mechanical rules
Test construction does not necessarily require the same set of skills as does item writing.
Not all teachers make good item writers.
The statements above have implications for training and staffing.
Training required: Item writing, test construction, item analysis (e.g. IRT\textsuperscript{19}), how to recognise and train a good item writer |
| 2                          | Procedures and practices for pre-testing item accuracy and reliability and developing item banks | Need wider item selection for balanced test and multiple forms. Over time, develop a calibrated item bank. Item banking creates permanent rather than seasonal work for test developers as their time is directed to future tests and the “populating” of a calibrated item bank. This would go some way to invigorating the working milieu in PEC/PEAS.
Panelling (item review) is content validation and requires attention to technical, linguistic and editorial aspects of the items.
Extra panels at different stages of the Test Development Cycle (see Annex 4) would provide extra scrutiny of the papers.
Trialling and item analysis must occur before test construction.
Analytic marking schemes for open-ended questions must be |

\textsuperscript{19} Item Response Theory combines psychology and mathematics in determining the probability \( p \) that an examinee with ability \( \theta \) correctly answers an item. Modern test theory (after Rasch) estimates item parameters and person ability, placing items and examinees on the same scale.
<table>
<thead>
<tr>
<th>Discussion points</th>
<th>Area of operation</th>
<th>Comments and needs identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope #</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 3 | Ensuring exam/test security (during, before, and after the exams/tests) | Excellent product of a just-in-time strictly enforced timetable Possibly unbalanced compared with attention given to test construction |
| 4 | Scoring and coding of tests, test calibration, scale construction, data analyses and communicating results to various audiences | Training required: Psychometrics, Item Response modelling, statistical modelling. Comparability of results across tests requires the application of test-equating techniques: common-item equating, common-person equating techniques. Comparability across grades needs vertical equating. Comparability within the same grade across years or across test forms needs horizontal equating. Other skills: setting benchmarks, reporting, item banking, data display and report writing for various audiences; statistical modelling such as multilevel modelling, structural equation modelling. Present analyses are elementary despite energy and commitment of PEAS. |
| 5 | Communication of information at various stages to (potential) test-takers, looking at relevant, completeness, and timeliness | The marking criteria for the examinations must be decided at the time the item is written and panelled. The polished marking criteria (rubric) must be available even before the marking proper but as a tool for marker training (and later monitoring). |
| 6 | Helping schools and students familiarise themselves with test contents and test-taking requirements and procedures in advance of test administration | It is vital to familiarise students with test content and conduct to ensure that a lack of “test-wiseness” does not disadvantage them |
| 7 | Preparation and training of district education administrations for administering the tests | The training of exam supervisors is appropriate but the model for selection and training of markers is too simplistic (see mentions of accreditation and monitoring elsewhere in this report). |
| 8 | Test administration procedures and practices by district education departments (all stages) | Consider balance in technical and operational activities. To what extent do strict time-lines affect quality of test and its marking? Also, less testing would save money not just from the point of view of test development but also of test administration, the cost of the latter being high due to huge student population in Punjab. |
| 9 | Special mention (Governance) | Reconceptualise structure and functions of PEC and PEAS. Refer to Figure 1 and associated discussion later in this report. |
| 10 | Capacity building and enhancement | Commit to long-term capacity building in the form of “built-in” activities. Create permanent positions for Head of Test Development, accountable position (sign off of all stages of test development) Undertake research on system using multivariate analyses (not simply bi-variate analyses), thus providing useful information for stakeholders Consider using PEC student results for a sample of students as outcome measure for part of PEAS research. This would save the
### 3.2.1 Issue 1: Item and instrument development

In order to use data obtained from learning assessment programmes, whether monitoring learning outcomes at Grades 4 and 8 at district and province level over time or making decisions about promotion of students out of Grades 5 and 8, high-quality assessment instruments have to be prepared.

The box labelled “Refine” in the Test Development Cycle diagram (Annex 4) is embellished to indicate the iterative nature of processes here (e.g., panelling of items). The validation panel that traditionally occurs at item level is situated here.

The process of item review (or panelling) is appropriate at test level as well as item level, and collections of reviewers require different sets of skills. Effective remits for other panels at the test level (e.g. in the box labelled “Reflect”) are Equity, Editorial, and Space/time/difficulty.

Test developers need to have information regarding the functioning of the items that make up their tests so that they can evaluate the effectiveness of the testing situation and, more importantly, so that they can improve the quality of their instruments for future occasions. Item Response Theory (IRT), particularly the Rasch models for uni-dimensional measurement, is considered to be “state-of-the-art” for underpinning the testing work. Some personnel in Lahore are currently using IRT, some are familiar with its application, and others wish to learn about it. The training programme suggested below includes activities on IRT and test and item analysis.

The items should be trialled on carefully selected samples of students (about 300 students attempting each item would be sufficient for the trial) and the data used to calibrate the items and operationalise the underlying scale. It might be possible to trial the items at least two years in advance (starting immediately) and build up the psychometric properties of the items in this way.

Preliminary item characteristics obtained from pilot testing (or “trialling”) through the application of classical and modern test theory would then be used to decide if items required revision or outright rejection (industry standard has approximately 30% of items rejected at this stage). Items accepted immediately after their development and items accepted after revision would then be available for a selection to be made for inclusion in examinations/tests each year.

Training materials not sufficient in their present form. The training programme appears to be a series of presentations starting with the topic of “assessment methods” and then MCQ writing rules sourced from a textbook and there are few if any resources for writers of OEQs which make up a significant proportion (40%) of the exams.

### 3.2.2 Issue 2: Communications strategy

The assessment reports we studied represent a considerable investment in time and expertise. The information in those reports could be distilled for various audiences. Reports could be provided to schools...
on main trends, sub-group differences, strengths and weaknesses of students, and evidence of students’ flawed thinking.

It would be a good idea if, within a new staff structure as envisaged by the authors of this report, there is a position with responsibilities for communications and liaison. The following strategies, among others, are recommended.

The use by education stakeholders of data generated from the assessments should be fostered by ascertaining what data are meaningful to the various stakeholders, optimising the use of data by schools, using data to support the professional development of teachers, and making available materials from PEC/PEAS in useable form for government (i.e., policy makers), agencies, districts, and schools.

Information from both trial and “live” tests should be incorporated into a test analysis report, which would be prepared each year by PEAS. In any given year, a subset of the tests (subjects) could be selected for both analysis and reporting, and scores on these tests collated at individual item level.

There are models from elsewhere on how to communicate information on student and item behaviour to different audiences.

It is vital that accountability is demonstrated by schools, districts and PEAS/PEC, otherwise the assessment programme will be unproductive and undervalued. One example of a break in the accountability chain is the reported under-utilisation (or complete lack of utilisation) of detailed reports on tests, provided to districts/schools by PEC/PEAS. Among a range of possible reasons for this is a lack of skill in interpretation of data from the report for classroom remedial practice; another is that districts simply do not take the reports seriously.

### 3.2.3 Issue 3: Calibrated items banks for assessments

A calibrated item bank is one element of international best practice in assessment and an aspect of good data management. If calibrated item banks were to be developed, maintained and used in Punjab, this would allow for on-going professional training and skills development of staff and help to ensure quality of outputs.

Calibrated item banks hold together the various elements of an assessment system so the issue above is embedded in the discussion below.

A calibrated item bank (i.e., a collection of items measuring the same construct on the same measuring scale) has the potential to provide a source of items from which a selection can be made for use as required. Tests with known characteristics can be quickly assembled using the known item calibrations and histories of item use.

A calibrated item bank would provide the basis for designing the tests for the Punjab assessment programme. Items would be written, trialled, calibrated onto a common scale and deposited in the bank together with their history and item characteristics.

A calibrated item bank would provide substantial saving in time and resources in the annual test development cycle. To construct a test, test developers would merely pull out items with required content and measurement characteristics from the bank rather than commencing a new cycle of item writing and trialling. This assumes that, at some stage, sufficient items have been deposited in the bank. Item writing
would be continuous throughout the year not merely aimed at the upcoming test for a particular year. Another assumption is that the items are kept secure and not published. This assumption makes it more appropriate for PEAS than for PEC.

As discussed above, PEAS should have an item bank for use in constructing large-scale assessments. The aim would be to have a Rasch-calibrated item bank, for which there would be one measurement scale per subject area. The scale would be constructed based on calibrated and equated tests from Grades 4 and 8.

The item bank should be developed first for Language (English and Urdu) and Mathematics while progress is being made on other subject areas. Test developers would be able to access the item bank and develop potential tests independently of IT staff. This means that they would be able to type in the characteristics of the items they require for their tests, have the items appear on the screen, select the items they require to meet the psychometric properties of the tests they are constructing, format the tests on screen and print the master copy of the test.

The item banks could be partitioned so that items that have already been used (extant items) can be catalogued and deposited in the bank. In the long term the PEAS item banks could allow the tracking (monitoring) of the progress of individual cohorts of students (and sub-groups thereof).

Some items could be taken out of the secure bank and used when training item writers. Others could be used as the basis of feedback to teachers.

A calibrated item bank with items located on the same scale would provide not only a theoretically defensible method for designing tests for the Punjab assessments in the immediate future but also a rigorous basis for interpreting student performance on those tests and providing feedback to government, schools, teachers, the education community generally and, if desired, the wider community.

3.2.4 Issue 4: Governance

The Punjab Examination Commission Act 2010 called for the establishment of a commission for improving the examination system, conducting examinations of elementary education, and providing for ancillary matters (see Annex 5 for extracts from Act XI of 2010).

Such a policy umbrella provides the opportunity to explore the model for governance of learning assessment in Punjab to an extent not achieved by present arrangements. The status report delivered at the “wrap-up” meeting included a model for a new agency based on an amalgamation of PEAS and PEC, thus signalling the need for a fresh approach to roles and functions of the new agency and its staffing and resource requirements. The new agency with a new name would be responsible for all assessments (examinations, tests, surveys etc.) and all evaluation, research, development and monitoring. It is acknowledged that re-naming PEC is not a simple matter, given that it was set up under an Act of Parliament. Nevertheless, a new agency that was perceived to be the product of the absorption of PEAS into PEC would not be conducive to high productivity or high levels of staff morale, for obvious reasons.

Despite possessing a kernel of expertise in test development and analysis, a demonstrated capacity to undertake research and produce reports, and a passion for their work, we believe that PEAS is lacking an identity (as attested to by many of their comments). The new agency would be the destination of the expertise presently residing in PEAS thus complementing the current expertise in PEC.
The emphasis in Figure 3.1 is on the two boxes on the right-hand side of the diagram and the box on the bottom left-hand side. The tiny label on the latter reads “Exams/Psychometrics & Reporting”. The Exam/Test Development Unit and the Exams/Psychometrics & Reporting Unit (for want of better names) would absorb all positions in the PEC and PEAS staff structures that are related to the technical side of the assessment function. The tentatively named Research & Development Unit would absorb all positions in PEC and PEAS that are related to evaluation, research, development, and monitoring.

![Diagram](image)

**Figure 3.1:** Re-conceptualisation of functions (re-engineering PEC)

An efficient use of funds might be to set up a unit rather than a new entity and locate this unit within PEC. This unit would have clearly stated responsibilities, degree of autonomy and accountability, and be led by a person with expertise and experience in educational measurement. The Unit (not called PEAS) would be accountable for open and transparent processes, demonstrate best practice in large-scale assessments, and operate in an environment that is conducive to high productivity. The PEAS Unit could include all the personnel required to fulfil the functions of instrument development, data management, item banking, data collection, sampling, analysis and reporting, communications, and desktop publishing.

### 3.2.5 Issues not directly related to this review

Which subjects and when? – reconsideration of the testing cycle

Punjab presently has sample-based system-wide assessments that test the same subjects at Grades 4 and 8 as does PEC’s examinations at Grades 5 and 8. This, therefore, involves more than one assessment in the same subject. In other places this might be considered excessive.

An alternative approach worthy of consideration is to maintain the essence of the PEAS programme but have tests delivered on a defined schedule – say on a triennial basis. Another possible approach is to test the same subjects as at present on a staggered basis but only one subject per year per grade. These subjects could be different, commencing with say Mathematics for Grade 8 and English for Grade 4 in 2013. A third possibility is to proceed with Option 2 for Grade 8 but for Grade 5 replacing subject-specific tests with tests of basic skills (literacy and numeracy). This sort of testing has become common in (developed) countries. Long-term costs to the system would be significantly lower than costs for the current state of affairs and would provide more and different information.

Reconsidering which subjects and when may lead to less test development activity and therefore less expense generally. It may also mean less test taking and therefore less stress on students (and teachers).
4. Conclusions about where technical skills do not meet required capacity

The position of the reviewers on the predominant issues was consolidated by using a framework with the following components: Enabling Context, Technical Quality, and System Alignment. The conclusions we reached form the basis of recommendations for the design of a three-year capacity-enhancing programme for staff of PEC and PEAS.

What follows is an elaboration of each of the components of the framework that are directly relevant to a discussion about the capacity of PEC and PEAS.

4.1 Enabling Context

The World Bank uses the term “enabling context” to refer to the broader context in which assessment activity takes place in a country and the extent to which that context is conducive to, or supportive of, the assessment activity.

Critical to transforming the environment for examinations and district-wide large-scale assessments in Punjab is the organisational development of PEC/PEAS; specifically:

- Governance structure for assessments
- Institutional development of PEC and PEAS or an amalgamation of the two
- Role clearly nominated for each accountable officer (e.g., on signing off final form of tests)
- Provision of equipment and computer/electronic hardware to perform activities associated with assessment for research and assessment for reporting
- Increased staffing levels and appropriately qualified staff.

None of these are eligible for inclusion in the costings for capacity building as they are not “technical skills” per se.

4.2 Technical Quality

Existing assessment practices would reflect best practice if the following received attention:

- Item writing
- Reducing the surplus categories in assessment frameworks
- Exploration of the Open-Ended Question (OEQ) format
- Strengthening the MCQ format
- Test construction (as distinct from item writing)
- Inserting an editing loop in the test development cycle thus enhancing validity and ensuring quality
- Inserting a language loop in the test development cycle to ensure the appropriate use of language on tests
- Devising a test construction matrix to illustrate range and balance of skills and difficulties over the complete test
- Trialling of items before inclusion in a test
- Equating of tests across forms and across years

---

20 System assessment and benchmarking for education results (SABER) (2011)
21 See Annex 4 (Test Development Cycle).
22 Ibid
4.3 System Alignment

The World Bank\textsuperscript{23} uses the term “system alignment” to refer to the situation in which different programmes or components in an education system are coherent or mutually supportive. There is evidence of alignment when the components refer to the same conceptual model or reference framework or when strengthening one component results in strengthening the other. There is misalignment when, in the case of assessment, the tasks do not cover a well-balanced and representative sample of the topics covered in the curriculum, or when the test includes topics not covered in the curriculum.

Recommended PEC/PEAS activities for (re-) alignment include the following:

- Development of an effective communications strategy to operate at central and district levels to achieve the purpose of providing information and feedback to schools and ensuring that this is received and acted upon
- Appointment of an Information Officer within PEC/PEAS to implement and monitor the communications strategy
- Releasing some used items as models of good assessment instruments
- Releasing items for use in test preparation activities
- Internally reviewing the volume and nature of assessment in Grades 4, 5 and 8
- Employ more sophisticated methods for analysing data, displaying information and reporting results.

\textsuperscript{23} Ibid
Developing optimal capacity in assessment and associated processes will require a considerable allocation of funds, a significant amount of work, long-term vision, long-term planning, and an injection of training in the immediate future. PEC and PEAS staff have different needs, and the differences are not so much related to place of employment as to specific areas of operation and interest. These different needs would be best served by taking a modular approach to training within the wider plan for enhancing capacity.

Training programmes typically break-down when technical advisors deliver the programme but do not carry out any follow-up activities or when several technical advisors deliver inconsistent messages. CE (WB) should ensure that neither of those models of technical advice prevails. Any resulting programme should be needs-based, continuous and coherent, producing internally consistent messages, and an integral part of the productive capacity of PEC/PEAS. The options proposed meet these requirements.

PEC/PEAS could encourage practising teachers and others in the community with appropriate qualifications who have an interest in standardised testing to participate in the training activities, thus contributing to the formation of a pool of possible future employees. This is one way to help with the maintenance of skill availability for the work of the PEC/PEAS.

CE (WB) might consider funding scholarships for study overseas and international attachments.

If it is agreed that the workplace competencies necessary for delivering high-quality assessment programmes in Punjab are deficient, urgent steps must be taken to remedy the situation. It is suggested that a small group from PEC/PEAS with appropriate qualifications and/or aptitudes be provided with training by a recognised provider outside the country. Recommendations for the design of a capacity-enhancing programme follow.
6. Recommendations for the design of a capacity-enhancing programme

Based on the findings of our review, PEC is in need of capacity building in both the test construction area and the psychometrics, reporting and statistical analysis area. PEAS is better placed in terms of test construction but would still need to continue staff professional development in this area to ensure continuity of good practice. Both organisations are in urgent need of training and capacity building in the psychometrics area, especially in basic principles in psychometrics, equating techniques and scale construction to move towards item banking.

Capacity building should include training of current staff and a plan to increase staffing level. The training should involve intensive workshop for current staff and potential recruits, as well as on-going mentoring to ensure quality assessment instruments and psychometric procedures for the calibration and scaling of tests for valid and useful reporting of results. Data collected should be used for research to inform policy and to improve practice. PEC and PEAS quantitative staff needs to be trained in research methods and statistical modelling to perform effective data analysis for reporting research findings.

The training and mentoring is best achieved through an institutional linkage with an organisation that has expertise in all aspects of assessment, details of which are outlined in the below sections.
7. Institutional linkage and training programme

7.1 Institutional linkage

It is recommended that an institutional linkage (sometimes called a partnership or ‘critical friendship’) with an organisation such as ACER be developed in order to strengthen technical exchange and knowledge building. In order for such a linkage to be as sustainable and useful as possible, this should be an ongoing role that could be renewed from year to year, or for a longer period, and is aimed at the provision of broad professional support, including peer review of documentation, and consultancy on assessment, measurement, and data analysis. This linkage would exist alongside the workshop program (outlined in the following sections).

Such a linkage is based on an understanding of the “necessity of ensuring a clear sense of shared ownership and the maintenance of ‘open communication’ between all partners”. Rather than acting as ‘evaluators’, ACER experts would work together with PEC and PEAS in a reciprocal relationship that promotes continuous learning. Critical friends use “two complementary approaches: (i) they generate information through discussions that offer new perspectives or ideas, and (ii) they ask challenging questions or suggest possible changes. By acknowledging one another’s unique contributions and engaging in constructive dialogue the critical friend process thus promotes the production of new knowledge and insights.”

The Memorandum of Understanding (see attachment) will be the underlying agreement for all work undertaken between ACER and PEC/PEAS and will form the basis for both the staff development workshops and also the critical friend role. In this spirit, this document will be devised and agreed on by all parties – the attached document is a draft only, awaiting review and comment from PEC/PEAS. In the MoU, PEC/PEAS and ACER agree to use their best joint endeavours to facilitate staff development, capacity building and technical assistance to support educational development in Pakistan. This will be achieved through the establishment of a strong and enduring inter-organisational relationship.

ACER has developed several valuable and ongoing partnerships with international organizations, and has signed MOUs with the Central Board of Secondary Education India, UNESCO Beirut and the Vietnam National Institute for Education Strategy and Curriculum Development. ACER has assisted with development of regional frameworks, staff development and capacity building for these and many other organizations, including the National Centre for Educational Research and Training in India, which received technical assistance from ACER of the kind proposed in this document.

Critical friend consultancy work would be undertaken by expert staff and would be performed on a needs basis. The need of PEC/PEAS for various types of consultancy depends on the expertise of staff and the number of reports and assessments for which assistance is required. As work will be requested and agreed from time to time by PEC/PEAS and ACER, the extent of support required will need to be assessed and

---

26 Please note, ACER acknowledges that it may be more appropriate for the MoU to be with another legal entity (e.g. PMIU) rather than PEC/PEAS.
costed from time to time, although it is estimated that approximately 24 days of consultancy support spread over the training period, which runs from November 2012 to August/September 2015, will be needed.

7.2 Training programme

7.2.1 Qualifications of proposed trainees – item writing and test construction

Each workshop will build on the knowledge and skills gained in the previous workshops. Therefore, it is anticipated that all PEC and PEAS item writing staff will participate in all the workshops over the two to three year period where possible. Currently PEC and PEAS engage a number of item writing staff, and all these staff should undertake the training in Lahore. These staff should have education qualifications with knowledge of school curriculum and preferably with teaching and test development experience. It will be the responsibility of PEC/PEAS to identify these potential candidates for training. It may be decided that only key item-writing and test development staff attend the final workshop in Australia, but this is at the discretion of PEC/PEAS management.

7.2.2 Qualifications of proposed trainees – Psychometrics, data analysis and reporting

Each workshop will build on the knowledge and skills gained in the previous workshops. Therefore, all workshop participants will undertake all the workshops over the three year period, preferably including the workshop in Australia.

Currently, there is one data analyst working at PEC, and two at PEAS. For longer term capacity building, at least three more trainees with mathematics and statistics qualifications should be recruited to attend the workshops. All the participants should have at least basic statistical data analysis knowledge and experience. The trainees may be identified from maths and statistics teachers or from the universities. It will be the responsibility of PEC/PEAS to identify these potential candidates.

7.3 Training modules

In-country workshops and training are able to affect a much greater number of staff and as a result the costs are substantially less. Further, as the focus is on capacity building within PEC/PEAS, it is logical for the required training to occur on location and in the relevant context. However, at the end of the 2.5 year training period, it is recommended that one workshop be undertaken at ACER’s offices in Melbourne, Australia for all PEC/PEAS staff. ACER staff in Australia will be able to show the participants our best practice and the participants can take part in our test panelling and data analysis and reporting sessions.

Further, an initial study tour will take place in Australia for decision makers and senior officers.

Modules A and B are training workshops designed specifically for test developer group and the data analyst group, respectively. Workshop participants will receive mentoring and technical support if needed during the periods between face-to-face training (Module C). For the purposes of Module C, it is proposed that a team leader be elected for each of the two training groups. The team leader will then convey any questions or issues to the ACER mentor if necessary, and all participants will be copied on the communications.

All workshops in a given Module are intended to build on the knowledge and skills developed in the previous workshops in that Module, so it is essential that all relevant staff attend all (Lahore) workshops. It
may be decided that it is beneficial for selected senior staff to participate in all workshops as well as the study tour.

**Study tour – Decision makers**

Study tour in Melbourne, Australia for policy makers/senior officers overseeing assessment programmes, with opportunities for arranged visits to assessment authorities and education departments.

Duration: 1 week consultancy

Timing: November 2012

**Module A – Item Writing and Test Construction (for test developers, item writers and new trainees)**

All training sessions will be undertaken by 2 senior assessment specialists (one Maths/Science specialist, and one Humanities specialist), ideally in conjunction with PEC/PEAS head test developers. Engaging PEC/PEAS staff in the running of the workshops builds the capacity of these staff to conduct ongoing training.

Work will be suggested for the participants after each workshop for completion before the subsequent workshop. The completed work will be discussed during training/mentoring under Module C.

**Year One**

Workshop 1: Development of Assessment Framework (Lahore) – 1 week duration (with two ACER consultants), December 2012:
- Development of Assessment framework
- Principles of test development and assessment
- Item writing – Maths, Science (with Maths/Science assessment specialist)
- Item writing – English (with Humanities assessment specialist)

**Year Two**

Workshop 2: Item Writing – Maths, Science, English (Lahore) – 1 week duration (with two ACER consultants), August/September 2013:
- Test development process and methodology – active participation of trainees
- Item writing – Maths, Science (with Maths/Science assessment specialist)
- Item writing – English (with Humanities assessment specialist)
- Refinement of items written in workshop 1
- Use of item statistics in the evaluation of item quality – discussion will make use of items and data collected in Punjab exams

Workshop 3: Item Writing Consolidation (Lahore) – 1 week duration (with two ACER consultants), November/December 2013:
- Item writing consolidation

---

27 Any reference to a ‘week’ in this document refers to five (5) consecutive working days, each consisting of 8 hours. Please note that the duration and timing of workshops will be finalised in discussion with PEC/PEAS and other relevant stakeholders.
Use of item statistics in the evaluation of item quality – discussion will make use of items and data collected in Punjab exams
Evaluate alignment with assessment framework
Development and use of learning standards for reporting

Year Three

Workshop 4: ACER Visit – Item Writing best practice (Australia) – 1 week duration, August/September 2014
Showcase test development process undertaken at ACER on various international and Australian projects
Workshops and briefing sessions with test development experts
Optional extras include trips within Australia to an examinations “board” (e.g. The New South Wales Board of Studies in Sydney)

Module B – Data Analysis (for data analysts and new trainees)

All training sessions will be undertaken by a senior psychometrician, with support from an experienced technical statistician/data analyst, who will assist PEC/PEAS staff in the ‘hands-on’ sessions.

Exercise will be assigned to the participants after each workshop for completion before the subsequent workshop. The completed work will be discussed during training.

Year Two

Workshop 1: Introduction to measurement theory and equating methods (Lahore) – 1 week duration (with two ACER consultants), August/September 2013:
Psychometrics and measurement theory
Item response theory (IRT) and applications in educational measurement
Item and test calibration and evaluation of fit based on Rasch modelling
Learning ACER Conquest software, hands-on participation
Interpretation of analysis output and statistical results
Test equating methods using Rasch modelling

Workshop 2: Test calibration and equating, formulation of research questions and statistical modelling (Lahore) – 1 week duration (with two ACER consultants), November/December 2013:
Test calibration and equating using ACER Conquest
Equating of PEC/PEAS tests using real data, hands-on participation
Formulation of research questions
Statistical modelling for answering research questions
Interpretation of analysis output and statistical results
Reporting of statistical results

Year Three

Workshop 3: Consolidation of analysis and modelling skills, reporting (Lahore) – 1 week duration (with two ACER consultants), August/September 2014
Consolidation of psychometric/analysis skills
Introduction of advanced statistical modelling
- Report writing skills

Workshop 4: ACER visit – Data analysis best practice (Australia) – 1 week duration, November/December 2014
- Showcase and demonstrate analysis undertaken at ACER on various international and Australian projects
- Workshops and briefing sessions with ACER data analysis, psychometricians, assessment experts
- Optional extras include trips within Australia to an examinations “board” (e.g. The New South Wales Board of Studies in Sydney)

Module C – Mentoring and Technical Support (for all PEC/PEAS staff)

In between workshops, there will be ongoing contact between ACER experts and PEC/PEAS staff via video-link, phone and email in order to provide support to test development and item writing, data analysis and interpretation and report writing in the following ways:
- Support to real data analysis and test development/item writing while PEC/PEAS staff are undertaking this work
- Share assessment lessons and practice
- Answer questions and provide support
- Undertake remote training sessions
- Peer reviewing documentation and reports

Two ACER trainers – one for psychometrics and another for item writing – will be assigned to PEC/PEAS so there will be specific points of contact in Australia. These mentors will communicate with the team leaders for data analysis and test development/item writing nominated by PEC/PEAS.

The extent of mentoring requested will of course vary at different times of the year, and will likely reduce from year to year as capacity is increased at PEC/PEAS, but it is estimated that 30 working days (15 for psychometrics and 15 for test development/item writing) would be needed over the training period – November 2012 to August/September 2015.

Please note: the mentoring will only start once the first workshop has been held. It will not be practicable or pedagogically sound to have remote mentoring between Australia and Pakistan before the parameters and scope of the mentoring relationship have been established at the initial workshop.
### TRAINING SCHEDULE

<table>
<thead>
<tr>
<th>Workshop topic</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>November</td>
<td>December</td>
<td>August/ September</td>
<td>November/ September</td>
</tr>
<tr>
<td>Study Tour for Policy Makers and Senior Officers*</td>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module A - Item Writing and Test Construction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workshop 1 Development of Assessment Framework</td>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workshop 2 Item Writing – Maths, Science, English</td>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workshop 3 Item Writing Consolidation</td>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workshop 4* ACER Visit - Item Writing best practice</td>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module B – Psychometrics and Data Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workshop 1 Introduction to measurement theory and equating methods</td>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workshop 2 Test calibration, formulation of research questions &amp; statistical modelling</td>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workshop 3 Consolidation of analysis and modelling skills, reporting</td>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workshop 4* ACER visit – Data analysis best practice</td>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module C – Mentoring and Technical Support</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Please note that the duration and timing of workshops will be finalised from time to time in discussion with PEC/PEAS and other relevant stakeholders.
Capacity Review of the Punjab Examination Commission (PEC) and the Punjab Education Assessment System (PEAS)

<table>
<thead>
<tr>
<th>Ongoing throughout training period (24 days in total spread according to need)</th>
<th>Support to real data analysis and test development/item writing while PEC/PEAS staff are undertaking this work</th>
<th>Share assessment lessons and practice</th>
<th>Answer questions and provide support</th>
<th>Undertake remote training sessions</th>
<th>Peer reviewing documentation and reports</th>
</tr>
</thead>
</table>

Please note: Mentoring will only commence after the first face-to-face workshop has occurred

* Asterisked workshops to take place in Melbourne, Australia, all others to be conducted in Lahore, Pakistan
Annexes

Annex 1: Memorandum of Understanding ................................................................. 33
Annex 2: Terms of Reference (TOR) ......................................................................... 44
Annex 3: PEC Work Breakdown Structure ................................................................. 48
Annex 4: PEAS Organisational Chart ....................................................................... 49
Annex 5: Test Development Cycle ............................................................................. 50
Annex 6: Extract from *The Punjab Examination Commission Act 2010* .................. 51
Annex 1: Memorandum of Understanding

MEMORANDUM OF UNDERSTANDING

[DATE]

Between

CAMBRIDGE EDUCATION
Demeter House, Station Road, Cambridge CB1 2RS, UK
(‘CE’)

AUSTRALIAN COUNCIL FOR EDUCATIONAL RESEARCH LIMITED,
[ABN 19 004 398 145], 19 Prospect Hill Road, Camberwell, Victoria, 3124, Australia
(‘ACER’)

And

Programme Monitoring and Implementation Unit/Punjab Examinations Commission/Punjab Education Assessment System (PMIU/PEC/PEAS)
PMIU, 22B, Near Ayubiya Market, New Muslim Town Lahore, Pakistan
PEC, Link Wahdat Road (near Pilot School) Lahore, Pakistan
PEAS, College Road, Township, Lahore
1. BACKGROUND/PURPOSE

The Programme Monitoring and Implementation Unit/Punjab Examinations Commission/Punjab Education Assessment System (PMIU/PEC/PEAS) and the Australian Council for Educational Research Limited (ACER)/Cambridge Education (CE) agree to collaborate in the areas of staff development, capacity building and technical assistance to support educational development in the region.

The Programme Monitoring and Implementation Unit (PMIU) was created in 2003 to monitor the implementation of the Punjab Education Sector Reform Programme, with the aim of increasing access and improving quality and governance of education in Punjab.

Under the Punjab Education Sector Reform Programme (PESRP), PEC was established in 2006 to be responsible for designing, implementing and reporting on assessments for all Grade 5 (primary school) and Grade 8 (middle school) students in Punjab. This includes students in private institutions offering Grade 5 or 8. The assessment domains for Grade 5 candidates are Maths, Science, English, Urdu, Social Studies, and Islamiat. For candidates in Grade 8 there is also Computer Sciences or Arabic, plus one other subject from a list of electives.

The Punjab Education Assessment System (PEAS) is part of the National Education Assessment System (NEAS). PEAS is responsible for monitoring the educational levels of the Punjab student population through assessment surveys.

ACER was established in 1930 with the support of the Commonwealth Government of Australia, and is an independent not-for-profit Company registered under the Corporations Act 2001. ACER's mission is to create and promote knowledge and tools that can be used to improve learning. Underlying this mission is a belief in the importance of ongoing and lifelong learning both for the fulfilment of individuals and for the well-being of society and a commitment to the use of systematic investigation, evaluation and critical reflection in the search for ways to improve learning. ACER has substantial experience in designing and conducting national assessment programs in a wide range of countries, and leads the
A consortium that holds the major contract to deliver the International PISA project on behalf of the Organization for Economic Cooperation and Development (OECD).

Cambridge Education is currently implementing the three year technical assistance package of the WB Punjab Education Sector Project (PESP) which is aimed to support the Government of Punjab as it embarks on developing and implementing second generation reforms in the education sector. The overall objective of the project is to improve access and equity, and the quality and relevance of education in Punjab.

All organizations acknowledge the benefits that international cooperation in research, assessment, capacity building and professional learning can bring to educational standards and aim to further this cooperation as outlined in this Memorandum of Understanding (MoU).

2. SCOPE OF MoU

2.1- The parties have entered into this MoU with the intention of working together in good faith and using their best endeavours for that purpose.

2.2- The parties agree that this MoU does not represent a binding commitment by either party to enter into a business relationship with the other party, and that implementation of specific activities and projects will be subject to the execution of contractual agreements which will set out in detail the terms and conditions under which these activities will be carried out, including payment of any fees.

2.3- The parties agree that this MoU is not intended to prevent either party collaborating or otherwise co-operating with any other third party.

2.4- The areas of collaboration discussed in this MoU have been identified but may be extended or modified over time, by written agreement of the parties.

2.5- No part of this MoU will limit or affect the operation of any contractual agreements signed by the parties.
3. AREAS OF COOPERATION

The parties agree to use their best joint endeavours to facilitate staff development, capacity building and technical assistance to support educational development in the Pakistan. This may be achieved through joint work in some or all of the following proposed areas for collaboration, conditional on the execution of contractual agreements outlining in detail the responsibilities of both parties for particular activities:

3.1- Developing a strong and enduring inter-organizational relationship through staff exchanges and capacity building, regular meetings and joint research and project work.

3.2- Building organisational, regional and national capacity within PMIU/PEC/PEAS, the Punjab region and Pakistan for effective implementation of assessment work in the following areas:

   3.2.1. Data analysis, including Item Response Theory methodologies and Rasch modelling techniques
   3.2.2. Development of standards and reporting scales
   3.2.3. Development of item and test construction
   3.2.4. Establishment of best practices and processes for data collection, analysis, report writing and interpretation

3.3- Staff development through activities such as:

   3.3.1. Workshops and training programs run by ACER in Pakistan and Australia
   3.3.2. Feedback and review surveys

3.4- Providing technical assistance, including in the following areas:

   3.4.1. Item writing and test construction, including development of an assessment framework;
   3.4.2. Educational Measurement, for staff development activities related to test construction including linking and equating models, scale development for each
domain and year level, and standards setting techniques to measure performance against the curriculum framework;

3.4-3. Report writing, for staff development activities related to appropriately constructed reports to cater for a variety of stakeholders, interpretation of data from PMIU/PEC/PEAS reports to inform the dissemination of appropriate information in a user-friendly manner;

3.4-4. Effective action research, for staff development activities related to identification from assessment results of topics for further contextual analysis through qualitative studies, and classroom action research to understand the implications of the data analysis at appropriate levels;

3.4-5. Other technical assistance in any area of capacity building, educational assessment or evaluation or research or other matter relating to education in Pakistan that may be identified from time to time.

3.5- Exchanging and sharing educational research, information and expertise.

3.6- Applying research evidence and developing best practices to support educational policy development.

3.7- Organizing and participating in jointly sponsored seminars and workshops.

3.8- Cooperation in such other areas that may be agreed by both parties from time to time.

4. IMPLEMENTATION OF MoU

4.1- To implement this collaboration between the parties, there will be regular correspondence between PMIU/PEC/PEAS and ACER/CE to consider, develop and approve each program of cooperative activity.

4.2- The parties’ respective general areas of expertise are outlined as follows, which will guide the implementation of any joint activities undertaken by the parties:
4.2-1. ACER has expertise in educational research, development and implementation of assessment instruments, staff professional development and capacity building including running workshops and training for implementing assessment surveys and processing and analysing assessment data.

4.2-2. CE is responsible for procuring, managing, and providing in-country technical and operational support to the project, and has expertise in organising and covering costs of technical and policy meetings, facilitating workshops and training programs.

4.2-3. PMIU/PEC/PEAS has expertise in organising and promoting the results of assessment and research, advocacy and fundraising, and regional policy implementation.

4.3- Any other clause of this MoU notwithstanding, contractual agreements will be executed to cover any specific activities to be carried out by the parties, and will set out in detail the terms and conditions under which these activities will be carried out, including payment of any fees. This MoU does not and is not intended to take the place of contractual agreements between the parties.

5. INTELLECTUAL PROPERTY RIGHTS

5.1- In this MoU “Background Intellectual Property” means pre-existing or independently developed intellectual property made available by either party for the purpose of this MoU.

5.2- The ownership of Background Intellectual Property remains with the party that introduces it to the other party for the purpose of this MoU.

5.3- ACER/CE and PMIU/PEC/PEAS may use each other’s Background Intellectual Property only with the written consent of the party which owns it.

6. CONFIDENTIAL INFORMATION
6.1- In this MoU “Confidential Information” of a party means information that is confidential or which is treated by that party as confidential and which relates either directly or indirectly to the business of that party regardless of the form in which that information is constituted, and which is not lawfully in the public domain.

6.2- Each party must keep all Confidential Information of the other party confidential, which becomes known to it during the operation of this MoU.

6.3- Each party must not disclose Confidential Information of the other party to any person without the prior written consent of the other party.

6.4- On termination or expiry of this MoU, each party must immediately return to the other party or delete or destroy all Confidential Information of the other party in its possession or control.

6.5- The obligations of the parties under this clause survive termination or expiry of this MoU for any reason.

7. COSTS AND EXPENSES
Cambridge Education will be responsible for the costs and expenses of implementing the MOU determined under executed contractual agreements between the CE and ACER.

8. LIABILITY
To the extent permitted by law, neither Party has any liability for any indirect or consequential expenses, losses, damages or costs incurred by the other Party in relation to any matter, act, omission or thing arising out of, connected with, or in relation to, this Agreement.

9. ANNOUNCEMENTS
9.1- Any major public announcement in connection with this agreement must be agreed by the parties before it is made, except if required by law or a regulatory body, in which case the party required to make an announcement must, to the extent practicable, first consult with and take into account the reasonable requirements of each other party. For clarity, a major public announcement means an announcement in the press or on a website that announces the parties’ general collaboration under this agreement or announces the parties’ collaboration on particular new programs and initiatives.

9.2- Minor announcements relating to activities undertaken under this MoU that have been validated in writing by both parties as accurate, and not involving information about a new initiative or substantial change in the nature of the parties’ relationship, and not containing confidential information of any party or information belonging to a third party, may be made routinely available on the parties’ websites or in other documentation produced by either party.

9.3- This clause does not limit or affect the operation of any contractual agreements signed by the parties.

10. RELATIONSHIP OF THE PARTIES
Neither party shall represent itself, and shall ensure that its employees or subcontractors do not represent themselves, as being an employee, partner or agent of the other party, or as otherwise able to bind or represent the other party.

11. TERM AND TERMINATION
11.1- This MoU will commence upon the date of execution and continue until terminated under the provisions stated in this clause.
11.2- Either party in their sole discretion terminate this MoU upon 60 days written notice to the other party.

11.3- The termination of the MoU will not affect any other signed written agreement between the parties.
This MEMORANDUM OF UNDERSTANDING is made BETWEEN the Australian Council for Educational Research Limited/Cambridge Education Limited and PMIU/PEC/PEAS,

Punjab Examinations Commission/Punjab Education Assessment System (PMIU/PEC/PEAS)

Signed by

sign here ►

Director

print name

AND

Australian Council for Educational Research Limited (ACER)

Signed by

sign here ►

Director International Development

print name
Capacity Review of the Punjab Examination Commission (PEC) and the Punjab Education Assessment System (PEAS)

Cambridge Education Limited (CE)

Signed by

______________________________
Director

______________________________
print name
Annex 2: Terms of Reference (TOR)

Australian Council for Educational Research

Capacity review of PEC and PEAS

12.5.12

Introduction
This TOR is for experts from an education institution to assist Cambridge Education (CE) by conducting a capacity review of the Punjab Examination Commission (PEC) and the Punjab Education Assessment System (PEAS). The experts will provide CE with recommendations for the design of a program to strengthen the capacity of PEC and PEAS to deliver the assessment programs and associated processes for which they are responsible.

Background
Cambridge Education (CE) is currently implementing the three-year technical assistance package of the World Bank’s Punjab Education Sector Project (PESP). This is under the Punjab Education Sector Reform Programme I (PESRP I), which is aimed at supporting the Government of Punjab as it embarks on developing and implementing second-generation reforms in the education sector.

CE is working with government institutions – the Programme Monitoring Implementation Unit (PMIU)/School Education Department (SED), Directorate of Staff Development (DSD), Punjab Examination Commission (PEC), Punjab Education Assessment System (PEAS) – and with donors – German Agency for International Cooperation (GIZ), United Nations Children’s Fund (UNICEF), World Bank (WB), and Department for International Development (DfID). CE is providing a diverse range of technical services to i) enhance fiscal sustainability and the effectiveness of public expenditures including education; ii) enhance the quality of school education; iii) improve and expand access through improvements in school participation and completion rates and reduction in gender and regional disparities; and iv) strengthen school management and governance within the sector.

CE intends to present a detailed case to donors, seeking support for capacity-building activities for, and institutional strengthening of, PEC and PEAS.

As a preliminary step, CE requires a “thorough root-and-branch review of PEC and PEAS in order to make an accurate assessment of the capacities to deliver quality assessment programs and associated reports to accurately assess their real technical needs” – an exhaustive review of PEC and PEAS in order to make an accurate assessment of their capacities to deliver quality assessment programs and associated reports. The review will identify the support that PEC and PEAS need in both the short and long term to ensure that they effectively and efficiently fulfil their roles and responsibilities. Specifically, an analysis of what capacity currently exists will be conducted in order to identify where strengthening is required.
Scope of Work
To comply with CE’s specific requirements, ACER will undertake a thorough, three-stage review. ACER will provide recommendations for the design of a program to strengthen the capacity of PEC and PEAS to deliver the assessment programs and associated processes (including reports) for which they are responsible. Advice will be given on policy and professional development that will protect, retain and grow capacity in system assessments.

The scope of work, lasting 36 days, will include an examination of the following areas of the operation:

1. Curriculum frameworks (assessment objectives) and test item accuracy, validity and reliability
2. Procedures and practices for pre-testing item accuracy and reliability and developing item banks
3. Ensuring exam/test security (during, before, and after the exams/tests)
4. Scoring and coding of tests, analysing and communicating results to various audiences
5. Communication of information at various stages to (potential) test-takers, looking at relevant, completeness, and timeliness.
6. Helping schools and students familiarize themselves with test contents and test taking requirements and procedures in advance of test administration
7. Preparation and training of district education administrations for administering the tests.
8. Test administration procedures and practices by district education departments (all stages).

ACER will review both PEC and PEAS in light of a possible merger between the two organisations and the report will consider the possible issues attendant on such a merger.

Tasks
1. Desk Review

The first stage will be a desk review of available documentation to establish a detailed, formal understanding of PEC and PEAS – their structures, mandate, ethos and staff capacity. This requires provision of key documentation in soft copy to ACER by both institutions.

These documents will include:

- Strategic plans
- Organisational charts, showing lines of responsibility and staff numbers
- Budget documents
• Staff CVs

• Significant examples of deliverables (for example, tests, data sets and analyses, public and technical reports).

The initial phase of the review may also include discussion by telephone with key personnel from CE and other stakeholders. The review of formal documents and initial telephone discussions will be conducted jointly by senior research staff in ACER’s Assessment & Psychometrics program.

2. In-Country Visit

The second stage will consist of intensive participatory meetings in Lahore with the heads/directors and technical staff of PEC and PEAS. These discussions are essential for assessing the following:

• How formal policies and processes are enacted;

• Institutional and individual capacity to fulfil their mission; and,

• Constraints that reduce the capacity of the institutions to deliver their mission.

The second stage of the review will be jointly conducted by senior staff in ACER’s Assessment & Psychometrics program.

Elements of the previous discussion will be allocated to three main analytical areas: Enabling environment for assessment in Punjab; Technical quality of existing assessment practice; and Alignment of assessment with other aspects of the system such as curriculum, text books, and teacher training.

3. Report Writing

The third stage of the review will involve the preparation of a report of no more than 30 pages (excluding annexes), which covers the findings from the needs assessment and proposes a series of capacity building strategies and activities, in the form of programme options.

The report will offer three options. Each activity proposed within each option will be ranked according to priority: i) essential; ii) desirable; and iii) optional. Each option and each activity within each option will be individually costed. The budget will further be broken down into fees and reimbursables and will include other costs associated with the provision of service by ACER.

An implementation plan will be prepared for each option. The plan will clearly indicate timing and sequencing of the activities within that option. The report will be prepared by two senior staff in ACER’s Assessment & Psychometrics program together with ACER’s Senior Accountant.
**Days Required**
A total of 40 working days including:

- 36 expert fee days (18 days for each expert) comprising:
  - 4 travel days (2 for each expert),
  - 5 days in Australia for desk review,
  - 14 days spent in Pakistan (6 for one expert, 8 for second expert),
  - 5 days for analysis and recommendations, and
  - 8 days for costing of options and writing the report.

- 4 days for project management and coordination, including drafting and negotiation of TOR, flight and accommodation bookings, organisation of visas, and project meetings and discussions.

The duration of the assignment is 4 May to 4 July 2012.

All travel and accommodation expenses, including business class flights and per diems will be in addition to these days: To be reimbursed by Cambridge Education on provision of receipts, invoices or credit card statements by ACER staff. Where such documentation is not available (e.g. for local travel), ACER staff will provide a declaration and breakdown of costs endorsed by ACER’s chief financial officer for reimbursement purposes.

**Outputs of Consultant**
Report of no more than 30 pages (excluding annexes) identifying areas for improvement in the capacity of PEC and PEAS, and containing recommendations and costed options that address those gaps.
Annex 3: PEC Work Breakdown Structure

Overall conduct of examination 1

Research 2

Database management 3

Capacity building 4

Financial management 5

Item development 1.1

Registration of students 1.2

Selection of centres 1.3

Supervisory staff for examination 1.4

Printing /delivery of material 1.5

Result formation 1.6

Research on item development 2.1

Research after result on available data 2.2

Development of software /web site and maintenance 3.1

Training supervisory staff 4.1

Training of item developer 4.2

Training of marking staff 4.3

Budget preparation 5.3

Payment to examination staff 5.2

Payment for printing 5.1
Annex 4: PEAS Organisational Chart

Director

Admin Section

Deputy Director

Steno Typist
Key Batch Operator
Clerk
Driver
Dispatch Rider
Naib Qasid
Sanitary Worker
Gardener

Academic Section

Senior Subject Specialist

Assessment Expert

Account Section

Accounts Officer

Accountant

Technical Section

System Analyst

Computer Programmerr

PEAS Strength
1. Director BS-19 [No 1]
2. Deputy Director BS-18 [No 1]
3. Assessment Expert BS-18 [No 1]
4. Senior Subject Specialist BS-18 [No 5]
5. System Analyst BS-18 [No 1]
6. Accounts Officer BS-18 [No 1]
7. Computer Programmer BS-17 [No 1]
8. Accountant BS-14 [No 1]
9. Steno Typist BS-12 [No 1]
10. Key Batch Operator BS-10 [No 1]
11. Clerk BS-05 [No 1]
12. Dispatch Rider BS-4 [No 1]
13. Driver BS [No 1]
14. Naib Qasid BS-1 [No 2]
15. Gardener BS-1 [No 1]
16. Sanitary Worker BS-1 [No 1]
Annex 5: Test Development Cycle

Create

Refine

Bank

Select

Assemble

Reflect

Administer

Produce

Mark

Combine

Score

Grade

Certify

Scale
Annex 6: Extract from *The Punjab Examination Commission Act 2010*

“An Act to establish a Commission for improving the examination system and for conducting examinations of elementary education and to provide for ancillary matters”

**Functions of the Commission**

a. Design, develop, implement, maintain, monitor and evaluate a system of examination for elementary education

b. Formulate policy and programme for the conduct of examinations relating to elementary education

c. Collect data for research to improve curricula and teaching methodology

d. Recommend measures for capacity building of the teachers and educationists for improving the system of assessment of students

e. Identify the areas where improvement in training of the teachers or educationists is required

f. Promote public discussions on issues pertaining to elementary education

g. Advise the Government on all policy matters relating to the objectives of the commission

h. Approve annual research programme and annual budget of the Commission; and

i. Perform such other functions as may be ancillary to its functions, or as may be prescribed or as may be assigned by the Government.

The Commission may

a. Issue instructions or guidelines to a local government for data collection and conduct of an examination of elementary education;

b. Register students for the purpose of examination of elementary education and maintain their data; and,

c. Conduct and outsource research in the sphere of an examination of elementary education