Journal of Education & Social Sciences

ISSN: 2410-5767 (Online)
ISSN: 2414-8091 (Print)

Review of the Role of Large Scale Standardised Assessment Practices in Sindh Pakistan

Affiliation:
Razia Fakir Mohammad
Aga Khan University Institute for Educational Development.
Email: razia.fakir.m@gmail.com
Anil Khamis
The Aga Khan University and University College London Institute of Education

Manuscript Information
Submission Date: November 28, 2020
Reviews Completed: February 23, 2021
Acceptance Date: February 27, 2021
Publication Date: March 9, 2021

Citation in APA Style:
Mohammad, R. F., & Khamis, A. (2021). Review of the Role of Large Scale Standardised Assessment Practices in Sindh Pakistan, *Journal of Education & Social Sciences, 9*(1), 28-41.

DOI: https://doi.org/10.20547/jess0912109103

.

Journal of Education & Social Sciences
Vol. 9(1): 28-41, 2021
DOI: 10.20547/jess0912109103

Review of the Role of Large Scale Standardised Assessment Practices in Sindh Pakistan

Razia Fakir Mohammad [*] Anil Khamis [†]

Abstract: *This paper reviews two large-scale assessments (LSAs) with reference to their purported contribution to improvement of education. Documentary analysis is used to review and analyse publicly accessible documents, which describe and/or explain LSA practices in Sindh. Additionally, episodic data were collected through routine fieldwork with head teachers of primary and secondary schools who had participated in policy dialogue forums where LSAs were discussed. This provides rich contextual data with ground level perspectives. A fundamental question in the approach of LSAs that this paper considers is: does measurement of student performance, with no other supporting or triangulating data sources, allow us to judge the quality of education at a provincial or national level? In the context of Sindh, one of four provinces in Pakistan, with a population of well over 60 million, how valid are the findings of LSAs and what should be done to further strengthen the effort that now has near global uptake? Corollary to the above question is what has been the role of LSAs in promoting quality education in the classrooms. The paper has reviewed gaps in the practice and policy implications of current LSAs and argues for greater validation and critical use of LSAs to judge the current system of education. The paper also calls for further research, particularly systematic reviews and meta-analyses of LSAs.*

Keywords: Educational assessment reform; large scale assessment; document analysis.

Introduction

Over the past decade, governmental and non-governmental organisations have accelerated data collection efforts focused on student performance on achievement tests that assess content knowledge using a variety of tools and instruments. The purpose of such large-scale assessments (LSAs) and the accompanying widespread dissemination is to reform educational policies and practices with the overall aim of increasing the quality of education. In this paper we review two LSAs with reference to their purported contributions to the improvement of education. The primary focus of this paper is Sindh Pakistan, where LSAs have been prominent for the last decade. Whilst the paper presents a critical review, the intent is to consider practical ways to further LSA data use beyond abstract argumentation that is so prevalent in the literature. Documentary analysis is used to review and analyse publicly accessible documents, which describe and/or explain LSA

[*]Aga Khan University Institute for Educational Development.
[†]The Aga Khan University and University College London Institute of Education.

Received November 28, 2020; Received in revised form February 23, 2021; Accepted February 27, 2021; Available online March 9, 2021

practices in Sindh. Additionally, episodic data collected through routine fieldwork with head teachers of primary and secondary schools who had participated in policy dialogue forums where LSAs discussion inform the analysis presented in the paper. This provides a combination of rich contextual data with ground level perspectives. A fundamental question in the approach of LSAs that this paper considers is: does the measurement of student performance, with no other supporting or triangulating data sources, allow us to judge the quality of education at a provincial or national level? In the context of Sindh, one of the four provinces in Pakistan, with a population of well over 60 million, the question that should be asked is: how valid are the findings of LSAs and what should be done to further strengthen the quality of the assessments as well as the effort required to conduct such assessments? Corollary to the above question is what has been the role of LSAs in promoting quality education in Sindh. We review how prominent LSAs, determines their influence, and examines their provenance in the context of Pakistan.

Theoretical Framework

The term large-scale assessment (LSA) has its origin in the notion of standardised assessment and refers to initiatives by public or private organisations, in which a large number of students based on (representative) sampling of a specific age and level are assessed through tests or other instruments. The LSA results are then used to draw inferences about the quality of learning, programme, and resource allocation. Implicit in the use of LSAs are the assumptions that a common assessment method of students, across schools, is an appropriate means to measure the progression of all students, and make informed decisions with regards to quality enhancement and reforms (Phelps et al., 2008). With the growing trend in the use of assessment for improvement of educational outcomes, accountability has been added as an overarching consideration. This resulted in a discourse established by the early 20th century that coupled student assessment to judgements of schools' efficacy (Nagy, 2000). Popham (2001) directs 'the annual collection of students' achievement scores will allow the public and educational policymakers to see if educators are performing satisfactorily' (p. 34). Lewis (2000) discussed LSA as a "feedback loop" that inputs updated, reliable, and accurate information into the policy process so that programmes and initiatives are monitored and improved. Almond (2002) and other like-minded commentators consider LSAs' utility related to programme and system-level decision-making for further improvement, effectiveness and accountability to relevant stakeholders. Others note that curriculum designers and policy makers can utilise large-scale assessment data to review quality of curriculum. The presence of results can also invite and generate discourses on reform in education and teaching and learning practices. Today, such assessment programmes are used to rationalise and effect changes in policy (Clarke, 2012), curriculum, and educational practice especially in a milieu of data-driven decision making (Klinger, Rogers, Anderson, Poth, & Calman, 2006). LSA results include information on school infrastructure, pedagogical approaches and processes, and longitudinal studies. From the theoretical perspectives, LSAs have come to be considered as largely beneficial and unproblematic – a means of gaining a direct link to the process

of schooling via tests, that measures both student and school performance, provides information regarding curricular and policy guidelines, and monitors effectiveness and a roadmap to reform measures.

Contrary to the above, an analysis of the literature indicates that large scale tests provide a minimalist perspective of system results - such test scores increase the risk of education failure, hold teachers responsible for results, despite having different levels of support and resources and narrow and distort the curriculum (Kohn, 2000). LSA scores are not relevant indicators of student achievement as "most published tests favor economically and socially advantaged children over their counterparts from lower socio-economic families (p. 67)." The scores make stakeholders assume that test scores are direct evidence of student learning and they learn to ignore the potential impact of other external variables on student achievements (Koretz & Hamilton, 2000).

As a result of LSA results, schools and systems do not appear to make explicit attempts to restructure or re-organize the schools, instead time, money and effort is redirected to some parts of the system to speedily increase their students' test scores and showcase schools' effectiveness.

Emergence of LSAs in Pakistan: Historical Background

Based on the unfolding discourse and assumptions on the effective value of change purported to be taking place at school -level, in teaching-learning, and in policy and governance modalities, the culture of large -scale assessment has grown exponentially in Pakistan. Governmental and non-governmental organisations have redoubled their efforts to collect data on student performance through various instruments including achievement tests, school-based surveys and household-based surveys. LSAs are de facto viewed as a critical parameter to judge the education system, to monitor its quality, and to make decisions on strategies and processes to improve the quality of teaching and learning.

Historically the practices of external high stake examination were introduced in the period of colonialism under British rule, which continued after the independence of the country. In 1947 there was no independent examination board in Pakistan. The first Board of Secondary and Intermediate Examination (BISE) was established in 1951 in Karachi, and later more Boards were established. The Sharif Commission (1959) introduced internal school based examinations in the early grades to improve students' assessment results in the external examinations. However, it was reported that teachers limited skills, poor infra structures, and limited capacity in assessment and examinations has not changed the situation over the years. According to NEP, 2017 report, this situation invited the attention of policymakers resulting in the establishment of the National Education Assessment System (NEAS) at the federal level and Provincial/Area Education Assessment Centres (PEACs/AEACs).

The National Education Assessment System (NEAS) was established in 2003 with bilateral donor funding, to introduce a coherent mechanism of monitoring schools' performance. The main purpose was to track overall system efficiency as well as individual student performance through large scale assessment, and identify key areas for intervention that would lead to improvement in the quality and effectiveness of the education system.

NEAS was launched in response to the country's rising need for reliable data to improve the education system. This was emphasised in stakeholder consultations conducted by UNICEF in 1999, which presented a concept paper (Education Monitoring 2016). With the support of the Pakistan government and further funding from the UK Department of International Development (DFID), and a grant from the World Bank, the NEAS assessment project was extended by its counterparts in the provinces with the development of Provincial Educational Assessment Centres (PEACe). PEACs were initially supposed to conduct low-stakes, independent sample-based assessments, with NEAS providing the conceptual and methodological approaches as well as capacity-building, tools of development, and ensuring compliance with international assessment practices. Thereby, the aim was to inform policy specifically with regard to the variation in student performance; monitor standards of education in terms of curriculum knowledge and skills by way of assessment tools, identify the correlates of the (principal) determinants of student performance, and assist teachers to use the emerging data and information to raise student performance in subsequent LSAs.

The new Constitution's (2010) 18th Amendment devolved responsibility for education to the provinces from the earlier national oversight. The Standardised Achievement Test (SAT) was created in 2011 by the Resource Support Unit and the Provincial Education Department and was used to evaluate Grade 5 (primary, before the end of elementary schooling) and Grade 8 (secondary, two years prior to matriculation) students' performance in public and private schools. The RSUs were established as World Bank funded units in Sindh, the Sindh Education Sector Project (SEP) being established for this purpose.

In 2008, another LSA organisation, the Annual Status of Education Report (ASER), sponsored amongst others by USAID and Open Societies Initiative, started the largest 'citizen led' non-governmental household survey. The survey claimed to provide reliable estimates on the schooling status of children aged 3-16 years across Pakistan in urban as well as rural and urban areas. ASER piloted its first survey in 2008/2009 and since then has repeated the annual survey nationally with a focus on children literacy/reading and numeracy measurement.

Concurrently, an American-based organisation, Research Triangle International (RTI), funded almost entirely by USAID, established and promoted its Early Grade Reading Assessment (EGRA) and Early Reading Mathematics Assessment (EGMA) initiative. RTI in partnership with the Global Partnership in Education (GPE), post 2015, with the declaration of Sustainable Development Goals, is the preferred partner of the Sindh government for carrying out assessments of primary students' reading ability in the local language: Sindhi, and the national language, Urdu. EGRA is an oral assessment administered to individual children in their homes. RTI's baseline studies (2013) form the base-line empirical data that informed the planning of the Sindh Primary Reading Project (PRP).

Hence over the course of the last 15 years, as this brief historical review has striven to illustrate, LSAs have increased in number and scope to aggregate more and more the voice of the state of education virtually unchallenged: In the light of the history and uptake of LSAs in Pakistan, in particular in Sindh, this paper now presents a review of standardised assessment practices.

Data Collection and Analysis Methods

The authors reviewed two examples of LSA which are the most prominent and influential in sustaining the current discourse of education and schooling: SAT (Standardised Achievement Test), and household survey ASER (Annual Status of Education Report). Document analysis was used to review and analyse publicly accessible documents, which describe and/or explain LSA practices in Sindh. Additionally, episodic data collected through routine fieldwork with head teachers of primary and secondary schools who had participated in policy dialogue forums where LSAs was discussed.

The analysis presented begins with a thorough descriptive analysis identifying self-defined aims, features, purpose, methodological approaches and the subsequent planning, design, conduct, and reporting of the findings which are summarised in Table 1 attached.

Following the descriptive analysis, a critical discourse analysis is presented using a comparative approach. The analysis focuses on the use and design of the LSAs, within, and across, three areas identified. Taking a Foucauldian approach, the issue of interrogation is the narrative presented across LSAs with regard to areas that are reinforced by one-another in addition to the rigor of the empirical findings. Thus, questions such as, whether the results presented relate to the LSA purpose and conceptualisation and whether the findings reinforce each other are implicit. Thus, it is argued here, that the analysis supports the identification of the (methodological) strengths and weakness of each of the individual assessments, and supports a more critical and nuanced understanding of the uses of LSAs in the context of Pakistan.

Descriptive Analysis: An overview of SAT and ASER

A comparative analysis of the LSAs revealed that they are low-stakes and census based, administered annually. There are some variations in terms of stated purposes and structure or design approach; however, both claim methodological veracity with the use of appropriate methods to construct well-designed test items, maintain test security, transparency and fairness of administration to reduce bias or confounding variables.

Standardised Achievement Test (SAT)

SAT targets all students in Grades V and VIII (n 300,000) in public schools and selected private schools across Sindh (SAT Report, 2016). The SAT tests students in Mathematics, Science and languages - items are developed following the framework of the national curriculum and textbooks in use. The language component tests students in three languages: Sindhi - the provincial language, Urdu – the national language, and English – the official language of the Republic of Pakistan medium schools. Depending on locality, not all three language tests are administered to all students. Background information on factors supporting and/or hindering student achievement such as student socio-economic status, their parental education, teacher qualification is gathered via surveys in the districts.

The item writers - subject specialists- are hired and oriented to write test items, which are then reviewed by IBA. Test items are developed at knowledge, understanding, and application level as per the weightages of 40%, 35% and 25%. For each topic/concept, four versions of items, in English, are developed, reviewed and refined. A team is recruited for the implementation of tests. The test is administered, under supervision of SAT staff, sequentially by coordination of education officers at sub-district levels, called Talaqua Officers (TOs). The TOs coordinate with people on the ground and finalize procedural matters such as field plans, invigilation, security, and logistics. In order to achieve smooth administration, they are provided with a 2-day orientation training by SAT staff. Tests are e-marked by providing the markers with E-markers 'scoring scheme which contains an analytical rubric as well as an example of an answer for every CRQ that appears on screen' (Education Monitor III , 2016). The reports mainly include detail descriptive analysis of scores - school wise, taluka wise, district wise, region wise, and subject wise. Moreover, comparative analysis of results that is, comparison of schools within same district and/or sub-district is also provided; which however to not clearly explain the tools that are used to analyse the results. The reports indicate that teachers and head teachers are informed about test venue and schedule a week before the test administration (SAT Report 2015/16 p: 20). An overview of overall results, province wise, indicates almost no variation among the subject results over the last three years – the students' performance in mathematics, science and language has been remained very low over the last five years (SAT 2015 /16; P: 81/32/33) – see Table 2 attached.

Annual Status of Education Report (ASER)

ASER Pakistan is managed collaboratively by Idara-e-Taleem-o-Aagahi (ITA) and other international partners. They design and administer nation-wide house hold surveys of children aged 5-16 years. ASER surveys aim to measure basic literacy and numeracy skills of students of primary grades from private and public sectors, annually.

ASER surveys are administered as face to face individual interviews and individual pencil and paper tests. Unlike the previously discussed assessments, ASER tests are conducted in the household environment during the weekends. Data is mainly collected by volunteers who are fresh graduates/students of universities. They are provided with three days training to administer the tests and surveys and gather the required data. Background information is also collected through questionnaire administration. In addition, schools are visited to collect school level data (Annual Statutes of Education Report, ASER Pakistan , 2017). ASER results report various information such as student learning gaps, gender wise and province wise, in language and numeracy, enrollment rate in public and private schools, gender equity and disability (in hearing, seeing, walking, memorising) (Annual Statutes of Education Report, ASER Pakistan , 2017). ASER also indicates, as mentioned in Table 3, students' low performance in reading and numeracy repeatedly for the last several years.

Critical Discourse Analysis: Uses and Usefulness

Becoming Data Rich

Our document review indicated that SAT and ASER reports have been presenting consistently low results and repeated information about students' low achievements on an annual basis since they have been delivered. Their data has been presented in annual reports and high profile conferences/seminars. This data has generated a discourse around learning gaps, potential limitations in the development of education with suggested approaches to ameliorate the situation, and increase social mobility to achieve quality learning in the local context (ASER, 2015). Donors, such as the European Union and World Bank also support public sector departments to initiate projects including the review textbooks and activation of the Professional Institute for Teacher Education (PITE's) role in monitoring and enhancing teaching practices. The increased media attention given to LSAs also indicates the importance of assessments and their purported findings and the narratives being created have become de facto accepted (Open Society Foundation, 2012).

However, it is important to notice that these finding are not unique. The National Achievement Test (NAT, 2014) also supported the perceived decline in achievements of students over the years. For example, '79% students in Science obtained scaled mean score below the mean of 500, and only 21% of students could cross the mean score of 500. Students' performance in Science declined during 2006 to 2014. Similarly, the results of achievement in English (writing) also present alarming figures that emphasize the need of urgent remedial measures' (National Education Policy , 2017).

Several other assessment studies have been conducted with support of international donors and agencies to improve the education system by an affective examination and assessment system: the World Bank's Primary Education Project in 1984, the Basic Research and Implementation in Developing Education Systems (BRIDGES) Project of the Harvard Institute of International Development in 1988-89, UNESCO, in collaboration with Curriculum Wing of Ministry of Education, Islamabad. Yet findings of these assessment studies indicated low learning achievement levels of students (National Education Policy , 2017), as well as malpractice in examination systems and test design (Education Monitor III , 2016). But there is no reform in schooling since then and so far.

Pakistan, at one time, had a dearth of evidence to qualify or disqualify education policies, to decide upon intervention and allocation of resources with reference to educational reform. The LSAs, as the data sources, can certainly be in used for policy makers to generate discourse around educational reform initiative, but in a country, with 58% literacy rate and 51 million out of school children of age (5-16) along with low learning progress and many other issues related to human and material resources may not need repeated information and statistics about student achievement in the LSAs on a regular basis. The analysis also indicates that in the main the students, teachers and parents do not receive the results; and, if received, schools don't receive the results until two or three months later, by which time the students have usually moved on to other grades. Due to frequent testing of students' achievement, we are becoming data rich, which has been used for organising seminars and presentations, but not necessarily in a helpful or informed manner examining where the students are, what are the missing areas, and what is required to

achieve improvement.

Becoming Instrumentals

The landscape of Sindh is diverse; living conditions, norms and contexts are complex and influenced by many environmental, administrative and social factors. Coverage of the content and/or competencies, teaching styles also may vary from school to school within and across a district. Students' anxiety of assessment of learning along with outsiders'/externals' involvement, in schools and/or households' tests, could add to their cognitive, social and emotional stress. A deficit in the LSAs findings is that they do not highlight any other related issues and their possible impact on students' performance. A reflective approach to question appears missing: how the validity of these assessment results were established and the item profiles prepared and how well these assessments function during the real setting or whether an item tests what it is expected to assess.

Our conversation with a group of head teachers indicated that there was a lack of coordination and communication between the LSA organisations and the schools. Schools were not informed about the LSA schedule in advance, therefore the LSA schedule overlapped with the schools' final examination plans. For example, in many schools in a district the students came to take their final examination, but they were asked to appear in the LSA assessment. In some cases, students came to take the final examinations at their schools but were moved to SAT examinations centers for SAT assessments. Also it was found that the students were not familiar with the examination/test format adapted by the large scale assessment organization.

The LSA reports do not discuss if external and/or internal factors maximise the measurement errors, and whether the student scores represent true achievement of the learning outcomes; if they do, they do not say to what extent. No information is given regarding the methods, logistics and psychological factors that might impede students' responses on these LSAs nor is there any information on the conditions and circumstances that influence students' performances in any kind of administered test. Questions of diversity of the geographical landscape, population, and culture are not addressed and neither is the effect of this diversity on the reliability of the findings presented. Reports remain silent on the potential issues impeding students' performance in their assessment settings. Assessment data in the absence of detailed explanations may raise issues of accuracy and authenticity of the data.

Singular Judgments

LSA organisations hold teachers accountable for student performance, without providing supporting evidence. Other factors (such as parental interest, students' interest in the subjects, students' backgrounds, nutrition, school environment, textbooks, quality and relevance of learning standards, psychological and emotional factors of experiencing overwhelming test culture) that may affect students' performance are ignored in the analysis of their findings. The LSA reports repeatedly inform a continuous decline in student learning resulting from poor quality of teaching (ASER, 2014, 2015, 2016, 2017, 2018,

2019, 2020; SAT-I, 2013; SAT-II, 2014; SAT-III, 2015; SAT-IV, 2016; SAT-V, 2017; SAT-VI, 2018). For example, in a policy dialogue seminar SAT suggested, 'Removal of teachers who perform below 60% in their subject test via golden hand shake ....' (SAT report on Recommendations for Policy and School Improvement Reforms, July 2017). However, would it be better to provide better training for these teachers rather than remove them? If they do not teach and are removed, who is going to take their place? The analysis remains very generic and vague, and mainly emphasises on a singular perspective. Moreover, suggested claims such as, '...Content strand based analysis of districts' class VIII scores revealed a pattern of achievement similar to class V.

There is no evidence if the teachers and/or schools were able to understand the findings presented in the LSAs' reports. Any single agenda of reporting about students' scores and offering relevant opinions about teaching may not bring improvement and development in the absence of a systematic and holistic approach to gathering and analysing data with a holistic approach to reform in teaching and learning.

Moreover, the LSAs narrative seems to critique governmental school systems as part of a larger assessment of government ineffectiveness in (education) service provision and thus call for change, reform, and support by non-governmental agencies and institutions, both local and external, since mainly the findings, in one way or the other, appear to claim that the public sector struggles to provide quality learning opportunities. For example, ASER (2016) indicates the better performance of private schools as compared to government schools over the years (P:6). ASER (2015) indicates that '94% of third graders are unable to read two line sentences as compared to 91% in the previous year, in public sector. 92% children enrolled in Class 3 could not solve two-digit division in 2015 similar to 92% of children in 2014'. SAT results also indicate consistent low learning performance measured in the last four sets of assessment. Similarly, based on ASER 2014 survey, UN reports that Pakistan's education is 50 years behind the world [The Express Tribune, September 7th, 2016]. Many 10-year-olds have never been to school, have already left (often because of not benefiting from the experience) or were in a lower grade and had not yet developed reading skills. While only 89% of grade 10 students could read very simple text, only 64 per cent of sampled 14-year-old could do so, a difference of 25 percentage points.'

The results have put students of mainstream schools at a low standard as compared with other countries across the globe –are such inferences beneficial for our students and for their future ambitions, self-esteem and motivation? The current findings promote a culture of distrust towards the wider national education system, and in turn produce defensiveness rather than the desire to improve and motivate the education system and student achievement. Implicitly, their message is that the public sector struggles to provide good quality learning opportunities. Despite their production of huge data, the major concerns regarding improvement in teaching and learning remain unanswered so far at national and/or provisional level. Our analysis is that if we continue to keep routine of traditional testing and reporting declines in students' results, we will be lost generation in the 21st century. Development and implementation of meaningful assessment practices, that can directly support individual student learning outcomes, is a necessity for the long-term benefits. Sindh Pakistan, in 2020 is going to get into another LSA (TIMSS),

it is important to ask the 'why' and 'how' to assesses and ensure that the results of the LSA are actually used to improve teaching and learning practices rather than promoting a culture of deficit analysis.

Discussion

Large scale assessment examination is an educational tool whose purpose is to monitor and improve quality education and thus, has a potential to be a key element in transforming schools into places of quality learning for all students. We argue that LSAs in the context of Pakistan must be used as a highly reflective and educational process to support and improve learning practices rather than to continually evaluate the learning performance and report the similar negative findings over time. Additionally, since several reports by the government have also reported limited capacity and quality in the area of assessment and examination, the LSAs, in the context of Pakistan, appear to impose another weak layer of assessing students' performance, generating replicate findings, and reinforcing the culture of teach for test. Hence, if the learning does not occur as per standard expectations then the idea of testing and measurement of learning makes no sense (Gorard, 2010).

The demands of the today's world require students learn many skills. A knowledge - based, highly technological economy requires that students master higher-order thinking skills and that they are able to see the relationships among seemingly diverse concepts. Skills such as teamwork, collaboration, communication and moral character are critical. There is evidence from observations that assessment in the current scenarios have added to students' experiencing depression, anxiety, low self-esteem and poverty. The key point highlighted in this paper is that the greatest responsibility of assessment is not judging success and failures but help students realize their human potential and grow emotionally as well as intellectually. Collaborative assessment quality questioning, skilled feedback and self/peer assessment strategies – in place of ready-made solutions and/or factual recall test make students think, ensure learning with understanding, establish safe classrooms with few learning risks, and prepare students for future challenges (Wiliam, 2011; Hattie, 2012). Hence, the real essence of assessment is formative.

Theoretically and ethically, assessment plays an important role in improving learning and teaching practices; therefore, it is important to question whether the LSAs, the way they are currently designed and executed, serves this purpose or not. There is, thus, an urgent need for policy makers and educators to engage in reflective and critical dialogue on uses and usefulness of the LSAs.

These questions will guide the agenda to develop relevant policies and take stronger actions that are needed to improve classroom based, school based and large scale assessment practices. Moreover, no matter how well the tests are conducted, the focus of the large scale assessments (in the form of paper-pencil test), by its very nature, remains limited. Perhaps, in the long term, summative assessment should also integrate authentic assessment tasks. This will not only help in assessing high level learning outcomes but will also, in turn, encourage high/deep learning (the backwash effect) in classrooms. Another

very critical way forward could be initiating and facilitating classroom based research in the areas of assessment and learning reforms. The schools, instead of being used as being assessed sites; and knowledge receivers could have a more active role in knowledge generation and its dissemination in exploring feasibility and outcomes of the LSAs to nurture and improve learning.

Moreover, the existing LSAs, in the contexts, has been supported mainly by Western based foreign funding agencies. Perhaps, the sponsors had assumed contextual relevance of these assessment methods in the developing country, like Pakistan. Our analysis is that the underpinning assumptions of these LSAs are not contributing to any substantive educational reforms. Perhaps the cost and effort could be used for more effective ways.

We recommend that the assessment organisations, in collaboration with other educational institutes, should provide practical support and training to facilitate teachers and teacher educators to understand and learn to use assessment information to raise learning standards. Since mainly these assessments have been conducted in public school context, where accountability measures mainly have no meaning to the public sector; therefore, these assessments should need to revisit their agenda from reporting to providing quality support to teachers and schools. There is a need for consistent- hands on training to help teachers understand how to use assessment data for formative purpose, how to use feedback effectively, how to write assignments and/or design good tests, and, how to record student results simply and clearly in order to stimulate each student's learning progress. For this to happen there is a need for all, schools, assessment organisations and the policy makers, involved to: Be clear about the learning goals and the success criteria by which learning will be judged, sharing them with students using student-friendly language; use alternative assessment methods, including authentic assessment, that provide opportunities for assessing students' deeper knowledge and understanding; provide teachers with skills to understand, analyse and address students' errors adequately; develop new forms of assessment; and embed emerging technologies into the assessment schemes. Hence, we conclude that LSAs role to be rethought and transformed towards facilitating learning from its existing reporting function.

Table 1
Descriptive Analysis of Large-Scale Assessments in Sindh

AREAS	SAT	ASER
Purpose	Know students' achievement Link educational reforms to output/outcome rather than input Inform parents, educational administration, civil society and government about the status of education in the province.	Provide data on students cognitive and competencies Increase social mobility and enrolment Outreach all children across the country
Sample Size	Grade 5 and Grade 8 Entire Population Expected 300,968, Appeared 177,474 -58.97% [Grade 5) Grade 8 Expected 214,616 Actual 108408 (50.51%)	Age 5- 16 children Since 2008 – 2015 421, 735 household survey administered; 1,271,606 children assessed and 29,168 schools (public and private) profile collected. 42 874 volunteers citizens involved in the administration
Subjects Tested	Three subjects Urdu/ Sindhi/ English (depending on the medium of school. Mathematics and Science :	Numeracy and Reading skills
Test Administration	In one day 3- hour paper for four subjects -at the selected examination centres. Examination conditions	On weekends in home environment
Rationalisation for the Achievement Test	International trends TIMSS, PIRLS and PISA with a caveat that these test are held over a cycle of 4- 5 years. National trends Punjab Examination Commission that test grades 5 and 8	International agenda – education in developing country Ranking countries
Test Strands	Each subject test comprised of several content strands or main content areas of learning. Two main content strands for Language Were Reading comprehension and Writing. Math test was based on five content strands including Numbers and Operation, Algebra, Measurement, Geometry, and Information Handling. Science test comprised of three content strands, Life Science, Physical Science, and Earth and Space Science.	Grade 2 is focused to measure the linguistic and numeracy companies
Test Results subject wise	Language scores for all districts are significantly higher than math and science scores. With a few exceptions, most of the districts appear to be performing between 38% and 45% language average scores. Districts' math and science average scores appear to be concentrated between 15% and 20%,	Government fails to achieve enrolment targets -education in Sindh is poor
Dissemination	Report cards to parents.	Reports /media coverage and international seminars
Frequency	Annually	Annually

Table 2
SAT % Results 2013/14- 2015/16

Subject	Grade 5			Grade 8		
	SAT 1V (2015/16)	SAT III (2014/15)	SAT II (2013/14)	SAT 1V (2015/16)	SAT III (2014/15)	SAT II (2013/14)
Language Average %	32.32 /32.80	32.81	30	37.58/36.93	40.48	37.01
Mathematics Average %	23.61/24.09	18.22	17	22.37/21.95	17.62	13.73
Science Average %	23.18/23.68	15.26	18	24.87/24.37	17.17	17.07
Overall Average %	26.37/26.87	22.1	21	28.27/27.75	25.09	22.6

Table 3
An overview of ASER information (%) in Province Sindh (ASER 2016: 11)

Year	% of children able to do at least subtraction		Children able to read at least a sentence in Urdu/ Sindhi		Children able to read words in English		Enrolment of gender in Gov schools		Enrolment of gender in private schools	
	B	G	B	G	B	G	B	G	B	G
2014	32	25	36	29	31	25	65	35	62	38
2015	37	31	40	33	36	31	64	36	62	38
2016	32	24	34	25	26	19	62	38	64	36

References

Annual Statutes of Education Report, ASER Pakistan . (2017). *ASER Report.* Retrieved from http://aserpakistan.org/document/aser/2016/reports/national/Annual-Status-of-Education-Report-ASER-2016.pdf

Clarke, M. (2012). What matters most for student assessment systems: A framework paper.

Education Monitor III . (2016). *Lahore: Society for the Advancement of Education.* Solution Tree Press.

Gorard, S. (2010). Serious doubts about school effectiveness. *British Educational Research Journal*, *36*(5), 745–766.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Klinger, D. A., Rogers, W. T., Anderson, J. O., Poth, C., & Calman, R. (2006). Contextual and school factors associated with achievement on a high-stakes examination. *Canadian Journal of Education/Revue canadienne de l'éducation*, 771–797.

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Heinemann Portsmouth, NH.

Koretz, D., & Hamilton, L. (2000). Assessment of students with disabilities in kentucky: Inclusion, student performance, and validity. *Educational evaluation and policy analysis*, *22*(3), 255–272.

Lewis, A. C. (2000). *High-stakes testing: Trends and issues*. Mid-continent Research for Education and Learning Aurora, CO.

Nagy, P. (2000). The three roles of assessment: Gatekeeping, accountability, and instructional diagnosis. *Canadian Journal of Education/Revue canadienne de l'éducation*, 262–279.

National Education Policy . (2017). *Ministry of Education Government of Pakistan.* Retrieved from http://www.moent.gov.pk/userfiles1/file/National%20Educaiton%20Policy%202017.pdf

Phelps, R. P., et al. (2008). The role and importance of standardized testing in the world of teaching and training. *Nonpartisan Education Review*, *4*(3), 1–9.

Popham, W. J. (2001). *The truth about testing: An educator's call to action*. ASCD.

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.