Author: By Newton Murani, Kenya National Examinations Council: *nmurani@knec.ac.ke* +254722597827

Title: # Comparison of Grading Approaches and Reporting Learner Achievement

Theme: **Assessment and Big Data**

Sub-theme: **Reporting on Progress and Achievement Using Big Data**

Abstract

In every Educational system reporting of learner achievement is a critical aspect that educationists strive to address and improve continuously. Accurate reporting of learner achievement on attained competencies such as knowledge, skills and attitudes provides relevant feedback information in both formative and summative assessments that is useful for deciding appropriate remediation, placement for transition, certification and policy.

The purpose of this study was to investigate styles of reporting learner's achievement in large scale testing. Focus was made on comparison of grading candidate's scores using raw marks as awarded by examiners and by use of statistically computed standardized scores in order to establish the most effective approach.

The large data for the study was obtained from summative examinations offered in the Kenyan education system for the Kenya Certificate of Secondary Education in five domain or subject areas of the year 2017. Descriptive statistics was used to analyze the quantitative and qualitative data obtained and reported using tables, frequency polygons and graphs following the use of MS excel application. The study established that there are more merits in the use of statistically computed standardized scores than the use of raw scores in reporting learner achievement. Further, use of rubrics to evaluate and express the measurement scores obtained was found to be highly meaningful and useful while disseminating results using big data in digital format was a great challenge. The study mainly recommends greater adoption and use of statistically computed standardized scores combined with descriptive rubrics in order to promote fairness and efficiency in reporting of learner achievements in various educational systems

1.0 Introduction

1.1 Background

Kenya National Examinations Council (KNEC) is an examination board which develops, administers and analyzes various national summative examinations for placements and certification purposes at different levels in the education sector. The assessment results from these examinations are consumed by a wide network of stake holder who includes the candidates, parents, teachers, curriculum designers, government policy makers and various employers.

1.2 Statement of the problem

The development of any nation or society is based on formulation of good policy which depends heavily on informed decision making from learners' and candidates' assessment results. Approaches to assessment, measurement and evaluation of learners and candidates' performance include those with either criterion or norm reference base and one wonders which approach is most suitable. In many situations the presentation for dissemination of assessment results is often ineffective resulting in failure to meet set objectives. This study aims at investigating efficiency and establishing effective grading approaches in assessment process involving measurement, evaluation, presentation of results and dissemination of learners' performance in results in Kenya.

1.3 Objectives of the study

The objectives of this study were to establish the:

- outcomes of grading candidates performance using raw marks by criterion referenced approach;
- outcomes of grading candidates performance using standardized scores by norm referenced approach;

- effectiveness of the use of assessment rubrics in reporting candidates' performance in an examination;

- effectiveness of the methods used to disseminate assessment digital data to stake holders.

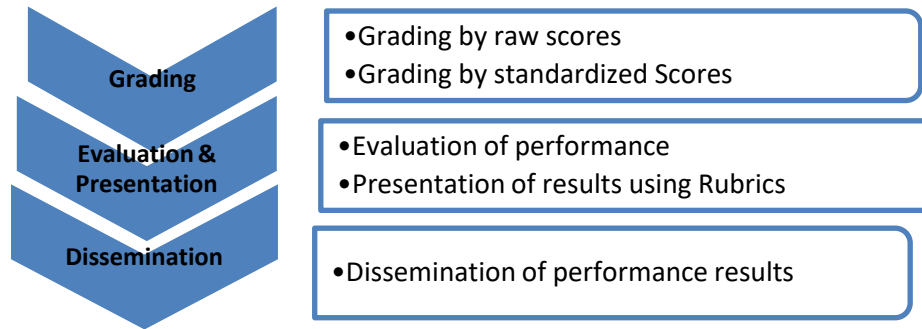## 1.4   Research questions

The specific research questions of the study were:

- What are the outcomes of grading candidates' performance using raw marks in KCSE examination?

- What are the outcomes of grading candidates' performance using standardized scores in KCSE Examination?

- What do assessment rubrics convey in communicating candidates' performance in the KCSE examination?

- Were the methods used to communicate assessment results to stake holders quick, timely and efficient?

1.5   Scope and Limitations of the study

The data for this study was collected from the records of Kenya National Examinations Council (KNEC) targeting the population of all the candidates' in the country who sat Kenya Certificate of Secondary Education (KCSE) examination in the year 2017 in five subject areas namely; Languages, Mathematics, Sciences, Humanities and applied subjects

1.6   Conceptual Framework: Grading, Evaluation, Presentation and Dissemination

Grading	•Grading by raw scores •Grading by standardized Scores
Evaluation & Presentation	•Evaluation of performance •Presentation of results using Rubrics
Dissemination	•Dissemination of performance results

1.7   Definition of operational terms

Assessment; Process of obtaining information about the extent to which learners have acquired desired learning outcomes that include; developing tests and examinations, administering those tests and examinations, marking and awarding measurement scores, analyzing and awarding measurement grades and evaluating the measurements to give them meaning.

Examination; A battery of tests or series of tests designed to measure different attributes of learners such as behaviour.

Measurement; Process of assigning numerical values such as marks and grades according to specified criteria of performance.

Evaluation;   Process of making value judgment on an assessment measurement score to interpret learning achievement.

Summative assessment: Process of obtaining learning outcomes at the end of a course, program of study or specified periods of time

Big data: information gathered on a large population by way of census or information gathered in successive intervals all leading to accumulation of huge information that requires speedy processing using electronic devices and software to handle the data in digital format.

Assessment rubric; A guide with descriptors that lists specific criteria for awarding marks, grades or scores in an, examination.

Standardized score; The value of statistical measurement obtained by use of the mean and standard deviation of a given distribution.

2.0   Literature Review

2.1   Assessments

The purpose of criterion referenced testing is to determine whether each learner has achieved specific concepts or skills while norm referenced testing aims at ranking each learner with respect to the achievement of others in broad domains of knowledge. Another main difference is that in criterion referencing each learner is compared with a set standard for acceptable achievement while in norm referencing each learner is compared against other examinees and may be expressed as a percentile. (Hutt, 1996)

2.2   Large Scale Assessments

In a study report on using large scale assessments of students` learning to inform education policy M Tobin et al (2015) described large scale assessments as follows:

- may be either at the sub-national, national or international levels;
- may be conducted in centralized or decentralized education systems;
- are intended to be representative of education systems;

- are usually **standardized** to facilitate comparison across learners, subjects/schools and countries;

- not purposefully meant for certifying individual learners` achievement;

- do not refer to classroom assessments used by teachers;

- are not meant mainly for purposes of selection, placement, graduation or university entrance.

2.3   Big data

Big data has been viewed as a term used to describe data sets so large that they can only be analyzed by computers, and analytics is used to describe how the data is collected, analyzed and used (Strauss, 2016). Four measurement challenges of big data in assessments that are enabled by technology have been identified as follows:

- dealing with change over time via time-based data;

- interaction of digital performance space's with learner actions, communications and products;

- interpretation formed from translations of atomistic data into meaningful larger units suitable for making inferences about learners performance and;

- representation of the dynamics of interactions between and among learners when being assessed through interactions with each other and digital resources and agents in digital performance spaces. (Gibson et al, 2015)

2.4   Use of Statistics in assessments

Commonly used in statistics for data analysis are measures of central tendency namely the mean, mode and median. Other useful values include variance, standard deviation standardized scores and Z- scores. Information extractable from such values and graphics are whether the data is uniformly distributed around the median mark or skewed and how well scores are spread between the minimum and maximum scores. A good examination is one in which scores of learners are spread out showing wide variability in order to discriminate different abilities among the learners. Measures of variation for numeric data include: range, inter-quartile range, mean absolute deviation, variance, standard deviation and coefficient of variation. (Jaisingh, 2000)

3.0   Research Methodology

3.1   Research design

A descriptive survey research design was adopted for this study. Inferences were made based on the independent variable of raw scores that had already been awarded to the candidates before the study commenced.

Study population

The target population of this study was all the candidates who sat Kenya Certificate of Secondary Education (KCSE) in five domain or subject areas namely; Languages, Mathematics, Sciences, Humanities and Applied subjects.

Sampling procedure and data collection

The study population was randomly picked from three main levels of education in Kenya namely; Basic Education primary, Basic Education secondary and tertiary education. In this case secondary education candidates were picked. Purposive sampling was then applied to select candidates of five domain or subject areas to have at least one subject represent each subject grouping. The subjects

picked were: English and Kiswahili for languages, Mathematics alternative (A) represented Mathematics group, Biology for Sciences group, History for the humanities group and Business Studies for the applied subjects.

The huge data collected was obtained from the records of Kenya National Examinations Council (KNEC) by use of a checklist in its raw form.

3.2 Data analysis

Microsoft excel application software was used. The huge data was grouped in class intervals and class midpoint obtained before determining measures of central tendency; mean, variance, standard deviation and Z-scores.

Grading levels can vary depending on the skill levels to be attained and can be as few as two and as many as twelve. KNEC uses a twelve level scale and this study adopted a four level scale to satisfy the objectives of the study. Thus four level predetermined aggregate scales for the raw sores on a criterion performance basis were used. A similar four level scale was used to grade the measurement scores by use of statistical measurement values of standard deviation and z-scores on a norm reference basis. The candidates' performance was further expressed using assessment rubrics for interpretation purposes. The reporting was done using tables, frequency polygons, Charts and graphs.

4.0 Analyzed Results of the Subjects in the study

4.1 Results of performance in paper 1

A Graph of Candidature Versus Score Class Interval for the Subjects Paper One (1)
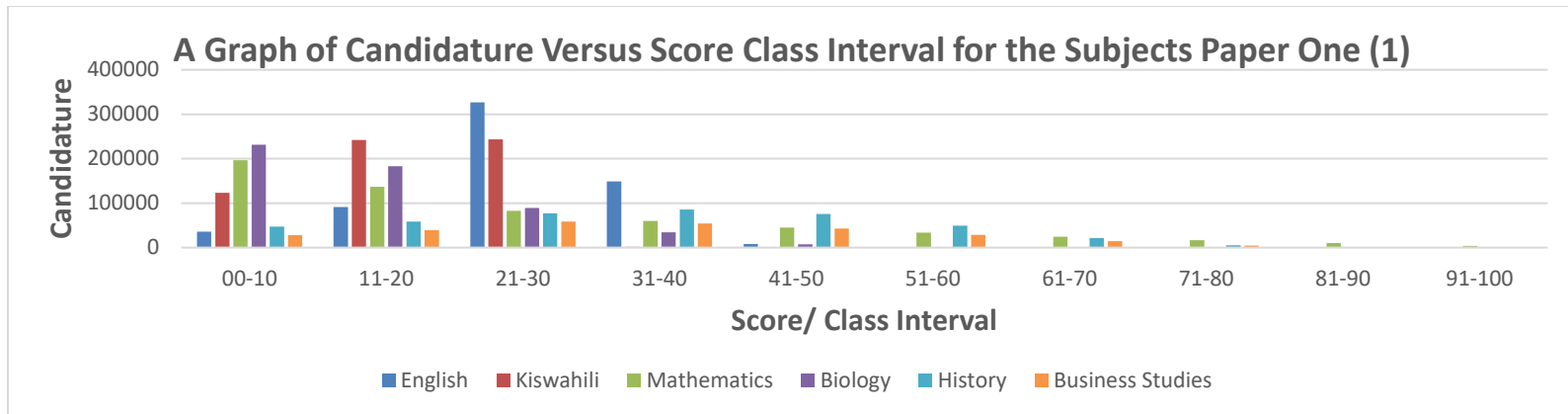
The bar chart above shows a distribution in which English, Kiswahili, Mathematics and Biology had shift to the left depicting positive skew. History and Business Studies had a near bell shape normal distribution.

4.2   Results of performance in paper 2

Graph of Candidature Versus Score Class Interval for the Subjects Paper two (2)
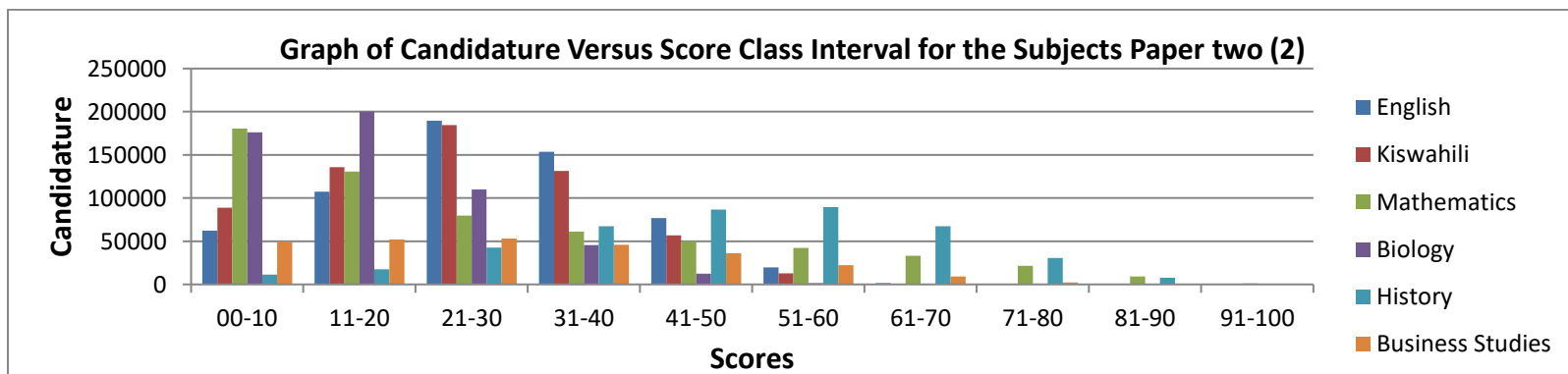
The bar chart above shows that in Paper two category just like in paper one, English, Kiswahili, Mathematics and Biology had a positive skew. History and Business Studies had a near bell shape normal distribution.

4.3 Grading by raw Scores – Criterion Referenced approach

The scores were graded in a quartile or four level of pre-determined scale as shown in the table below.

Quartile Score Range	Percentage Score (%)	Grade
4th	(76 –100)	A
3rd	(51 – 75)	B
2nd	(26 – 50)	C
1st	(00 – 25)	D

4.4 Results of Graded raw Scores

The graded results of the subjects in the study appeared as shown in the table below. As evident from the table most candidates fell in the lower quartiles or below the 50% score mark. Candidates were being measured against a fixed scale of reference. It is not possible to effectively compare and tell in which papers the candidates performed better by merely looking at these results.

| Grade	English Paper 1	English paper2	Math's paper 1	Math's paper 2	History Paper 1	History paper 2
A	00	00	3.47	3.21	0.42	4.65

B	0.01	3.47	11.04	14.40	17.70	41.88
C	54.20	53.30	23.36	24.19	47.89	42.31
D	45.79	43.22	62.12	58.20	34.00	11.16

Grade	Kiswahili paper1	Kiswahili paper 2	Biology paper 1	Biology paper 2
A	00	00	00	00
B	00	2.18	0.12	0.27
C	9.81	45.48	14.22	18.94
D	90.19	52.34	85.66	80.79

4.5   Grading By Use of Standard Scores – Norm Referenced Approach

The standardized scores were graded in a four level scale of reference that was determined from the computation of candidates' scores as shown in the table below.

Score Range (s)	Grade	Population %
$S > +1$ SD	A	16
$M \leq S \leq +1$ SD	B	34
-1 SD $\leq S \leq M$	C	34
$S < -1$ SD	D	16
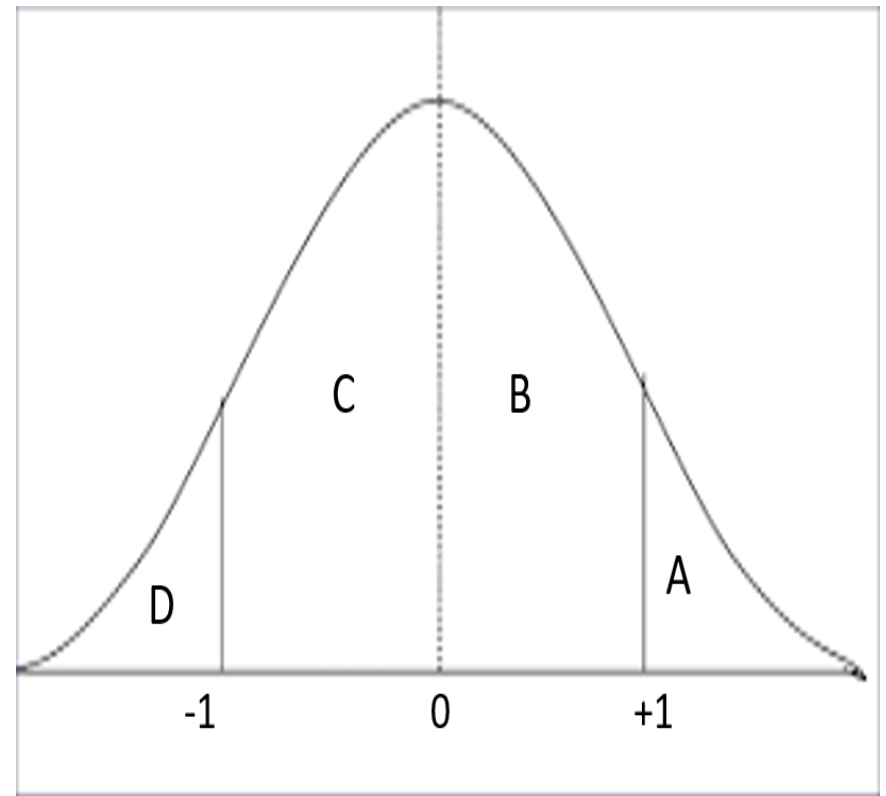
4.6 Results of Graded Standardized Scores

The first main step in this approach is to calculate the mean and the standard deviation of each examination paper from the obtained scores. The statistical mean is a value that allows comparisons and one can tell in which paper the performance was better. Even an individual candidate can

compare performance of self with the mean. The standard deviation indicates the spread or variability of the scores hence depicting in which paper discrimination of candidates in terms of ability was better. The table below shows the mean and standard deviation values for the subjects in the study.

| Subject	English Paper 1	English Paper 2	Kiswahili Paper 1	Kiswahili Paper2	Math Paper 1	Math paper2	Biology paper1	Biology paper 2	History paper 1	History paper 2	Business S Paper1	Business S Paper 2
Mean	25.7	28.0	18.2	25.2	24.8	24.7	14.1	16.6	33.7	47.9	33.5	28.5
Standard Deviation	7.6	12.0	6.3	12.1	21.8	22.2	10.1	10.3	19.4	27.5	19.2	17.5

4.7 Further the results of the graded subjects using standard scores were as captured in the table below. It is notable that it is possible to escalate the scale to more than four levels as may be desired. It is evident that unlike in the case of grading raw scores here one may not only make effective comparisons, but also obtain performance grades at all the levels of the scale in predictable measures. Proportions or percentages of the candidates at the top and at the bottom of the scale are equivalent. When this is applied to examination papers in a series of years the results would form predictable patterns which can be very useful for policy and planning purposes.

GRADE	MATHS PAPER 1	MATHS PAPER 2	HISTORY PAPER 1	HISTORY PAPER 2
A	97,511	49,599	67,405	67,404

B	195,023	194999	143,236	143,234
C	185,023	184,999	143,236	143,234
D	97511	49,499	67,405	67,404
TOTAL	**609,448**	**609,374**	**421,281**	**421,276**

4.8   Comparison of performance across subjects or papers

When graphs of scores versus frequency of candidates are plotted for different papers or subjects on the same scale the distribution can facilitate easy comparison like in the case below for History papers 1 and 2. It is easy to compare the range and the spread or variability of the distribution.
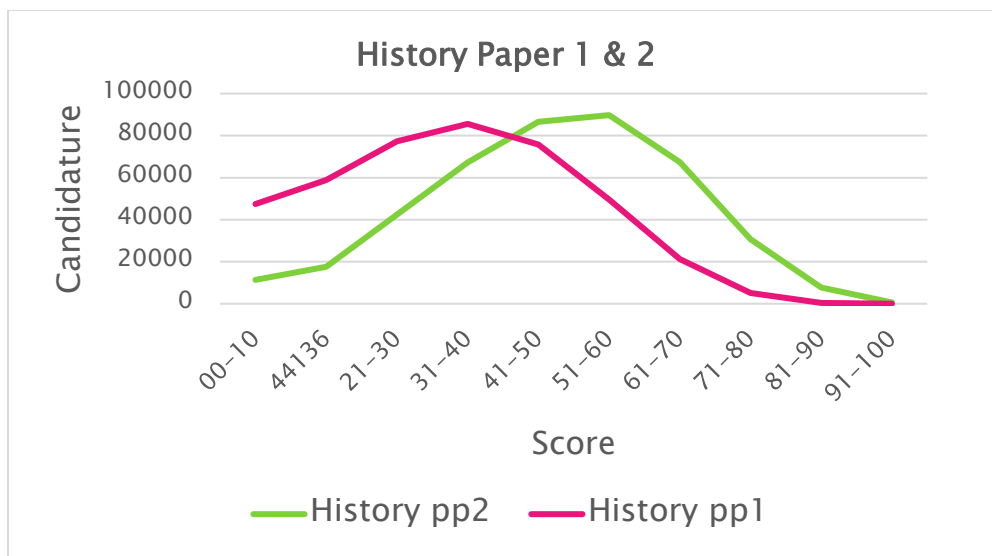
4.9   Assessment rubrics

Use of assessment rubrics in description of performance for History papers

Measurement Grade	Candidature Paper 1 (%)	Candidature Paper 2 (%)	Descriptor	Descriptor	Detailed Descriptor
A	0.42	4.65	Excellent	Exceeds Expectations	Learner is good at tackling higher order questions
B	17.70	41.88	Good	Meets Expectations	Learner is good at applying learnt concepts in real life
C	47.89	42.31	Fair	Approaches Expectations	Learner has good memory and comprehends learnt concepts well
D	34.00	11.16	Poor	Below Expectations	Learner can only tackle recall questions well

The rubric shown above have descriptors that provide more information about the performance when they accompany the grades than if the grades only were given. The more detailed the descriptor is the more it becomes informative and the better for the consumers of the information.

4.9   Release and communication of assessment results to stake holders

From the examination time table and program of handling the examination process the following were noted:

Activity	Execution	Comments
Marking of scripts	Completed on time; as scheduled. Within 3 weeks	Well executed
Transmission of raw marks from marking centre electronically	Done immediately marking was completed. Within minutes.	Efficiently executed
Processing of results and release	Done immediately marks were captured	Completed in good time
Access of results via mobile service provider	Delays of upto 48hrs	Technological system challenge with big data
Many stake holders requested for results via short messages (SMS) platform.	A total of 424,011 cases failed to get the results via SMS and were later refunded their payment.	The sms data platform could not handle the big data requested.

The process of marking and compilation of candidates' marks was well planned and executed thereby reducing time wastage. The master server computer was able to handle the big electronic data generated. However information dissemination with the contracted mobile service provider resulted in lengthy delays that were so acute that even after two days no feedback message of results was obtained. The disappointed customers ended up being refunded what they had been charged for sending the short message. The systems could not handle the big electronic data generated. To efficiently manage big data appropriate electronic hardware and software resources together with relevant expertise are needed.

5.0   Conclusion and Recommendations

5.1   Conclusion

From the above findings the following was deduced:

- Grading raw marks did not facilitate effective comparisons in performance;

- Grading using statistically computed scores enabled comparisons with other test takers and subject papers;

- Grading using standard scores maintained proportions of candidates falling in each grade which can be useful for policy and planning;

- Assessment rubrics with detailed descriptors present better and convey more useful information about performance than grades alone;

- Handling big data electronically improved efficiency in terms of the short time taken to process results;

- Transmission of results via mobile phone service provider experienced significant delay challenges**.**

5.2   Recommendations

The study made the following recommendations:

- Use raw scores for grading in a criterion referenced approach where candidates performance need not be compared with that of other test takers;

- Apply standardized scores to grade in a norm referenced approach where comparisons with other papers, subjects or test takers are required;

- Accompany the awarded measurement grades with assessment rubrics that are as detailed as possible;

- Manage big data with adequate electronic hardware and software resources to improve dissemination of results.

References

- Bennett R (2011). Assessment in Education: Principles, Policy and Practice Princeton, NJ, USA,

- Cope B & Kalantis M (2016). <u>Big Data Comes to School: Implications for Learning, Assessment, and Research </u>Washington DC; American Educational Research Association (AERA).

- Gibson D, Webb M & Ifenhaler D (2015). <u>Challenges of Big Data in Educational Assessment</u> Bentley WA 6102, Australia; Curtin University

- Great School Partnerships: <u>The Glossary of Educational Reform</u>: https://www.edglossary.org/assessment/

- Huitt, W. (1996). <u>Measurement and evaluation: Criterion- versus norm-referenced testing. Educational Psychology Interactive</u>. Valdosta, GA: Valdosta State University. <u>http://www.edpsycinteractive.org/topics/measeval/crnmref.html</u>

- **Jaisingh L (2000).** <u>Statistics for the Utterly Confused: </u>NewYork. The McGraw-Hill Companies, Inc

- **Moxley J (2013).** <u>Big Data, Learning Analytics, and Social Assessment1</u>, University of South Florida

- **Portela F, Lima L & Santos M** (2016) <u>International Symposium on Web of Things and Big Data (WoTBD 2016); Why Big Data? Towards a project assessment framework</u>, Portugal, University of Minho

- Research and Development, Educational Testing Service, http://www.informaworld.com/smpp/title~content=t713404048 F

- **Strauss V, (2016).** <u>Big data was supposed to fix education. It didn't. It's time for small data</u>.

  <u>https://www.woshingtonpost.com/people/valerie-strauss/</u>

- Tedesco J (2016). <u>Ten Notes on Learning Assessment Systems </u>Coordination and Production Team at the UNESCO IBE