Level of Agreement between Two Examiners:
A Reliability Study of E-Marking at the
Aga Khan University Examination Board

Dr Naveed Yousuf, Dr Shehzad Jeeva, Tooba Farooqui, Rabia Nisar

Aga Khan University Examination Board, Pakistan

Education System in Pakistan

- Education System
  - ➤ Primary (grade 1-5),
  - ➤ Middle (grade 6-8),
  - ➤ Secondary/Higher Secondary (grade 9-12)
  - ➤ Tertiary Education

- The system faces critical issue of access and quality **(25M Out of School Children)**

- High dropout rates

- High stake Board Examinations - major issues of quality of assessment and conduct of examination – *Cheating, leakage of exam papers, rote learning, poor assessment and marking etc.*

- 30 Government Examination Boards **(Operating District-wise)**

AKU Examination Board

- Establishing AKU-Examination Board

  ➢ Needs – Secondary and Higher Secondary Education

    ▪ Fair and transparent system
    ▪ Promote conceptual learning
    ▪ Quality of Assessment
    ▪ Improve classroom teaching

  ➢ Established in November 2002 – Ordinance of Government of Pakistan

    ▪ First private university-led national examination board
    ▪ Secondary School Certificate – SSC (Grade 9th/10th)
    ▪ Higher Secondary School Certificate – HSSC (Grade 11th/12th)

AKU Examination Board

Vision

AKU-EB aspires to be a model of excellence and innovation in education for Pakistan and the developing world.

Impact

a) Increase education quality for low and middle income students at secondary and higher secondary school level.

b) Inculcate a culture of higher order thinking in youth through high quality, affordable assessments and educational development.

Goal

Catalyze the market to impact quality of exam boards across the nation.

Aim

AKU-EB aims to make decisions based on best assessment practices for all examinations that are valid, reliable and fair

Valid, Reliable and Fair Assessment

- Item Development: valid, structured and objective model answers and marking criteria

- Peer Specialist Review

- Teacher Review: affiliated institution

- Multi-disciplinary Review

- Chief Examiner Review: external (non-affiliated), renowned, senior and respectable member in the discipline

Peer Review

Multidisciplinary Review

Expert Review

Valid, Reliable and Fair E-Marking

- Seeding: senior markers; sample-based

- Marker Orientation and Training

- E-marking in Groups: 5 markers, 1 senior marker

- Anonymous, Random allocation of items

- Senior Marker: Reviews 30% of all marked items

- Subject Specialist: Verify 10% of all reviewed items

Markers are trained

30% of marked items are reviewed

10% of reviewed items are verified

Research Objective

The present study aims to assess the inter-rater reliability of scores by two independent markers

Methods

01 This examination was already marked following the AKU-EB e-marking process and results were announced.

02 Independent re-marking of a representative sample of students was conducted using markers and senior markers.

03 The original scores were called May scores and the re-marking scores as QA scores.

04 Statistics used to analyze:
- Descriptive statistics
- Correlation
- Intra class correlation coefficient (inter-rater reliability)

Method: Sampling

05 Independent remarking of CRQs of the representative sample of :

- ➢ 300 scripts for Pakistan Studies
- ➢ 200 scripts in Mathematics
- ➢ 100 scripts for Banking

06 Selection criteria of candidates for all three subjects:

Subject Score	Candidates
90% +	7%
80% +	17%
70% +	27%
60% +	27%
50% +	17%
33% +	7%
Total	**100%**

Analysis for Pakistan Studies

Descriptive Statistics

Variable	N	Mean	SD	Minimum	Maximum
May Score	300	19.6	4.4	5	28
QA Score	300	16.2	4.3	4	25

Correlation and Regression Curve Fit
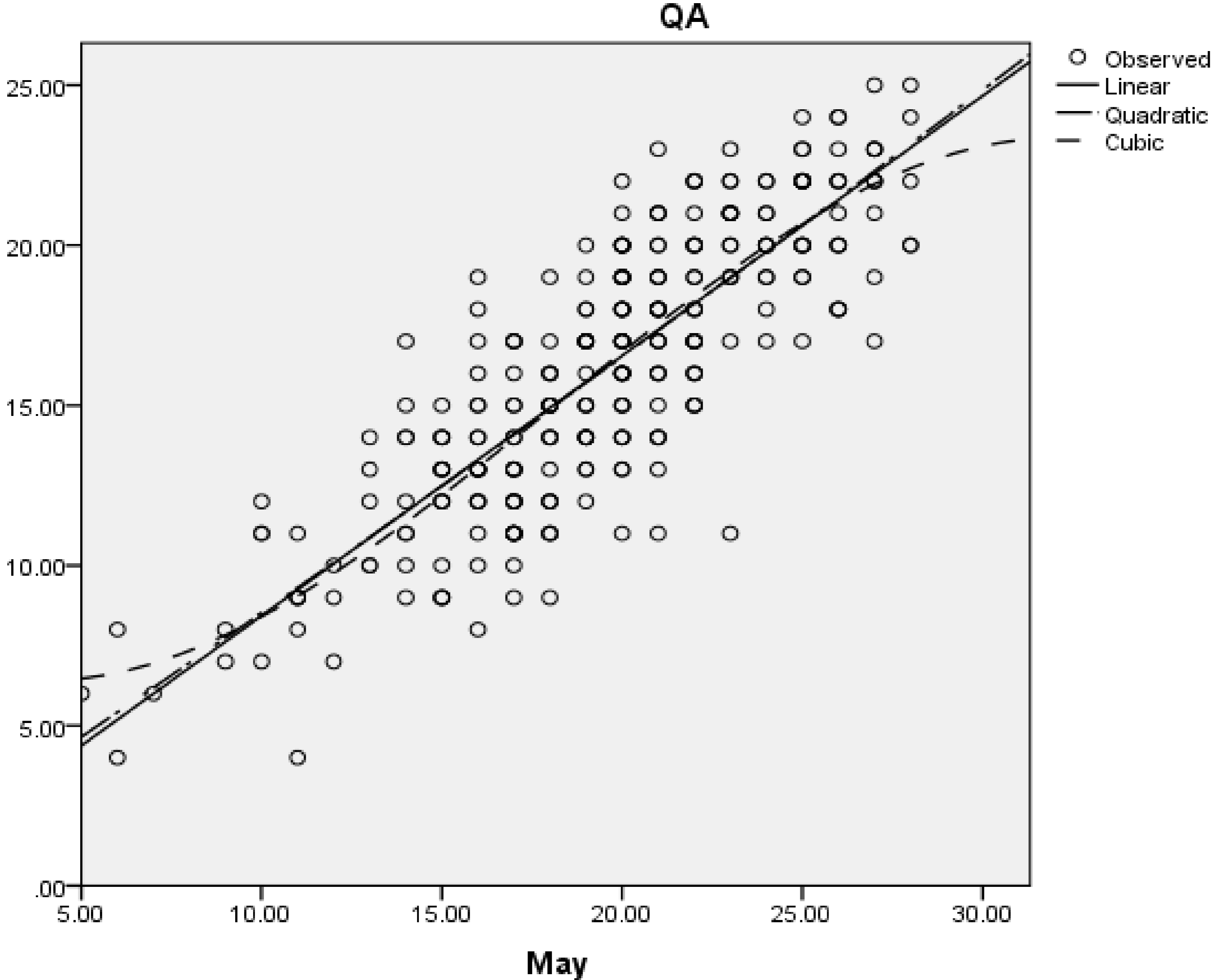
Variables	N	Correlation
May & QA Scores	300	0.84

Information:

Independent Variable=May

Dependent variable = QA

69.8% of variability in QA scores is due to the variability in May Scores

Intra-class Correlation Coefficient

Intra-class Correlation	
Single Measures	0.642

P value <0.01

Information:

➤ Two-way random effects model where both people effects and measures effects are random.

➤ Type A intra class correlation coefficients using an absolute agreement definition.
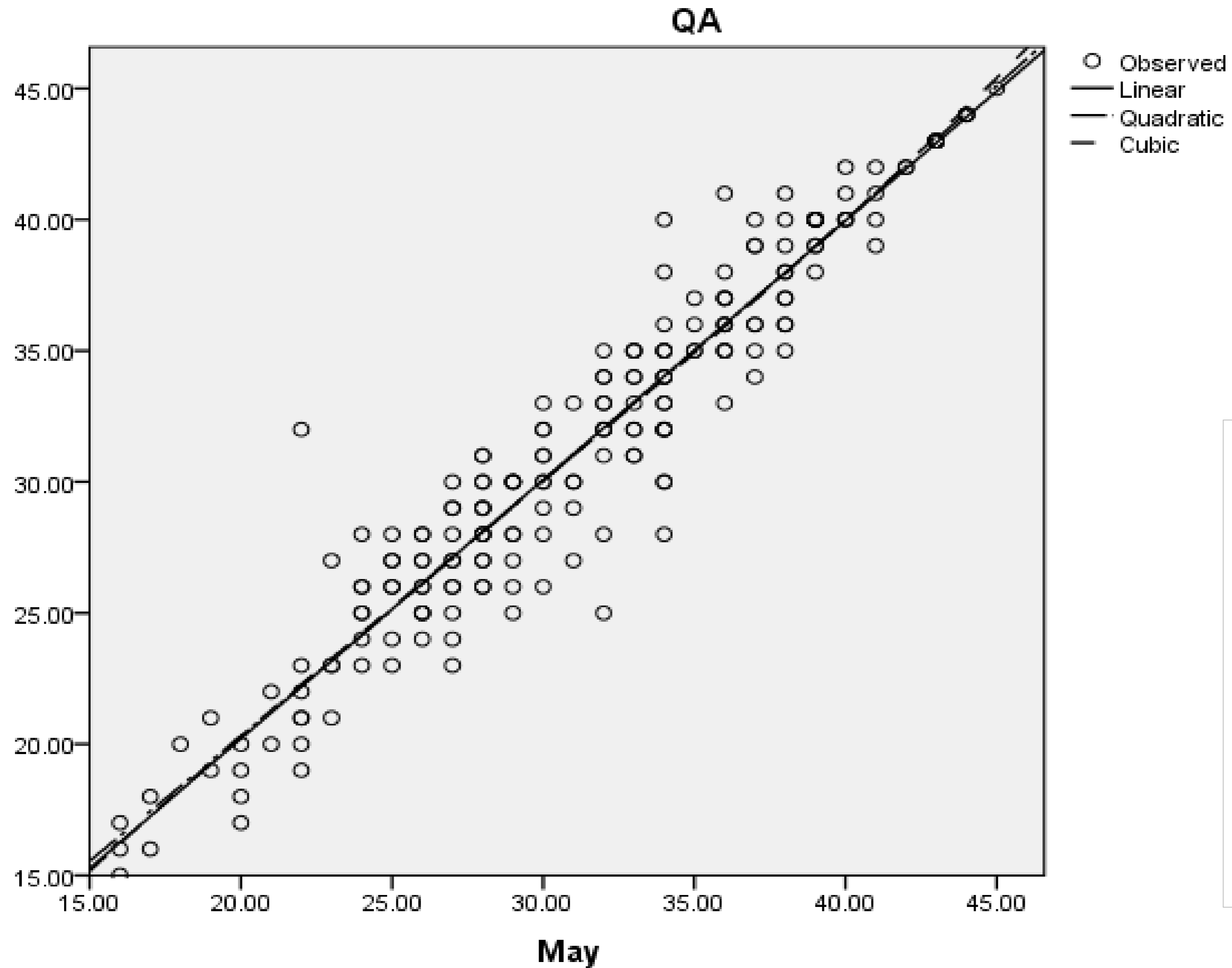
Analysis for Mathematics

Descriptive Statistics

Variable	N	Mean	SD	Minimum	Maximum
May Score	200	31.3	6.6	16	45
QA Score	200	31.3	6.8	15	45

Correlation and Regression Curve Fit

QA

Variables	N	Correlation
May & QA Scores	200	0.96

Information:

Independent Variable=May

Dependent variable = QA

91.4% of variability in QA scores is due to the variability in May Scores

Intra-class Correlation Coefficient

Intra-class Correlation	
Single Measures	0.955

P value <0.01

Information:

➢ Two-way random effects model where both people effects and measures effects are random.

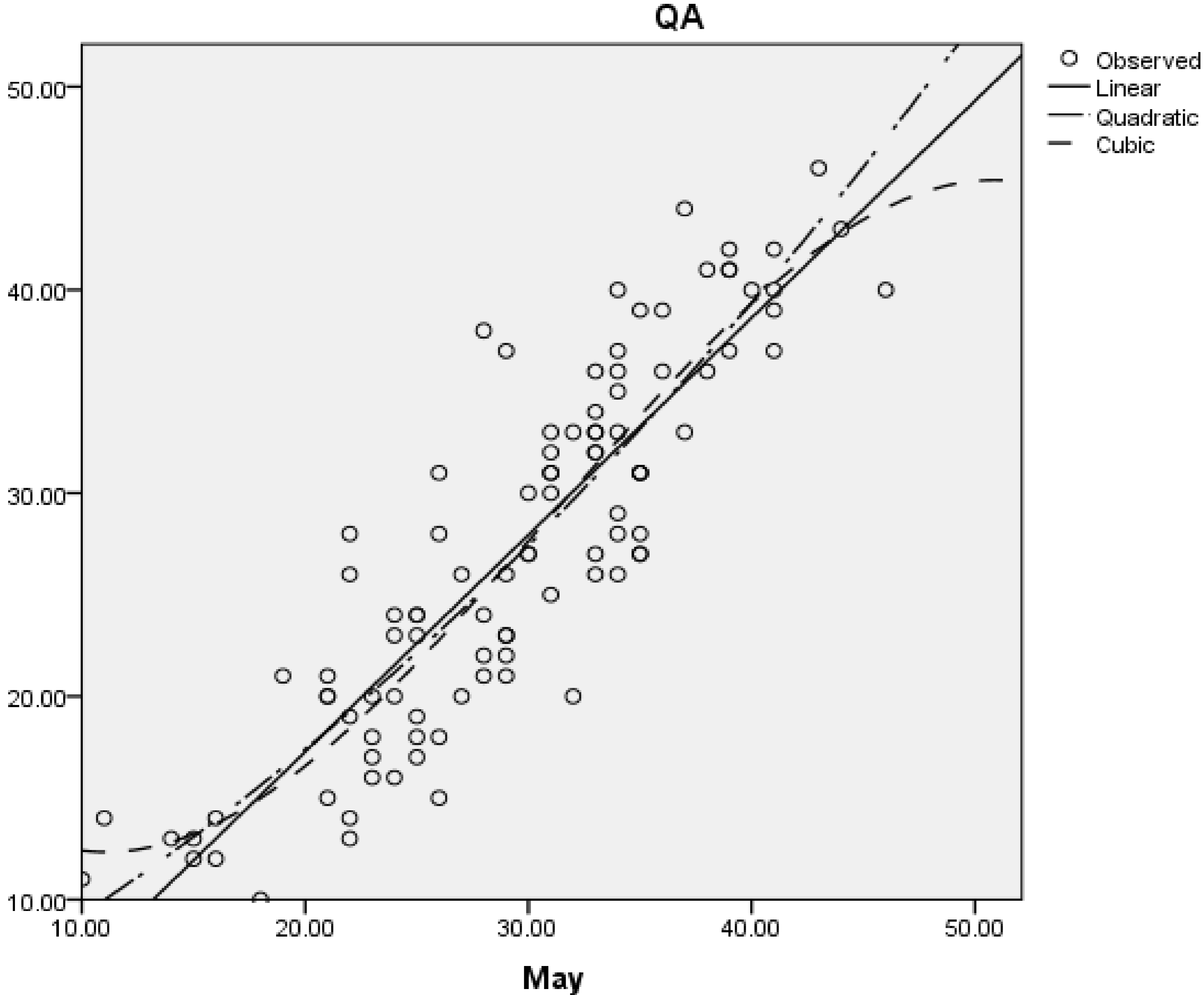➢ Type A intra class correlation coefficients using an absolute agreement definition.

Analysis for Banking

Descriptive Statistics

Variable	N	Mean	SD	Minimum	Maximum
May Score	100	29.5	7.5	10	46
QA Score	100	27.4	9.1	10	46

Correlation and Regression Curve Fit

Variables	N	Correlation
May & QA Scores	300	0.88

Information:

Independent Variable=May

Dependent variable = QA

77.2% of variability in QA scores is due to the

Variability in May Scores

Intra-class Correlation Coefficient

Intra-class Correlation	
Single Measures	0.837

P value <0.01

Information:

➢ Two-way random effects model where both people effects and measures effects are random.

➢ Type A intra class correlation coefficients using an absolute agreement definition.

Conclusion

- The intra class correlation coefficients between May and QA scores observed are:
    - Mathematics: 0.96
    - Banking: 0.88
    - Pakistan studies: 0.642

- The order of inter-rater reliability across subjects reflects their intrinsic subjectivity

- There quest for further improving the quality of marking shall continue at AKU-EB

- The lessons learned from the present study will facilitate to achieve further excellence towards valid, reliable and fair assessment decisions

Thank you

NAVEED.YOUSUF@AKU.EDU

SHEHZAD.JEEVA@AKU.EDU