

Contents

Part I Statistical Interpretation of Test/Exam Results

1	On Average: How Good Are They?	3
1.1	Average Is Attractive and Powerful	3
1.2	Is Average a Good Indicator?	4
1.2.1	Average of Marks	4
1.2.2	Average of Ratings	4
1.3	Two Meanings of Average	5
1.4	Other Averages	6
1.5	Additional Information Is Needed	7
1.6	The Painful Truth of Average	8
2	On Percentage: How Much Are There?	9
2.1	Predicting with Non-perfect Certainty	9
2.2	Danger in Combining Percentages	11
2.3	Watch Out for the Base	12
2.4	What Is in a Percentage?	13
2.5	Just Think About This	13
	Reference	13
3	On Standard Deviation: How Different Are They?	15
3.1	First, Just Deviation	15
3.2	Next, Standard	16
3.3	Discrepancy in Computer Outputs	17
3.4	Another Use of the SD	18
3.5	Standardized Scores	18
3.6	Scores Are not at the Same Type of Measurement	20
3.7	A Caution	22
	Reference	23
4	On Difference: Is that Big Enough?	25
4.1	Meaningless Comparisons	25
4.2	Meaningful Comparison	26

4.3	Effect Size: Another Use the SD	27
4.4	Substantive Meaning and Spurious Precision	29
4.5	Multiple Comparison	30
4.6	Common but Unwarranted Comparisons	31
	References	33
5	On Correlation: What Is Between Them?	35
5.1	Correlations: Foundation of Education Systems	35
5.2	Correlations Among Subjects	36
5.3	Calculation of Correlation Coefficients	37
5.4	Interpretation of Correlation	40
5.5	Causal Direction	41
5.6	Cautions	44
5.7	Conclusion	45
	Reference	45
6	On Regression: How Much Does It Depend?	47
6.1	Meanings of Regression	47
6.2	Uses of Regression	48
6.3	Procedure of Regression	49
6.4	Cautions	50
7	On Multiple Regression: What Is the Future?	51
7.1	One Use of Multiple Regression	51
7.2	Predictive Power of Predictors	53
7.3	Another Use of Multiple Regression	53
7.4	R-Square and Adjusted R-Square	54
7.5	Cautions	55
7.6	Concluding Note	56
	References	56
8	On Ranking: Who Is the Fairest of Them All?	57
8.1	Where Does Singapore Stand in the World?	57
8.2	Ranking in Education	59
8.3	Is There a Real Difference?	61
8.4	Forced Ranking/Distribution	61
8.5	Combined Scores for Ranking	62
8.6	Conclusion	63
9	On Association: Are They Independent?	65
9.1	A Simplest Case: 2×2 Contingency Table	65
9.2	A More Complex Case: 2×4 Contingency Table	67
9.3	Even More Complex Case	68
9.4	If the Worse Come to the Worse	70
9.5	End Note	71
	References	71

Part II Measurement Involving Statistics

10 On Measurement Error: How Much Can We Trust Test Scores?	75
10.1 An Experiment in Marking	76
10.2 A Score (Mark) Is <i>not</i> a Point	78
10.3 Minimizing Measurement Error	79
10.4 Does Banding Help?	80
Reference	81
11 On Grades and Marks: How not to Get Confused?	83
11.1 Same Label, Many Numbers	83
11.2 Two Kinds of Numbers	84
11.3 From Labels to Numbers	85
11.4 Possible Alternatives	87
11.5 Quantifying Written Answers	88
11.6 Still Confused?	89
Reference	89
12 On Tests: How Well Do They Serve?	91
12.1 Summative Tests	91
12.2 Selection Tests	93
12.3 Formative Tests	94
12.4 Diagnostic Tests	95
12.5 Summing up	96
References	96
13 On Item-Analysis: How Effective Are the Items?	97
13.1 Facility	98
13.2 Discrimination	100
13.3 Options Analysis	100
13.4 Follow-up	101
13.5 Post-assessment Analysis	102
13.6 Concluding Note	103
Reference	103
14 On Reliability: Are the Scores Stable?	105
14.1 Meaning of Reliability	105
14.2 Factors Affecting Reliability	106
14.3 Checking Reliability	107
14.3.1 Internal Consistency	107
14.3.2 Split-Half Reliability	109
14.3.3 Test-Retest Reliability	109
14.3.4 Parallel-Forms Reliability	109
14.4 Which Reliability and How Good Should It Be?	110

15 On Validity: Are the Scores Relevant?	111
15.1 Meaning of Validity	111
15.2 Relation Between Reliability and Validity	115
Reference	116
16 On Consequences: What Happens to the Students, Teachers, and Curriculum?	117
16.1 Consequences to Students	117
16.2 Consequences to Teachers	120
16.3 Consequences to Curriculum	121
16.4 Conclusion	122
References	124
17 On Above-Level Testing: What's Right and Wrong with It?	125
17.1 Above-Level Testing in Singapore	126
17.2 Assumed Benefits	127
17.3 Probable (Undesirable) Consequences	127
17.4 Statistical Perspective	129
17.5 The Way Ahead	131
17.6 Conclusion	132
References	132
18 On Fairness: Are Your Tests and Examinations Fair?	133
18.1 Dimensions of Test Fairness	134
18.2 Ensuring High Qualities	134
18.3 Ensuring Test Fairness Through Item Fairness	137
References	139
Epilogue	141
Appendix A: A Test Analysis Report	143
Appendix B: A Note on the Calculation of Statistics	149
Appendix C: Interesting and Useful Websites	153