

Index

Entries marked in *italics* refer to definitions given in the Glossary.

- ABEEB, 5,
- achievement tests, 12, 286
- administrators, training of, 115–8
- Alderson, J.C., 4, 6, 23, 46, 55, 73, 97, 174, 175, 176, 182, 221, 225, 227–8
- Algina, J., 76, 77, 85, 89, 91, 96, 132
- Allan, A., 45
- American Educational Research Association, 237–42
- American Psychological Association, 171
- analysis of variance, 286
- analytic scale, 108, 109, 286
- Anastasi, A., 85, 89, 96
- Angoff, W., 97
- Associated Examining Board (AEB), 19, 35, 99, 101, 109, 122–3, 140–3, 191
- Association of Recognised English Language Schools (ARELS), 31–2
- Bachman, L., 14, 17, 19, 171, 172, 174, 186, 223, 225
- BIGSTEPS, 91
- BILOG, 91
- biserial correlation, 84
- Buck, G., 6, 45, 73, 187
- C-test, 44, 56
- Campbell, D., 185, 223
- Carroll, B., 5, 172
- central marking, 129
- central tendency, 94
- City and Guilds of London Institute (C & G), 68, 100
- Clapham, C., 23, 174, 182
- classical item analysis, 80–7
 - discrimination index (D.I.), 80
 - facility value (F.V.), 81–6, 289
- cloze, 44, 55
- Cohen, A., 3, 176
- compositions and essays, 59
- computer programs
 - BIGSTEPS, 91, 284
 - BILOG, 91, 285
 - ITEMAN (Microcat), 85, 100–1, 284
 - QUEST, 91, 285
 - SAS, 85, 132, 284
 - SPSS, 85, 132, 284
- computers, uses of, 225
- concurrent validity, 177–80, 286
- consensus scripts, 111
- construct validation, 17, 183–6
 - 286–7
- constructs, 17
- content validation, 173, 287
- convergent-divergent validation, 185
- correlation, 77–80
 - biserial correlation, 84
 - coefficient, 287
 - Pearson product moment correlation, 80, 287
 - point biserial correlation, 84
 - rank order correlation, 80, 278
- Council of Europe, 28
- Criper, C., 23, 182
- criteria, 13

Language test construction and evaluation

- criterion-referencing, 76, 157, 287
- criterion validity, 171
- Crocker, L., 76, 77, 85, 89, 91, 96, 132
- Davidson, F., 14, 222
- Davies, A., 23, 97, 182
- descriptive statistics, 92
 - dispersion, 95
 - mean, 92, 291
 - median, 92, 291
 - mode, 92, 292
 - 'negatively skewed', 93, 292
 - 'positively skewed', 93, 292
 - range, 92, 293
 - significant difference, 294
 - standard deviation (S.D.), 92, 294–5
- descriptors, 287–8
- diagnostic tests, 12, 288
- Diamond, E., 242
- dichotomous items, 51
- dictation, 57
- discrimination index (D.I.), 80–6, 274, 288
- dispersion, 95
- double-marking, 132
- Douglas, D., 5
- E_{1-3} formula, 84
- Ebel, R., 149, 150, 158, 183
- editing, tests of, 53
- Educational Testing Service (ETS), 17, 229, 246
- empirical validity, 171, 288
- English for Academic Purposes (EAP), 23
- English for Business Purposes (EBP), 32
- English for Specific Purposes (ESP), 22
- English Speaking Board, 27, 31–2, 122
- English Speaking Union, 5, 28
- equivalent tests, 97, 288
- 'eyeballing', 131
- examiner, 105, 288–9
 - training of, 110–15
- exercises, 41
- external validity, 171, 177–83, 270
- face validity, 172–3, 289
- facility value (F.V.), 80–6, 289
- factor analysis, 186, 289
- Faerch, C., 176
- feedback, 200–1
- Fiske, D., 185, 223
- Fremer, J., 242
- Frisbie, D., 149, 158, 183
- Fruchter, B., 80, 89, 132
- gap-filling, tests of, 54
- generalisability theory, 289
- Gronlund, N., 183
- Grotjahn, R., 176
- Guilford, J., 80, 89, 132
- Hagen, E., 171
- Hambleton, J., 91
- Hamilton, J., 97
- Heaton, B., 3, 46, 49, 50, 235
- Henning, G., 75, 91, 170, 173, 222
- holistic scale, 107–8, 289–90
- Hudson, T., 77
- Hughes, A., 3, 23, 46
- Hutchinson, T., 22
- Hymes, D., 22, 225
- impression scale, 108, 290
- information transfer, 52
- information-gap activities, 62
- Ingram, E., 155, 172
- interlocutor, 105, 290
- internal correlation, 183
- internal validity, 171–7, 290
- International English Language Testing System (IELTS), 23
- item analysis, 80–6
- item banking, 92
- item banks, 91, 290–1

Index

- Item Response Theory (IRT),
 - 89–100, 291
 - BIGSTEPS, 91, 284
 - BILOG, 91, 285
 - item characteristic curve, 90, 291
 - logit scale, 90
 - one-parameter model, 91
 - QUEST, 91
 - three-parameter model, 91
 - two-parameter model, 91
- item weighting, 149
- item writing, 40
 - cloze, 55
 - C-test, 56
 - compositions and essays, 59
 - dichotomous items, 51
 - dictation, 57
 - editing, tests of, 53
 - essays, 59
 - gap-filling, 54
 - information transfer, 52
 - information-gap activities, 62
 - matching, 51
 - multiple choice, 47
 - objective tests, 106
 - objective types, 51, 292
 - oral interviews, 62
 - ordering tasks, 52
 - problems with particular types, 46
 - short-answer questions, 57
 - subjective tests, 107
 - subjective types, 59, 295
 - summaries, 61
 - types, 44
- ITEMAN (Microcat), 85, 100–1
- Joint Committee on Testing Practices, 242–5
- Joint Matriculation Board, 25, 32–3, 153, 165, 208–15
- Kasper, G, 176
- Kerlinger, F., 173, 186
- Klein-Braley, C., 54
- Krahnke, K., 4,
- Kuder Richardson 20 & 21 (KR20/21), 88–9, 103, 291
- Kunnan, A., 174
- kurtosis, 291
- Lado, R., 45
- Lancaster Language Testing Research Group, 5, 73
- levels of difficulty, 28
- Linacre, J., 223
- listening, 117
- logit scale, 90
- London Chamber of Commerce and Industry, 26, 32, 33, 99, 124, 140–1, 144, 164–5, 189–92
- Lopes, M., 97
- Lord, F., 91
- Lukmani, 174
- Lynch, B., 14, 76, 174, 222
- McNamara, T., 97
- Magnusson, D., 96
- main trialling, 75
- mark scheme, 107, 291
- marking
 - analytical scale, 108
 - at home, 133
 - at test centre, 134
 - central, 129
 - blind marking, 130
 - ‘eyeballing’, 131
 - reliability scripts, 131
 - routine double-marking, 132
 - sampling, 130
 - second markers, 130
 - t-test, 132, 295
 - descriptors, 107, 287
 - holistic scale, 107–8, 289–90
 - impression scale, 108, 290
 - key, 106–7, 291
 - mark scheme, 106–7, 291
 - monitoring, 140–4
 - objective, 106
 - rating scale, 107, 293
 - designing, 111

Language test construction and evaluation

- marking, (*cont.*)
 - scaling, 275
 - scripts
 - consensus, 111
 - problem, 111
 - standardisation meeting, 112
 - subjective, 107
- Masters, G., 91
- matching, 51
- Mathews, J., 106
- mean, 81, 92, 275, 291
- median, 92, 275–6, 291
 - central tendency, 94
- method effect, 44
- mode, 92–94, 275, 292
- monitoring, 140–4, 218–222
- Morrow, K., 172
- Munby, J., 19, 22
- multiple-choice, 45, 47
- multitrait-multimethod analysis (MTMM), 186, 292
- National Curriculum (UK), 252
- native speakers, 97
- needs analysis, 12, 21, 34–5
- ‘negatively skewed’, 93, 292
- Nevo, D., 167, 249, 251
- norm-referencing, 76, 156, 292
- objective items, 51, 292
- objective marking, 106
- objectively marked tests, 46–59
- Oller, J., 3, 19, 45, 56
- one-parameter (Rasch) model, 91
- oral interviews, 62
- ordering tasks, 52
- Oxford-ARELS, 28–29, 124, 143–4, 162
- Oxford, University of, Delegacy of Local Examinations (OUDLES), 34, 208–13
- Palmer, A., 186, 223
- parallel tests, 96–7, 292
- pass marks/rates, 155–9, 163, 165–6
- Pearson product moment correlation, 80, 287
- Peirce, B., 47
- pilot testing, 74, 292
- Pitmans Examinations Institute, 27–8, 208
- placement tests, 11, 292
- point biserial correlation, 84
- Pollitt, A., 235–6
- Popham, W., 158
- ‘positively skewed’, 93, 293
- post-test reports, 197–217
 - analysis of candidates’ scripts, 201–2
 - for the institution itself, 198–202
 - for other audiences, 206–7
 - for teachers, 203–6
 - results of feedback, 200–1
 - results of observations, 199–200
- predictive validity, 177, 180–2, 293
- Preliminary English Test (PET), 30
- pretesting, 73–104, 293
 - less formal ‘pilot testing’, 74
 - main trialling, 75
 - native speakers, using, 97
 - parallel and equivalent test versions, using, 96–7
 - reasons for, 73
 - test analysis, 77–96
- proficiency tests, 12, 293
- progress tests, 12, 274
- QUEST, 91, 285
- questionnaire, 266–273
- range, 92, 276, 293
- rank order correlation, 80, 278–9
- rating scale, 107, 111, 293–4
- rational validity, 171, 294
- reliability, 6, 87, 128, 186–8, 293–4
 - and validity, 186–8
 - Cronbach’s alpha, 87, 101
 - inter-item consistency, 88
 - inter-rater reliability, 129, 290

Index

- intra-rater reliability, 129, 135–6, 290
- Kuder Richardson 20 (KR20), 88–9, 103, 291
- Kuder Richardson 21 (KR21), 88–9, 103, 282–3, 291
- parallel-form reliability, 87, 292
- scripts, 131
- split half reliability index, 88, 280–1
- test-retest reliability, 87
- response validity, 176, 294
- Robinson, P., 22
- Rogers, H., 91
- Royal Society of Arts (RSA), 19
- sampling, 130
 - truncated sample, 182
- SAS, 85, 132, 284
- scaling, 294
- Schools Curriculum and Assessment Authority (SCAA), 252
- Schools Examination and Assessment Council (SEAC), 5, 252
- scores, 148
 - aggregation, 151
 - combining, 153–4
 - correction, 148
 - reported/reporting, 152–3, 163, 294
 - setting pass marks, 155–9, 163, 165–6
 - subtest scores, using to reach a decision, 154–5
 - transformation, 151, 295
- second marking, 130
- setting pass marks, 155–9
- Shohamy, E., 167, 249–51
- Sharon, A., 97
- Sheridan, E., 97
- short-answer questions, 57
- significant difference, 294
- speaking, 116
- Specific Purposes (SP) tests, 12
- specifications
 - for test users, 20
 - for test validators, 16
 - for test writers, 11
 - user, 20
 - validation, 19
- SPSS, 85, 132, 284
- standard deviation, 92, 276–7, 294–5
- standardisation meetings, 112
- standards, 235–59
 - conditional, 240
 - defined, 235
 - primary standards, 240
 - principles, 236
 - secondary standards, 240
 - setting, 111–2, 295
- Stansfield, C., 4
- Stevenson, D., 172
- Stone, M., 91
- subjective items, 59, 295
- subjectively marked tests, 59–62, 86
- subjective marking, 107
- summaries, 61
- Swales, J., 22
- Swaminathan, H., 91
- syllabus, 9, 295
- t-tests, 132, 295
- tests
 - appropriate texts, 43
 - comments from test users, 221–2
 - developing and improving, 218–34
 - item writing, 40
 - listening, of, 117
 - monitoring, 218–23
 - multiple-choice, 45, 47
 - needs analysis, 12, 21, 34
 - revision, of, 223–5
 - speaking, of, 116
 - state of the art in EFL, 255–8
 - vs. exercises, 41
- test editing/moderating
 - committees, 62–4
- test specifications, 9, 11–24, 294

Language test construction and evaluation

- test specifications (*cont.*)
 - criteria, 13
 - items, 13
 - language elements, 13
 - language skills, 13
 - learners, 12
 - methods, 13
 - rubrics, 13
 - sections/papers, 13
 - target language situation, 13
 - tasks, 13
 - taxonomies for, 14
 - test purpose, 11, 21
 - test user specifications, 20
 - text types, 13
- test types, 11–2
 - achievement tests, 12, 286
 - diagnostic tests, 12, 288
 - equivalent tests, 96–7, 288
 - parallel tests, 96–7, 292
 - placement tests, 11, 292
 - proficiency tests, 12, 293
 - progress tests, 12, 293
 - Specific Purposes (SP) tests, 12
- Thorndike, R., 171
- three-parameter (Rasch) model, 91
- transformation of scores, 151, 295
- trialling, 75
- Trinity College, 26–30, 164
- two-parameter (Rasch) model, 91
- University of Cambridge Local Examinations Syndicate (UCLES), 5, 17, 24, 28–9, 34–6, 65–7, 98, 100–2, 123–4, 141, 143, 151–3, 160, 162–3, 190, 191, 229, 232
- user specifications, 20
- validity, 6, 170, 186–8, 296
 - and reliability, 186–8
 - comparison with student's biodata, 185
 - comparison with theory, 183
 - concurrent, 177–80, 286
 - construct, 171, 183–6, 286–7
 - content, 171, 173, 287
 - convergent-divergent validation, 185
 - criterion, 171
 - empirical, 171, 288
 - external, 171, 177–83, 270
 - face, 172–3, 289
 - internal, 171–7, 290
 - internal correlations, 183
 - multitrait-multimethod analysis, 186, 291
 - predictive, 177, 180–2, 293
 - rational, 171, 294
 - response, 176, 294
 - types of, 171
- Vallette, R., 46
- Vanniarajan, S., 174
- variance, 277
- video
 - for training examiners, 124
 - used in testing, 224–5
- Wall, D., 46, 182, 221
- washback, 46
- Waters, A., 22
- Waystage level (of difficulty), 28
- weighting, 149, 162, 296
 - equal, 149
- Weir, C., 3, 35, 99, 101, 191
- West, R., 5, 46
- Windeatt, S., 225
- Wood, R., 186, 189, 192–3
- Wright, B., 91, 223