

## *Index*

- ability groups, 183
- Academic Grammar Test. *See* grammar test
- academic self-efficacy, 300, 340
- academic word list, 187
- academic writing construct, 121
- adjudication of divergent scores, 159
- air traffic control, purpose defined, 85
- air traffic controllers, placement test for, 73
- ALC, 155–156, 163, 172
- ALC Educational Research Institute, 160
- American Council on the Teaching of Foreign Languages (ACTFL)
- Oral Proficiency Interview (OPI), 163
- proficiency levels on, 298
- Applied Linguistics*, 47
- argument-based validation (A-BV) literature review (2000–2018), 46
- argument-based validity
  - advocacy usage, 343
  - appraisal stage, 154
  - appraisal usage, 343
  - complete argument, 8
  - defined, 2
  - limits, 341
  - research trend mapping, 49
  - variants, 33
- Assessing Writing (AW)*, 47, 51
- assessment, 2, 19
  - blueprint, 80
  - changes, 3
  - classroom-based, 10, 12, 261, 327
  - implementation, 80
  - needs for testing aviation English, 75
  - use of term, 326
  - of writing course, 326
- assessment tasks, 78, 82, 90, 237
  - characteristics, 212
  - integrated, 235
- Assessment Use Argument (AUA), 21, 33
  - component layers
    - claims, warrants, backing, 37
  - concepts and practices, 36
  - differs from IUA, 36
- four claims, 36
- measurement claim, 49
- assessments, integrated, 326
- assumptions, 25, 39, 60, 62, 78, 84, 90, 92, 101–103, 130–132, 139, 157, 159–162, 165, 169–170
- evaluation similarity, 230
- in validity argument, 90
- mapping, 145
- scores relationship, 230
- similarity of assignments, 230
- simulating writing tasks, 129
- studies of, 56
- to be tested, 172
- unstated, 63
- Audacity audio recording software, 190
- audience of writing, 219
- authenticity, 2, 21, 31, 92
  - versus generalizability, 38
  - improving, 90
- aviation English, 328, 331, 333
  - testing, 73
  - shortcomings, 74
- Bachman, Lyle F., 1
- backing, 37, 60, 62, 76, 90, 92, 101–102, 129–132, 145, 158–162, 238, 240, 297
  - defined, 24
  - EPT coordinator interview, 283
  - for Kunnan's rebuttals, 155
  - sources of, 56
  - for validity argument, 11
- basic argument units, 23
- British National Corpus (BNC)
  - academic files, 184
- Camtasia screen-capturing software, 117, 188–189, 239, 247–248
- certification testing, 2, 328
- Chinese translation of survey questions, 187
- claims, 6, 37, 87, 93, 98, 149, 155
  - about test use, 6
  - defined, 24

346 *Index*

- Classical Test Theory, 133
- classroom assessment, 53, 326
- CLIL (content & language integrated learning, 261
- coding, 190, 302
- COE (Committee of Examiners) for TOEFL research program, 28
- collocation frequency, 192
- collocational ability, 338
  - measuring, 177
- collocational ability construct, 177–178
- collocational ability test, 10, 60, 177, 184
  - test items, 184, 199
- Computer-Adaptive Test of Productive and Contextualized Academic Vocabulary Breadth in English, 55
- computer-scored responses, 331
- concept of validity, 342
- Conceptual Assessment Framework (CAF), 80
- ConcGram 1.0 software, 184
- confirmationist bias, 154
- consequence implication inference, 12, 57, 62, 102, 132, 161, 207, 240, 253, 286, 296, 327, 340
  - assumptions, 272–273, 297
  - placement test, 296, 340
  - results, 253
  - support, 283
- consequences, 265
  - of test scores, 161, 264, 287
  - test use, 77, 340
- consistency, 63, 97, 100, 162, 265
  - argument, 9
  - of performances, 55, 158
  - of rater severity, 101
  - of ratings, 238–239, 250
  - of results, 166
  - in scoring, 96, 156, 170
- constructs, 6, 31, 56, 61, 63, 131
  - defining, 10, 176–177, 328, 330
- interactionalist definition, 177
- interpretation, 26, 92, 329
  - relationship, 192
  - under-representation, 74, 155
- content analysis, 83
- content course
  - instructors, 205, 213–214, 220
  - writing assignments, 213
    - limitation, 215
- context cues, 76
- corpus linguistics, 178
- Corpus of Contemporary American English (COCA), 185
- correlation analysis, 61
- Cronbach, Lee J. validity definition, 26
- curriculum change prompted by A-BV, 261
- cut-off score for course exemption, 300
- data defined, 24
- design validity phase, 145
- development stage of A-BV, 154
- diagnostic descriptors, 100, 104
- domain analysis, 11, 79, 82, 87, 90, 92, 99, 331
  - of aviation English, 84, 87
- domain definition inference, 8, 29, 36, 38, 55, 58, 63, 77–78, 84, 99, 129, 157, 181, 237, 328, 333
  - claims, 331
  - investigative methods, 337
  - limitations, 90
  - results, 248
  - warrant for, 129
- domain modeling, 79–80
- ECD (evidence-centered design), 8, 76, 79, 92–93
  - domain modeling process, 90
- Educational Measurement*, 20, 26
- Educational Testing Service (ETS), 27
- English as a Second Language (ESL)
  - experiential essays, 219
- instructors
  - interview data, 309
  - interview questions, 319
  - interview responses, 314
- student interview
  - questions, 317
  - responses, 314
- English for Specific Purposes (ESP), 73, 78
- English Language Development Classroom Assessment, 31
- English Placement Test (EPT), 12, 134, 137, 145, 244, 265, 296, 298, 327
  - & Academic Grammar Test results, 275
  - essay test results, 275
  - essay writing
    - correspondence with grammar test, 279
    - interpretation & use, 315, 317
    - performance impact, 312
    - score data collection, 301

- error sources, 154, 156, 161, 337
- ESL (English as a Second Language)
  - assignment sheets (table), 215
- ESL placement tests, 204
- ESL writing
  - assignments & disciplinary course writing, 217
  - course syllabus, 204
  - criteria, 246
  - gap with FYC, 205
  - gaps with mainstream courses, 205
  - instructors, 214, 241
  - revising courses, 206
  - revising tasks/assignments, 230
  - tasks, 215
- ETS Research Report Series*, 47, 51
- evaluation argument, 26
- evaluation framework, 31
- evaluation inference, 9, 29, 58, 63, 98–99, 102, 158, 238, 332, 337
  - assumptions, 130
  - claims, 331
  - results, 249
  - scoring evidence, 55
  - warrant, 129
- evaluation methods, 326
- evidence, integrating, 20
- explanation inference, 10, 38, 56, 61, 63, 101, 131, 159, 180, 191, 239, 327–330, 338
  - assumptions, 181, 332
  - claim, 329–330
  - evidence, 196
  - results, 250
  - summary table, 196
  - warrant, 338
- explanatory sequential design, 103
- external evaluator, 154–155
- extrapolation domain, 76, 120
- extrapolation inference, 11, 28, 38, 57, 61, 63, 101, 155, 160, 180, 207, 230–231, 239, 326, 328–330, 339
  - assumptions, 229
  - claims, 329, 339
  - construct, 131
  - requires authentic tasks, 75
  - results, 251
  - for TOEFL iBT Speaking, 30
  - warrant, 207, 230, 329, 339
- eye-tracking, 117, 262
- false positive decisions, 272–273, 279
- feedback from experts, 331
- First-Year Composition (FYC) courses, 204
- framework, 5, 20, 27–28, 92
  - based on AUA, 77
  - conceptual, 173
  - development, 21
  - for discussing testing issues, 29
  - Messick characteristics, 20
  - for the TOEFL iBT, 77
  - using A-BV, 157, 328
  - for validation, 3, 64
  - for validity research, 4
- fully recursive model, 313
- G Theory, 9, 56, 129, 131, 133, 138
- General Service List (GSL) frequency band, 187
- generalizability, 36, 38, 300
- generalization inference, 9–10, 55, 61, 63, 100, 130, 132, 145, 149, 155–156, 158, 171, 181, 238, 327, 337–338
  - for score consistency, 49
  - not justified, 170
  - rebuttal, 154, 161, 169, 238, 326, 338
  - results, 250
  - for TSST, 172
- genre analysis of assessed writings, 205
- genre family classification, 217–218
- geographical spread of A-BV research, 51
- grade point average (GPA), 297, 311
- grammar, 226–227
- grammar test, 28, 264, 275–276, 289, 295, 327, 340
  - correspondence with EPT essay writing, 279
  - cut score, 279, 286
  - interview questions, 290
  - limitations, 287
  - positive consequences, 287
  - rater training, 278
  - results, 276
  - score use, 280–282
  - scoring rubric, 278
  - test items, 275
- IELTS, 25, 204, 212, 241, 303
  - Academic Writing Test 1 (AWT1), 121
    - rating rubric, 137
    - speaking section, 101
  - in-class testing, 300

348 *Index*

- inclusion criteria for writing test, 134
- inferences, 6, 13, 25, 39, 78, 97, 161, 342
  - in AUA, 36
  - based on test scores, 11
  - basic process of drawing, 19
  - chain of, 98, 328
  - consequence implication. *See* consequence implication inference
  - defined, 24
  - domain definition. *See* domain definition inference
  - evaluation. *See* evaluation inference
  - explanation. *See* explanation inference
  - extrapolation. *See* extrapolation inference
  - framework in coding, 49
  - generalization. *See* generalization inference
  - investigations of, 50
  - multi-part, 25
  - multiple, 7, 24, 62
  - of interpretation and use, 327
  - research methods used, 58
  - in published research, 54
  - types of assumptions, 333
  - use of, 37
  - utilization. *See* utilization inference in validity argument, 36
- International Civil Aviation Organization (ICAO) requirements, 73
- International English Language Testing System. *See* IELTS
- International Language Testing Association Guidelines for Practice*, 3
- international teaching assistants (ITAs), 9, 97, 102–103, 213
- internet literacy, 235
- interpretation/use argument (I/U/A), 9, 76, 80, 149, *See* validity argument
  - advocate's perspective, 205
  - confirmationist approach, 179
  - for grammar test, 265
  - grammar test, 267
  - for the OECT, 98
  - scores on writing tasks, 129
  - structure, 98, 180, 236, 266
  - for TSST, 157, 161
- interviews
  - content course instructors, 220
  - ESL instructors, 214, 219
- individual & focus group, 103, 273
- post-test & follow-up, 188, 190, 247, 252, 338
- protocols, 82
- responses, 340
- semi-structured, 82, 275, 301
- stakeholders, 297, 340
- students, 274, 276, 339
- writing instructors, 276
- item response theory (IRT), 58, 190
- Kane, Michael, 1, 20, 45, 64, 264, 328
  - argument approach, 33
  - confirmationist bias, 154
  - usage of validity argument, 342
- keystroke logging, 262
- Kunnan, Antony J. A-BV ethics-based structure, 154
- LangTest, 165
- Language Assessment Quarterly (LAQ)*, 3, 21, 47
- Language Assessment Research Conference (LARC), 173
- Language Learning*, 47
- Language Testing (LT)*, 3, 21, 47, 51
- Language Testing in Asia (LTiA)*, 47
- lexical frequency, 181
- listening test, 53, 307
- measurement error, 162
- Messick, Samuel, validity defined by, 19–20, 33, 39, 45, 342
- methodologies, range of, 340
- Michigan English Language Assessment Battery (MELAB), 53, 62, 186
- mixed-methods research design, 12, 27, 129, 134, 162, 182, 241, 273, 288, 297–298, 325, 333
- Modern Language Journal*, 47
- Moodle online learning platform, 186, 246
- Motivated Strategies for Learning Questionnaire (MSLQ), 301
- motivation level factors, 310
- multifaceted Rasch measurement (MFRM), 9, 55, 60, 129, 131, 133, 138, 149, 161, 338
- multitrait-multimethod (MTMM) analysis, 61
- multiword linguistic units, 177

- needs analysis, 82, 85, 97
- New English as a Second Language
  - Placement Exam (NESLPE), 154
- nomological network, 178–179, 181, 192, 338
- non-parametric tests, 166
- Ontario Secondary School Literacy Test (OSSLT), 28, 30
- operationalization of validity concept, 6, 342
- Oral Communication Certification Test (OCCT), 328
- Oral English Certification Test (OECT), 8, 97–98, 103, 337
  - explanation claim, 329
  - rating speaking ability, 102
  - scores, 102
  - TEACH section, 104
- Oral Proficiency Interview (OPI), 104
- Palmer, Adrian S., 1
- Papers in Language Testing and Assessment (PLTA)*, 47
- parametric methods, 166
- partial credit model (PCM), 139, 190
- partial-credit scoring, 185
- Patterns of Adaptive Learning Scales (PALS), 300
- Pearson Test of English Academic (PTE Academic), 29
- performance, 7, 9, 101, 155, 158, 160, 265, 329, 331, 333, 337
  - comparisons, 61
  - consistency, 181
  - context influence on, 177
  - in the ESL courses, 297
  - descriptors, 100
  - digital data, 4
  - on OECT, 100
  - sample, 328
  - speaking assessment, 96
  - target, 329
  - in target domain, 57, 132
  - TBLA, 93
  - test versus course, 251
  - tests comparisons, 190
  - on TSST, 157
  - using web sources, 330
  - on VIAET, 78
  - writing test, 129
- performance assessment, 55, 148
  - strong and weak sense, 75
  - technological simulation for, 75
- persons (*p*) variance, 145
- phraseology research, 178
- placement, 62, 129, 272
  - essay test for, 327
- placement decisions, 132, 145, 160, 295, 328, 340
  - & academic achievement, 310
  - acceptance, 306
  - composite scores, 149
  - ESL instructors' perceptions, 305
  - ESL student perceptions, 305
  - frustrations, 303
  - hypothesizing grammar test use, 280
  - impact, 302
  - impact on learning, 307
  - mid-stakes, 265
  - positive attitudes, 304
  - positive impact, 316
  - student attitudes, 306
  - test score use for, 265
- placement testing, 2, 121, 133, 143
- placement tests, 53, 155, 176, 294, 327
  - institutional, 12
  - locally developed, 294
- proficiency levels, 100, 104, 120, 137, 181
- proficiency testing, large-scale, 261
- prompts, 106
  - difficulty, 100
  - for essays, 300
  - limited fidelity of, 120
  - for OECT, 99, 101
  - parallel, 250
  - TSST, 164
  - writing, 246
- proofreading services, 213
- ProQuest Dissertation and Thesis Global database, 47
- psychometric properties, placement test, 295
- putative validity argument, 29
- qualifier (Q) defined, 24
- Qualtrics, 135
- questionnaires
  - analysis of, 162
  - online background, 137
  - post-test & follow-up, 247, 252
  - for TSST, 164
  - student, 213

350 *Index*

- Rasch model analysis, 62
- Rasch-Kappa index, 139
- rater training, 96, 105, 249
  - for essay reading, 138
  - lack of, 149
  - online self-training, 97
  - for writing test, 137
- Rater-Platform (R-PLAT), 8, 97, 100, 102, 104–114, 332
- raters, 254, 331
  - behavior & performance, 259
  - consistency, 130, 159
  - essay, 134, 244
  - negotiations, 55
  - perceptions, 100
  - responses, 103
  - scoring, 61
  - severity, 130, 133, 139, 144, 149–150, 156
- rating, 99, 332, 337
  - criteria, 246
  - essays, 241, 249
  - paper-based, 97
  - practice, 138
  - process of, 98, 102
  - web-based, 98, 104
- rating scales, 53, 139, 148, 158
  - analytic, 147
- reading comprehension, 121, 176, 179
- reading-to-write task, 148
- rebuttals, 27–28, 155
  - by external evaluators, 154
  - by external examiner, 155
  - data, 170
  - defined, 24
  - for G theory, 10
  - grammar test objectivity, 287
  - by Kunnan, 155
  - response to, 172
- regression to the mean, 157
- reliability, 2, 6, 10, 21, 26, 31, 60, 101, 236, 338
  - aspects of, 156
  - claims about, 36
  - comparing score, 56
  - Cronbach's alpha used, 61
  - estimates, 133
  - G-theory-based analysis, 61
  - inter-rater, 9, 149
  - of raters, 159
  - test-retest, 173
- research design, 162
  - document analysis, 337
  - evidence centered, 336
  - for explanation inference, 182
  - interviews, 337
  - mixed methods, 212, 333, 341
  - mono-methods, 341
  - univariate, 138
- research questions of studies in this volume, 46, 78, 132, 162, 182, 212, 242, 272, 280, 297
- rhetorical functions, 223, 225, 231
- RMIT English Language Test for Aviation (RELTa), 75
- rubric, 96, 106, 132, 337
  - analytic rating, 135, 137, 246
  - comparison, 251
  - EPT scoring, 302
  - essay scoring, 238, 298, 300
  - essay writing, 307, 339
  - improvement, 249
  - for placement test, 148
  - for reading-to-write task, 137
  - revising, 250
  - scoring, 55, 60, 176, 265
  - systematic development, 249
- sample size, 239
- sampling, 163, 171, 299
- score correlations, 167
- score dependability, 139, 148
- score descriptors
  - improving, 253
- score interpretation, 2, 7, 9, 12, 21, 30–31, 96, 98, 120, 132, 149, 161, 191, 236, 288, 325, 328
- caution, 171
- of collocational test, 177
- specifics, 343
- TSST, 161
- validity, 283
- score report, 164, 340
  - feedback, 170
  - online, 164
- score use, 240, 264, 326
  - grammar test, 278
  - intended, 264
  - validity of, 264
- score variance, 130, 147
  - components, 143
  - source, 148

- total, 140
- in TSST, 169
- scores. *See* test scores
  - composite, 138
  - observed linked to expected, 130
- scoring methods for collocational test, 185
- screen-capturing technology, 188, 194, 332
- Second Life, 78, 80, 83–84, 90, 331
- self-efficacy, 313
- simulation of speaking tasks, 99
- sociocognitive approach, 20
- source material use, 218, 230, 251
- source-based writing, 235, 240, 249
- speaking proficiency, 101, 154–157, 163, 165
  - changes in, 160, 164
  - levels, 100
  - test scores correlation, 160
- speaking test, 10, 61, 104, 326, 329
- stakeholders, 264, 277, 340
  - benefits of grammar test, 286
  - consequences to, 265
  - EPT coordinator, 272, 274, 276–278
  - EPT test takers, 274
  - ESL instructors, 294, 299
  - ESL students, 294, 298
  - interview questions, 299
  - judgments, 339
  - perceptions, 272, 283
  - perceptions of EPT decisions, 303
  - potential impact on, 288
  - test takers, 272, 277
  - writing instructors, 272, 274, 277
- Standard Speaking Test (SST), 163
- Standards for Educational and Psychological Testing*, 2, 343
  - validity definition, 5
- strategic competence, 76, 178
- strategies, 182, 332
  - cognitive & metacognitive, 190, 194
  - test-taking, 156
- structural equation model (SEM), 302, 340
- success rate in ESL courses, 309
- surveys
  - student, 214, 217, 225
  - test reflection, 187, 194
- System*, 47
- systematic review method, 46
- target construct, 76
- target domain, 157, 328, 339
  - defined, 328
  - evaluated, 160
  - of web-source-based writing, 237, 337
- target language use (TLU), 76, 96, 99
  - simulations, 92
  - situations, 86
- target performance, 329
- target scores, 101
- target tasks, 84, 87, 89
- task
  - instructional, 339
  - interactive assessment, 76
  - test compared to instructional, 251
- task design, 79, 337
- task difficulty, 121, 150, 156
- task model, 87, 92
  - rubric types, 87
  - variable elements, 87
  - variants, 87
- task sequences, 86
- task shells, 79, 84, 87, 92
- task situations, 83
- task specification, 87, 104, 338
- task types, 248
- Task-Based Language Assessment (TBLA), 75–76, 92
- technologies, process-tracing, 262
- technology use, 4, 97, 327, 330, 332
  - adaptive speech, 90
  - testing grammar, 327
  - web-based, 328
- Telephone Standard Speaking Test (TSST), 10, 154–155, 326, 329, 331
  - holistic scale, 163
  - procedures, 163
- TESOL Quarterly*, 47
- test administration, 156
- test delivery, 331
- test designs, 148
- test developer role, 342
- test development, 5, 329, 342
  - piloting phase, 340
  - process of, 326
  - stages of, 325
- test evaluator role, 342
- test factors, 295
- test fairness, 155
- test familiarity, 170
- test forms, 158
  - correlation, 171
  - differences, 162, 167
  - relationships, 168

352 *Index*

- test interpretation, 6
- test level analysis, 162
- test of collocational ability, 331
- Test of English as a Foreign Language. *See* TOEFL & TOEFL iBT
- Test of English as a Foreign Language (TOEFL), 53
- paper-based, 241
- Test of English for International Communication (TOEIC), 53
- Listening & Reading Test, 160
- Test of English Language Level for Controllers and Pilots (TELLCAP), 75
- Test of English Language Proficiency for Aeronautical Communication (ELPAC), 75
- test performance, 251
  - ESL writing course, 236
  - test score dependability, 143
  - test score differences, 168–169
  - test score use, 11, 57, 62, 340
  - test scores, 7, 162, 337
    - consistency, 9, 101
    - decision-making consequences, 295
    - fluctuation, 157, 168
    - in-class, 301
    - individual, 164
    - observed to expected, 100, 158, 161, 238
  - test specifications, 159, 250
  - test tasks, 158–159, 164, 265, 275, 333, 337, 339
    - for collocational ability, 184
    - development, 331
    - parallel, 338
    - parallel versions, 260
    - prototype, 336
    - prototypes of, 89
    - rating, 337
    - specifications, 330
    - that miss target construct, 74
    - on TSST, 157
    - TSST, 163
  - test topics
    - for essay writing, 135
  - test use, 1, 5, 11, 21, 25, 30, 96, 98, 102, 120, 132, 161, 173, 241, 265, 325, 333, 342
    - data, 12
    - in classroom, 326
    - interpretation, 172
    - of VIAET, 77
    - specifics, 343
- test usefulness, 20
- test users, 4, 7
- test-taking strategies, 170
- text borrowing, 120
- text comprehension
  - graphic included, 121
- Thompson, John, 288
- TOEFL, 204, 212
- TOEFL iBT, 2, 27, 167, 241
  - extrapolation inference, 155
  - fairness argument, 29
  - interpretive argument framework, 129
  - section scores, 56
  - speaking section, 28
  - validity argument, 28, 30, 63
  - writing section, 28, 56, 63
- Toulmin, Stephen, argument concept of, 23, 37
- unitary validity, 25
- utilization inference, 12, 27, 36, 57, 62, 102, 132, 160, 207, 240, 327, 339
  - assumptions, 273, 283
  - grammar test score use, 266
  - results, 252
  - support, 283
  - warrant, 340
- validation
  - framework, 13, 28, 33
  - process, 341
  - research, 1, 6, 20
  - usages of, 343
- validity
  - unitary but multifaceted, 20
  - yes or no decision about, 341
- validity argument, 98
  - as conceptual reference, 29
  - chain of parts, 25
  - diagram, 26
  - differs from AUA, 36
  - functions, 31
  - inferences, 36–37
  - limit to, 170
  - multifaceted nature, 25
  - parts, 24
  - usefulness of, 260
  - uses, 29
- variability source, 96
- verb-noun collocations, 177–179
- video coding scheme, 248

- Virtual Interactive Aviation English Test (VIAET), 73, 329
  - test interpretation & use, 76
- virtual simulation, 76, 78, 83–84, 90
- vocabulary
  - for aviation English, 85
  - strength of, 179
- vocabulary test, 53, 58, 177, 179, 186
- warrant, 25–27, 37, 39, 78, 90, 98–102, 131–132, 155, 157
  - defined, 24
  - studies of, 56
  - unstated, 63
- washback effects, 132, 241, 253, 260
- web-researching-to-write ability, 330, 332
- word pairs (concgrams), 184
- writing assessment, 31, 97
  - ESL versus content courses, 205
- writing assignments, 226
  - comparison ESL & course, 214
  - in content course, 222
- differences, 227–228
- ESL versus content courses, 217, 222
- essay types, 206
- length, 224
- purposes, 220
- similarities, 226
- source-based, 260
- writing proficiency, 307
- writing tasks, 135
  - content course, 214
  - graphic-prompt, 120–121, 148
  - independent, 147
  - integrated, 147
  - paper- versus computer-based, 121
  - source-based, 120
  - topics, 218
- writing test, 9, 11, 56, 60–61, 176, 246, 265, 330
  - computer-based, 135
  - graphic-prompt, 149
  - integrated, 235, 239, 262, 332, 336
  - procedures, 247
  - single prompt, 12