

# INDEX

## A

- Accommodating special needs
  - ADA requirements for, 387–388
  - in context of test purpose, 388–389
  - reasonable accommodation in context of, 388
  - for testing, assessment, and evaluation for disabled candidates, 390–394
  - undue hardship in context of, 388
- ADA (Americans with Disabilities Act)
  - on accommodating special needs, 387–394
  - blanket exclusions prohibited by, 378–380
  - bringing legal claim under, 368–370
  - disabilities covered under the, 369–370, 390–394
- Administration. *See* Test administration
- Adverse impact
  - as defined by *Uniform Guidelines*, 380–381
  - description of, 372–373
  - practice statements on, 381–383
  - record-keeping on, 387
  - sample summary of figures on, 381*t*
  - Uniform Guidelines*’ “bottom line” standard related to, 383–385
  - validity study assessment of, 400
- Affirmative action, 385–386
- Agreement ( $p_a$ ) coefficient
  - calculating inter-rater reliability using, 325*fig*–335
  - comparison of phi, kappa, and, 313–314
  - description of, 306–307
  - how high should it be?, 308
  - matrix for determining  $p_{chance}$  and, 309*fig*, 310*fig*
  - practice on calculating, 307
- Albemarle Paper Company v. Moody*, 375–376, 386
- Alexander, R. A., 386
- Algina, J., 296
- American Psychological Association, 394
- Analysis level (Bloom’s taxonomy), 93
- Analyze job content. *See* Job content analysis
- Anderson, L. W., 95
- Angoff technique
  - Angoff ratings of items, 263*t*–264
  - for establishing cut-off scores, 50, 266, 272–278
- Animated test item, 129
- Application level (Bloom’s taxonomy), 93
- Assessment
  - used for compliance, 58–62
  - definition of, 15–16
  - of disabled candidates, 390–394
  - practice statement on, 16
- Assessment Centers, 16
- Assessment stakes
  - costs of, 43
  - factors determining, 41–43
  - low-, medium-, and high-, 40–41
  - Question mark White Paper on, 43

The Association of Test Publishers, 69, 262  
 Attendance—Level E, 112–113*t*  
 Attitudes hierarchy, 97  
 Auerbach, M. A., 273  
 Ausman, T., 231  
 Australian Securities and Investment Commission (ASIC), 59  
 “Authentic” documentation, 415

**B**

BAR Scale (Behaviorally Anchored Rating Scale), 188–189*fig*  
 Barrett, A., 288  
 Barrett, G. V., 386  
 Barrett, R. S., 102–103  
 Barritt, C., 264  
*Bates vs. UPS*, 379–380  
 Bausell, R. B., 200  
 Belenky, A. H., 77  
 Bellezza, E. S., 237  
 Bellezza, S. F., 237  
 Bemis, S. E., 77, 78  
 Berk, R. A., 255, 265  
 Berry, T., 186  
 Blair, D., 115  
 Blanket exclusions, 378–380  
 Bloom, B., 91, 122, 123  
 Bloom’s taxonomy  
   essay items and level of, 148  
   fill-in items and level of, 147  
   matching items and level of, 132  
   multiple-choice times and level of, 133–135  
   for objectives, 123, 124*t*–127*t*  
   original version of, 91–94  
   practice statements on, 128  
   revised version of, 95–96*t*  
   short answer items and level of, 148  
   true/false items and level of, 131  
   used to validate a hierarchy, 94*fig*–95*fig*  
 Bookmark, 266  
 Borderline cut-off decisions, 282–285  
 Brennan, R. L., 295

Bristol-Myers Squibb, 180  
 Browning, A. H., 63  
 Bugbee, A. C., Jr., 63  
 Burton, S. J., 137, 159

**C**

Callahan, D., 232, 233, 234  
 Campbell, C. P., 190  
 Cascio, W. F., 370, 386  
 Case, S. M., 136, 159  
 Caveon Test Security, 69, 237, 240  
 Central tendency error, 321–322  
*Certification: A NOCA Handbook* (Browning, Bugbee, & Mullins), 63  
 Certification  
   definition of, 8  
   industry evolution of, 10  
   training purpose of testing for, 31  
 Certification programs, 7–9, 10  
 “The Certification Suite:” (Coscarelli, Robins, Shrock, & Herbst)  
   certification levels in the, 110–112  
   described, 109–110  
   how to use the, 113–116  
   Level A—Real World, 110–111, 113*t*, 117  
   Level B—High-Fidelity Simulation, 111, 113*t*, 117–118, 156  
   Level C—Scenarios, 111, 113*t*, 118, 156–158  
   Level D—Memorization, 112, 118, 150, 152–156  
   practice statements on, 155–158  
   quasi-certification levels in, 112–113  
   selecting the certification level, 115*fig*  
   summary of, 113*t*  
 Certified Production and Inventory Management (CPIM), 7  
 Chartered Life Underwriter (CLU), 7  
 Chauncey Group, 7  
 Cheating  
   consequences of, 231–232  
   protecting test integrity from, 231, 233–234

- using statistical methods to detect, 237–240
- test security measures to combat, 234–240
- understanding motivations for, 232–233
- The Cheating Culture: Why More Americans Are Doing Wrong to Get Ahead* (Callahan), 232, 234
- Cheating on Tests: How to Do It, Detect It, and Prevent It* (Cizek), 232, 234
- Checklists, 190fig–192fig
- City of St. Louis, Fire Fighters Institute for Racial Equality v.*, 102
- Civil Rights Act (1964), 368, 370, 374
- Cizek, G., 232, 234
- Closed-ended questions, 129
- Code of Federal Regulations Title 29, 371
- Coefficients
  - agreement ( $p_r$ ), 306–308, 309fig, 310fig, 313–314
  - comparing agreement, kappa, and phi, 313–314
  - deriving correlation, 292
  - Fisher's Z, 338–339t, 340
  - internal consistency of, 294–295
  - kappa ( $\kappa$ ), 308–314
  - Livingston's Coefficient kappa ( $\kappa^2$ ), 296, 297, 298–299
  - $p_{\text{chance}}$ , 309fig–310fig
  - phi ( $j$ ), 204, 255fig, 297–298, 302–306, 308, 313–314
  - single-test administration techniques for reliability, 293, 294–299
  - squared-error loss, 296
  - two-test administration techniques for reliability, 293, 294–313
  - understanding the reliability, 313–314
  - See also* Correlations
- Cognitive items
  - cookbook for working with SMEs on, 172–174
  - definition of, 121–122
  - description of, 46–47
  - documentation on creating, 65
  - guidelines for writing test, 158–170
  - intensional vs. extensional, 150–152
  - matching jobs to, 149–155
  - show vs. tell metaphor on writing, 152–155
  - test reliability and role of, 314–315
  - types of test, 129–149
  - See also* Items; Test items
- Cognitive strategies hierarchy, 97
- Cognitive tests
  - creating items for, 46–47
  - determining reliability of, 50–51
  - documenting reliability of, 67
  - guidelines for writing items for, 158–170
  - reliability of, 289–317
  - single-test administration techniques for, 293, 294–299
  - two-test administration techniques used for, 294, 299–313
  - See also* Tests
- Cohen, J., 308
- Collusion, 238
- Communication
  - language issues of documentation, 409–413
  - language issues of writing test items, 158
  - legal evidence of electronic, 409
  - See also* Documentation
- Compliance
  - effects of, 59–60
  - growing role of, 58–59
  - Uniform Guidelines* on documenting, 402–403
- Compliance assessment
  - compliance program role of, 60
  - described, 58
  - managing, 61–62
  - types and purpose of, 60–61
- Component Design Theory, 98–100fig
- Comprehension level (Bloom's taxonomy), 93
- Computer-based test items, 129–130

- Computer-based testing  
 assumptions behind, 176–177  
 business industry preference for, 177–178  
 cost-saving through, 178  
 description and advantages of, 176  
 medical sales training use of, 178–180  
 power of, 180
- Computerized Adaptive Testing (CAT), 180–181
- Computerized item banks, 262–264
- Concurrent validity  
 definition of, 22, 23–24, 202  
 determining, 202–206  
 example of, 204*t*  
 phi ( $\phi$ ) table for, 205*fig*  
 practicing, 206*t*–207*fig*  
*Connecticut v. Teal*, 384
- Construct validity, 185–186
- Content Validation Form, 199*fig*
- Content Validation Results Form, 199
- Content validity  
 comparing face validity and, 196–197  
 Content Validity Index (CVI), 200–202  
 as CRTD cornerstone, 23  
 description of, 22, 23, 197  
 determining, 197–200  
 documentation establishing items and instrument, 65–66  
 established for items and instruments, 47  
 establishing objectives, 46  
 importance of establishing, 196  
 summary comment about, 209  
*See also* Validity
- Content Validity Index (CVI), 200–202
- Content Validity Index Scales, 200, 201*fig*
- Content validity of objectives  
 documenting, 65  
 overview of process, 105–106
- Contrasting groups, 50, 266
- Contrasting Groups method, 278–281
- Correction-for-guessing  
 formula for, 286*fig*–287  
 problems with, 285–286
- Correlations  
 definition of, 290–293  
 graphic illustrations of, 291*fig*  
*See also* Coefficients
- Coscarelli, W., 43, 44, 53, 110, 288, 294, 296, 298
- Criterion-Referenced Testing Development (CRTD)  
 content validity cornerstone of, 22–23, 46–47, 65–66, 196–202  
 cookbook for working with SMEs on, 172–174  
 documentation of, 44–73, 68*fig*  
 inaccessibility of technology for, 4  
 overview of issues related to, 11–12  
 pitfalls related to, 3–5  
 practice statements on, 29–30, 31–32  
*See also specific issues*
- Criterion-referenced tests (CRTs)  
 balancing employment discrimination laws with, 376–378  
 checklist for ensuring legally defensible, 416–419  
 comparing NRTs and, 25–30  
 example of test-retest data for, 301*t*  
 importance of writing items above the memorization level, 105  
 internal consistency measures of reliability problem of, 295–296  
 interpretation of, 28*fig*–30  
 item analysis for, 251–255*fig*  
 objectives-based construction of, 34  
 outcomes of, 266*fig*–267  
 practice phi calculating using data from, 305–306  
 reporting scores of, 52–53, 67–68, 258–361  
 reporting scores of NRT vs., 358

- single-test administration of, 293–299*t*, 305–306
  - systematic process of, 39–40, 43–53
  - two-test administration for, 19, 51–52, 296–306
  - upper/lower index for a, 253*fig*–254
  - See also specific issues; Tests*
  - Critical elements, 193
  - Cronbach, L. J., 296
  - Cronbach's *alpha* index, 294
  - Cross-cultural issues, 158
  - CRT process
    - analyzing job content, 44, 46
    - conducting initial test pilot, 47–48
    - creating items, 46–47
    - creating parallel forms of item banks, 49
    - described, 43–44
    - determining reliability, 50–52
    - diagram of designing, 45*fig*
    - establishing content validity, 47
    - establishing content validity of objectives, 46
    - establishing cut-off scores, 49–50
    - instructional design process relationship to, 39–40
    - perform item analysis, 48–49
    - planning documentation of, 44–73
    - reporting scores, 52–53
  - Cut-off scores
    - borderline decisions, 282–285
    - calculating combined course, 360*t*
    - calculating overall course, 360*t*
    - described, 49–50
    - documentation of, 67
    - human judgment in setting, 267–269
    - misclassification of, 267–268, 284–285
    - performance data used in determining, 268–269
    - problem of saltatory, 287–288
    - problems with correction-for-guessing, 285–287
    - reversibility of, 268
    - squared-error loss and, 296
    - stakeholder opinions on, 268
    - three procedures for setting, 269–281
    - See also* Master/mastery decisions
  - Cut-off scores procedures
    - Conjectural Methods (Angoff Method), 50, 266, 272–278
    - Contrasting Groups method, 278–281
    - Informed Judgment, 50, 270–272
    - overview of, 269
    - practice statements on, 271–272, 276–278, 280–281
    - substitutability issue of, 269–270
- D**
- DACUM (Developing a Curriculum), 79–81*t*
  - Data Forensics, 237–240
  - Decision-Making Style Inventory, 15
  - Department of Transportation (DOT), 59, 379
  - Desmedt, J., 192–193, 194
  - Diagnostic tests, 31
  - Difficulty index, 48, 215, 248–249
  - Direct Consensus, 266
  - Disabilities
    - accommodating special needs test-takers with, 387–394
    - ADA definition of, 369–370, 390–394
    - ADA prohibition of blanket exclusions based on, 378–380
    - discrimination claims regarding, 368–370
    - hearing impairments, 390, 393
    - learning, 390, 391–392
    - medical issues, 390, 393–394
    - mobility impairments, 390, 393
    - visual impairments, 390, 392–393
  - Discrimination issues
    - ADA (American with Disabilities Act) on, 368–370
    - adverse impact as, 372–373, 380–383, 387, 400

## Discrimination issues (continued)

- avoiding blanket exclusions, 378–380
- balancing CRTs with employment laws on, 376–378
- “four-fifths” rule, 375
- Title VII (Civil Rights Act of 1964) on, 368, 370, 374
- See also* EEOC (Equal Employment Opportunity Commission); Legal citations
- Distractor patterns
  - described, 48, 215, 249–250
  - guidelines for writing for, 160–161
  - organizing the, 161–162
- Diversity issues. *See* Minority populations
- Documentation
  - of adverse impact and job-relatedness of tests, 387
  - “authentication” standard of, 415
  - compliance, 402–403, 412
  - used to defend against legal claims, 401–402
  - definition of, 401
  - effective word management in, 409–413
  - functions of, 402–407
  - illustration of CRTD process and, 68*fig*
  - keeping complete, 414–415
  - legally defensible, 400–415
  - list of questions/issues to address in, 64–68
  - planning, 44
  - reasons for, 57–58
  - retention policies and protocols for, 407–409, 413–414
  - standards for, 63–64
  - test security plan, 68–73
  - See also* Communication
- Documentation word management
  - avoiding inflammatory/off-the-cuff commentary, 412–413
  - described, 409–410
  - using objective terms, 412

- principles of, 410–411
- writing with accuracy and precision, 411
- Donath, C., 266
- Drag and Drop test item, 129
- Drake Prometric, 8
- Duke Power Company, Griggs v.*, 374–375

## E

- Ebel’s method, 266
- Educational Testing Service, 7, 159
- EEOC (Equal Employment Opportunity Commission), 376, 385, 387, 398, 409
- See also* Discrimination issues; *The Uniform Guidelines on Employee Selection Procedures*
- Employment selection laws
  - affirmative action allowed by, 385–386
  - balancing CRTs with discrimination and, 376–378
  - checklist for CRTs legally defensible under, 416–419
  - described, 368
  - legal claims made under, 368–370
  - See also* Legislation; *The Uniform Guidelines on Employee Selection Procedures*
- Entry tests, 30–31
- Environmental Protection Agency (EPA), 59
- Equal Employment Opportunity Commission (EEOC), 376, 385, 387, 398, 409
- Equivalence reliability
  - calculating, 299–300
  - described, 19
  - determining, 51
- Equivalency tests, 31
- Error of standards, 320
- Errors
  - misclassification, 267–268, 284–285
  - rating, 320–322
  - standard error of measurement, 282–284

Essay test items  
 description of, 148  
 guidelines for writing, 163–165  
 Ethnic/racial groups. *See* Minority populations  
 Evaluation  
 definition of, 16  
 of disabled candidates, 390–394  
 Kirkpatrick's levels of, 5–7  
 practice statement on, 16  
 Evaluation level (Bloom's taxonomy), 94  
 Extensional items, 150–152  
 Eyres, P. S., 57–58, 362, 367, 389, 394

## F

Fabrey, L., 8, 9, 31  
 Face validity  
 comparing content validity and, 196–197  
 definition of, 22, 23  
 performance test, 185–186  
 Fairness issue, 399–400  
 Feldt, L. S., 295  
 Fidelity performance test, 185  
 Fill-in test items  
 description of, 147  
 guidelines for writing, 162, 163–165  
 Financial Services Authority (FSA) [UK], 59  
*Fire Fighters Institute for Racial Equality v. City of St. Louis*, 102  
 Fisher's Z coefficient, 338–339*t*, 340  
 Flesch Readability Index, 219  
 Flowers, C. P., 256  
 Fluency, 193  
 Food and Drug Administration (FDA), 59  
 Foster, D., 240  
 "Four-fifths" rule, 375  
 Francis, R. W., 359, 360  
 Frequency distributions, 25–28

## G

Gael, S., 78  
 Gagné, R. M., 96, 122

Gagné's learned capabilities  
 listed, 96–97  
 used to validate a hierarchy, 97–98  
 Garbage In/Garbage Out (GIGO), 257  
 Godwin, J., 43  
 Graduate Record Exam (GRE), 47  
*Griggs v. Duke Power Company*, 374–375  
 Gronlund, N. E., 183, 187

## H

Hale, J., 31, 77  
 Halo error, 321  
 Hambleton, R. K., 255, 296  
 Hand scoring, 175  
*Handbook I: Cognitive Domain* (Bloom & Krathwohl), 123  
*Handbook of Task Analysis Procedures* (Jonassen, Hannum, & Tessmer), 78  
 Hannum, W. H., 78  
 hatcher, T. G., 190  
 Health Insurance Portability and Accountability Act (HIPAA), 59  
 Hearing impairment  
*Bates v. UPS* on discrimination of, 379–380  
 special accommodations related to, 390, 393  
 Herbst, P., 110  
 Hierarchies  
 analysis of tasks using, 87*fig*–88*fig*  
 ensuring validity of job task, 102–103  
 Gagné's learned capabilities to validate, 97–98  
 matching type of test to, 88–91  
 Hierarchy validation  
 Bloom's revised taxonomy for, 95–96*t*  
 Bloom's taxonomy for, 94*fig*–95*fig*  
 Component Design Theory for, 98–100*fig*  
 data-based methods used for, 100–102*fig*

Hierarchy validation (continued)  
 using learning task analysis for, 91  
 using posttest scores for, 101*fig*–  
 102*fig*  
*See also* Validity

High stakes tests, described, 10

High-Fidelity—Level B, 111, 113*t*,  
 117–118

“Holiday meal effect,” 227

Honesty/integrity issues  
 consequences of cheating, 231–232  
 protecting test integrity, 231, 233–  
 234  
 test security, 234–240  
 understanding motivation to  
 cheat, 232–233

Hotspot test item, 130

Hunt, M., 150, 151

Hurtz, G. M., 273

## I

Impara, J. C., 266, 269

Independent testing centers, 227–231

Index  $S_c$ , 296, 297–298

Inflation factors, 17–18

Informed Judgment Method, 50,  
 270–272

Initial test pilot. *See* Pilot tests

Instructional design process, 39–40

Instruments  
 documentation on content validity  
 of, 65–66  
 establishing content validity of, 47  
 rating, 47, 65

Integrity. *See* Honesty/integrity  
 issues

Intellectual skills hierarchy, 96–97,  
 98*fig*

Intensional items, 150–152

Inter-rater reliability  
 calculating/interpreting kappa  $\kappa$   
 for, 323–335  
 calculating/interpreting phi ( $j$ ),  
 335*fig*–344  
 description of, 19, 290, 322–323

example of performance test data  
 for, 324*t*

invalid, 349–353

practice statements on, 330–335

Internal consistency, 294–296

International Conference on Har-  
 monization of Technical Require-  
 ments (ICH), 59

“Is the Test Content-Valid: Or, Who  
 Killed Cock Robin?” (Barrett),  
 102–103

Item analysis  
 choosing techniques for, 255–257  
 for criterion-referenced tests, 251–  
 255*fig*  
 described, 48–49  
 documentation on, 66  
 phi ( $j$ ), 255*fig*  
 practice statement on, 252–253  
 upper-lower index, 253*fig*–254

Item banks  
 Angoff ratings for, 263*t*–264  
 computerized, 262–264  
 creating, 49  
 documentation on, 66–67

Item response theory (IRT), 256–257

Item statistics  
 choosing techniques for, 255–257  
 difficulty index, 48, 215, 248–249  
 distractor patterns, 48, 160–162,  
 215, 249–250  
 Garbage In/Garbage Out (GIGO)  
 principle of, 257  
 p-value, 249  
 point-biserial correlation, 48–49,  
 215, 250–251

Items  
 computerized item banks, 262–264  
 description of, 46–47  
 documentation on content validity  
 of, 65–66  
 establishing content validity of, 47  
 performing analysis of, 48–49  
 role of objectives in writing, 106–109  
*See also* Cognitive items

## J

- Jaeger method, 266
- Job analysis
- DACUM approach to, 79–81*t*
  - described, 44, 46
  - documentation on, 65
  - Kirkland v. Department of Correctional Services* on inadequate, 77
  - standard task analysis form for, 82*t*–86*t*
  - standards for, 75–77
  - summary of process of, 78–79
  - validity study and role of, 398
- The Job Analysis Handbook for Business, Industry, and Government* (Gael), 78
- Job analysis hierarchy
- Bloom's original taxonomy, 91–95*fig*
  - Bloom's revised taxonomy, 95–96
  - data-based methods for validation of, 100–102*t*
  - ensuring validity of the task lists in, 102–103
  - examples of using, 87*fig*–91
  - Gagné's learned capabilities, 96–98
  - Merrill's component design theory, 98–100*fig*
- Job performance-cognitive items matching, 149–155
- Job Task Analysis Form, 82*t*–86*t*
- Job tasks
- converting into objectives, 116–118
  - example of analysis form for, 82*t*–86*t*
  - hierarchical analysis of, 87*fig*–88*fig*
- Jonassen, D. H., 78
- Josephson Institute, 232, 234
- Judges
- Angoff method and role of, 273–278
  - Contrasting Groups method and role of, 278–281
  - high percent of agreement leading to negative, 353–356
  - inter-rater reliability of, 19, 290, 322–344

- invalid decisions by, 349–353
- organizational response to invalid decisions by, 352–353
- procedures for training, 347–349
- types of errors made by, 320–322
- See also* Master/mastery decisions

## K

- Kane, M., 265
- Kappa ( $\kappa$ ) coefficient
- calculating the, 309*fig*–310*fig*, 324
  - comparison of phi, agreement, and, 313–314
  - description of, 308–309
  - how high should it be?, 311
  - inter-rater reliability using, 323–335
  - practice calculating, 311–313
- Kelley, T. L., 253
- Kincaid Readability Index, 219–220*fig*
- Kirkland v. Department of Correctional Services*, 77
- Kirkpatrick, P., 5, 6
- Kirkpatrick's levels of evaluation, 5–7
- Kleeman, J., 288
- Knowledge level (Bloom's taxonomy), 92
- Knowledge tests, 184–186
- Krathwohl, D., 95, 123
- Kuder-Richardson 20 (K-R 20) index, 294
- Kuder-Richardson 21 (K-R 21) index, 294

## L

- Lack of variance, 295
- Language issues
- avoiding inflammatory documentation language, 412–413
  - effective word management in documentation, 409–411
  - objective terms used in documentation, 412
  - when writing test items, 158

Lathrop, R., 284, 344–345, 346  
 Learning disabilities, 390, 391–392  
 Learning theories, 186  
 Learning types, 99  
 Lee, E., 255, 294  
 Legal citations  
   *Albemarle Paper Company v. Moody*, 375, 386  
   *Bates vs. UPS*, 379–380  
   *Connecticut v. Teal*, 384  
   *Fire Fighters Institute for Racial Equality v. City of St. Louis*, 102  
   *Griggs v. Duke Power Company*, 374–375  
   *Kirkland v. Department of Correctional Services*, 77  
   *Myart v. Motorola*, 374  
   *Thigpen v. UPS*, 412–413  
   *Walls v. Mississippi State Department of Public Welfare*, 375  
   *Zubulake v. UBS Warburg*, 409  
   *See also* Discrimination issues  
*The Legal Handbook for Trainers, Speakers, and Consultants* (Eyes), 63  
 Legal issues  
   accommodating special needs of test-takers, 387–394  
   adverse impact as, 372–373, 380–383, 387, 400  
   affirmative action, 385–386  
   balancing CRTs with employment discrimination laws, 376–378  
   “bottom line,” 383–385  
   cautions against blanket exclusions, 378–380  
   checklist for legally defensible CRTs, 416–419  
   employment selection laws, 368  
   legally defensible documentation, 400–415  
   record-keeping of adverse impact/job-relatedness of tests, 387  
   regarding the archiving of tests, 362  
   validation criteria related to, 394–400  
   who may bring a claim, 368–370

*See also The Uniform Guidelines on Employee Selection Procedures*  
 Legislation  
   ADA (Americans with Disabilities Act), 368–370, 378–380, 387–394  
   Health Insurance Portability and Accountability Act (HIPAA), 59  
   No Child Left Behind (2001), 10  
   Occupational Health and Safety Act (OHSA) [South Africa], 59  
   Patriot Act, 59  
   Sarbanes-Oxley Act, 59  
   Title VII (Civil Rights Act of 1964), 368, 370, 374  
   *See also* Employment selection laws  
 Leniency error, 322  
 Leptokurtic distribution, 244  
 Leptokurtic (smaller) standard deviation, 244, 245<sup>fig</sup>  
 Level A—Real World, 110–111, 113<sup>t</sup>, 117  
 Level B—High-Fidelity Simulation, 111, 113<sup>t</sup>, 117–118, 156  
 Level C—Scenarios, 111, 113<sup>t</sup>, 118, 156–158  
 Level D—Memorization  
   described, 112  
   practice statements on, 155–158  
   quasi-certification, 118  
   show vs. tell metaphor applied to, 152–155  
   writing items above the, 150  
 Level E—Attendance, 112–113<sup>t</sup>  
 Lewis, D., 264  
 Lexile Framework, 220  
 Licensure, 8–9  
 Livingston, S. A., 265, 272  
 Livingston’s Coefficient kappa ( $\kappa^2$ ), 296, 297, 298–299  
 Logic error, 321  
 Lynn, M. R., 200

## M

Master/mastery decisions  
   described, 8  
   determining standard for, 265–266

- established by legitimate score reporting, 52–53
  - inter-rater reliability of decisions on, 19, 290, 322–344
  - invalid, 349–353
  - negative phi coefficient and high percent of agreement on, 353–356
  - organizational response to invalid, 352–353
  - repeated performance/consecutive success criteria for, 344–347
  - See also* Cut-off scores; Judges
  - Mastery curve, 247*fig*–248
  - Matching test items
    - description of, 132
    - guidelines for writing, 159–160, 163–165
  - Maynes, D., 237, 238
  - McGinty, D., 256
  - McKie, D., 296
  - Measurement
    - definition of, 15
    - practice statement on, 16
    - See also specific measurements*
  - Medical College Aptitude Test (MCAT), 28
  - Medical issues, 390, 393–394
  - Memorization—Level D. *See* Level D—Memorization
  - Merrill, M. D., 96, 122
  - Merrill, P. F., 137
  - Mesokurtic (normal) standard distribution, 244, 245*fig*
  - Metcalf, L., 150, 151
  - Microsoft, 7, 8
  - Millman, J., 285
  - Minority populations
    - adverse impact issue and, 372–373, 380–383, 387
    - affirmative action and, 385–386
    - avoiding blanket exclusions of specific, 378–380
    - compliance documentation regarding, 402–403
    - diversity issue when writing test items, 158
    - “four-fifths” rule and, 375
    - Title VII (Civil Rights Act of 1964) on equal rights of, 368, 370, 374
  - Misclassification
    - borderline cut-off decisions reducing, 284–285
    - consequences of, 267–268
  - Mississippi State Department of Public Welfare, Walls v.*, 375
  - Mobility impairments, 390, 393
  - Modified Angoff, 266
  - Monitoring tests
    - issues of, 223–225
    - tips on, 226
  - Moody, Albemarle Paper Company v.*, 375–376, 386
  - Moroz, R., 152
  - Motor skill hierarchy, 97
  - Motorola, Myart v.*, 374
  - Mulkey, J., 186
  - Mullins, M. A., 63
  - Multiple-choice items
    - advantages/disadvantages of, 136
    - cautions against specific variants of, 136–146
    - correction-for-guessing on, 285–287
    - described, 132–133
    - distractors, 48, 160–162, 215
    - examples at different Bloom levels, 133–135
    - guidelines for writing, 160–162, 163–165
    - test directions for, 218
    - See also* Test item types
    - Myart v. Motorola*, 374
- N
- Nathan, B. R., 370
  - National Board of Medical Examiners, 159
  - Nedelsky’s method, 266
  - Negatively skewed curve, 244, 246*fig*
  - 1974 Standards for Educational and Psychological Tests*, 394

No Child Left Behind (2001), 10  
 Norm-referenced tests (NRTs)  
   comparing CRTs and, 25–30  
   using different correlation coefficients for, 293  
   frequency distributions of, 25–28  
   Livingston's Coefficient kappa ( $\kappa^2$ )  
     used with, 296, 297  
   practice statements on, 29–30  
   reporting scores of CRT vs., 358  
   statistically based construction of, 33  
*See also* Tests  
 Normal curves. *See* Standard normal curve  
 Norton, R. E., 78, 79  
 Novell, 7  
 Novick, M. R., 296  
 Nuclear Regulatory Commission (NRC), 58

## O

Objectives  
   behavior component of, 107–108  
   Certification Suite and, 109–116  
   characteristics of good, 107–109  
   classification schemes for, 122–127*t*  
   conditions component of, 108  
   converting job task statement to, 116–118  
   documentation establishing content validity of, 65  
   establishing content validity of, 46  
   item writing and role of, 106–109  
   legal issues related to, 109  
   practice statements on, 128  
   standards component of, 108–109  
   test length and domain size of, 168–169  
   test length and homogeneity of, 169–170  
   three fundamental purposes of, 106–107  
 Objectives-based test construction, 34  
 Occupational Health and Safety Act (OHSA) [South Africa], 59

Occupational Safety and Health Administration (OSHA) [U.S.], 59, 401  
 Open skill testing, 192–194*fig*  
 Open-ended questions, 129  
 Optical scanning (OPSCAN systems), 175–176  
 Organizational culture, 221–222  
 Organizations  
   affirmative action policies of, 385–386  
   documentation retention policies by, 407–409, 413–414  
   employment selection laws followed by, 368–370  
   response to invalid decisions by judges, 352–353  
   test time limits mandated by, 221–222  
   *Uniform Guidelines* followed for employee selection by, 63, 76, 368–376  
 Oshima, T. C., 256

## P

P-value, 249  
 "Paper-and-pencil" items. *See* Cognitive items  
 Paper-and-pencil tests, parallel forms for, 260–262  
 Parallel forms  
   benefits of, 259  
   for computerized item banks, 262–264  
   creating, 49  
   documentation on, 66–67  
   for paper-and-pencil tests, 260–262  
 Patriot Act, 59  
 $P_{\text{chance}}$  coefficient  
   calculating cognitive test reliability using, 309*fig*–310*fig*  
   calculating inter-rater reliability using, 325*fig*–335  
 Performance test rating scales  
   behaviorally anchored, 188–189*fig*  
   checklists, 190–192*fig*

- descriptive, 188
- numerical, 188*fig*
- Performance test reliability
  - determining, 52
  - documenting, 67
  - inter-rater, 19, 299, 322–344
  - issues related to, 319–320
  - repeated performance/consecutive success for, 344–347
  - types of rating errors, 320–322
- Performance Testing Council (PTC), 184–185, 186
- Performance tests
  - cut-off scores using data from, 268–269
  - description and functions of, 183–184
  - difference between knowledge tests and, 184–186
  - inter-rater reliability using data from, 324*t*
  - open skill testing using, 192–194*fig*
  - product vs. process in, 187
  - rating scales used in, 187–192*fig*
  - reliability of, 52, 67, 319–356
  - special administration considerations for, 225–226
- Phi (j) coefficients
  - calculating inter-rater reliability using, 335*fig*–344
  - calculating single-test reliability using, 298
  - calculating two-test reliability using, 301, 302–306
  - comparison of agreement, kappa, and, 313–314
  - converting to Fisher's Z coefficient, 338–339*t*, 340
  - description of, 204, 297–298, 302, 335
  - high percent of agreement leading to negative, 353–356
  - how high should phi be?, 304–305
  - practice for calculating, 305–306
  - recommended over the agreement coefficient, 308
  - table for item analysis, 255*fig*
  - table for test-retest reliability, 303*fig*–304*fig*
  - when rater passes all test-takers, 349–352*fig*
- Physical test factors, 222
- Pilot test administration
  - 1: sequencing test items, 217–218
  - 2: test directions, 218
  - 3: test readability levels, 219–220*fig*
  - 4: lexile measure, 220
  - 5: formatting the test, 220–221
  - 6: setting time limits, 221–222
  - 7: physical factors, 222
  - 8: psychological factors, 222–223
  - 9: giving and monitoring the test, 223–225
  - 10: special considerations for performance tests, 225–226
- Pilot test steps
  - 1: determine the sample, 212–213
  - 2: orient the participants, 213–214
  - 3: give the test, 214
  - 4: analyze the test, 214–215
  - 5: interview the test-takers, 215–216
  - 6: synthesize the results, 216
- Pilot tests
  - actions prior to administering, 217–222
  - administration of, 222–231
  - benefits of, 211–212
  - conducting initial, 47–48
  - documentation on initial, 66
  - honesty and integrity in, 231–240
  - six steps in the process of, 212–216
- Piracy, 238
- Plake, B. S., 266, 269
- Platykurtic (larger) standard deviation, 244, 245*fig*
- Po*. See Agreement ( $p_o$ ) coefficient
- Point-biserial correlation, 48–49, 215, 250–251
- Positively skewed curve, 244, 246*fig*
- Posttests, 31
- Power tests, 221
- Practice analysis, 186

## Practice statements

- on adverse impact, 381–383
- on Bloom cognitive level of objectives, 128
- on calculating agreement ( $p_o$ ) coefficient, 307
- on calculating kappa ( $\kappa$ ) coefficient, 311–313
- on Certification Suite, 155–158
- comparing NRTs and CRTs, 29–30
- on concurrent validity, 206–207
- on cut-off scores procedures, 271–272, 276–278, 280–281
- on determining which type of CRT to use, 31–32
- on inter-rater reliability of performance tests, 330–335
- on item analysis, 252–253
- on reliability and validity, 21–22
- on test construction methods, 34–35
- on test data from two administrations of CRTs, 305–306
- on test item guidelines, 163–165
- on test scores (inflation and reduction factors), 18
- on testing, measurement, evaluation, and assessment, 16

## Predictive validity

- definition of, 22, 24, 208
- determining, 208
- phi ( $j$ ) table for, 208*fig*

## Prerequisite tests, 30

## Privacy issues

- of independent testing center, 227–228
- when writing test items, 158

## Psychological test factors, 222–223

## Pull down test item, 130

## Q

## Qualification, 9

## Question types

- closed-ended, 129
- open-ended, 129

## Questionmark, 43, 62

## R

Racial/ethnic groups. *See* Minority populations

## Rate (skill testing), 194

Raters. *See* Judges

## Rating errors

- central tendency error, 321–322
- error of standards, 320
- halo errors, 321
- leniency error, 322
- logic errors, 321
- similarity error, 321

## Rating instruments

- creating, 47
- documentation on creating, 65

## Rating scales

- behaviorally anchored, 188–189*fig*
- checklists, 190–192*fig*
- descriptive, 188
- numerical, 188*fig*

Readability levels, 219–220*fig*Real World—Level A, 110–111, 113*t*, 117

## Reasonable accommodation concept, 388

## Reasonable reconsideration concept, 376

## Reduction factors, 17–18

## Reliability

- of cognitive tests, 289–317
- definition of, 18–19, 289–290
- determining, 50–52
- documenting cognitive/performance tests, 67
- equivalence, 19, 51
- importance of, 317
- inter-rater, 19, 290, 293
- logistics of establishing test, 314–316
- performance tests, 52, 319–356
- practice statements on, 21–22
- relationship between test length and, 166–167
- test-retest, 51–52, 296–297
- types of, 293–294
- writing items for test, 166–167
- See also* Tests

## Reliability examples

- both reliable and valid, 21*fig*
- neither reliability nor valid, 21*fig*
- reliable but not valid, 20*fig*

## Reliability techniques

- recommendations for choosing, 316–317
- single-test administration, 293, 294–299
- two-test administration, 294, 299–313
- understanding the reliability coefficients used in, 313–314

## Reporting scores

- CRT vs. NRT, 358
- described, 52–53
- documentation of, 67–68
- legal issues regarding archiving tests, 362
- summing subscores, 258–361
- what to report to manager, 361–362

## Retake policy violations, 238

## Reusable instructional objects (RIOs), 264

## Reusable learning objects (RLOs), 264

## Revising cut-off scores, 268

## Robertson, R., 31

## Robins, D., 110

## Ruyle, K. E., 264

## S

## Saltatory cut-off score, 287–288

## Sapnar, G., 180

## Sarbanes-Oxley Act, 59

## Sarvela, P. D., 256

## SAT, 33

## Sc, 296, 297–298

Scenarios—Level C, 111, 113*t*, 118, 156–158

## Schaefer, M. A., 202

## Schmidt, B. S., 202

## Schnipke, D., 63

## Schonemann, P., 296

## Securities and Exchange Commission (SEC), 59

Security issues. *See* Test security

## Shepherd, E., 43

## Short answer test items

- description of, 147–148
- guidelines for writing, 162–165

## Show vs. tell metaphor, 152–155

## Shrock, S., 43, 44, 53, 110, 288, 294, 296, 298

## Similarity error, 321

## Singer, R., 193

## Single-test administration techniques

- calculating reliability for, 297–299
- internal consistency, 294–296
- listed, 293
- outcomes of using for CRTs, 298–299*t*
- practice using CRT data, 305–306
- squared-error loss, 296
- threshold-loss, 296–297

## SMEs (subject-matter experts)

- assessment planning by, 41
- content validity verification by, 197
- cookbook for writing cognitive test items with, 172–174

## Smith, C., Jr., 386

## Soder, D. A., 77

Special needs. *See* Accommodating special needs

## Speeded tests, 221

## Squared-error loss coefficients, 296

## Standard 1.1, 63–64

## Standard 1.6, 64

## Standard 3.2, 64

## Standard 3.11, 64

## Standard 14.9, 64

## Standard 14.10, 75–76

## Standard 14.14, 64, 76

## Standard deviations

- five most common, 244–246*fig*
- larger (platykurtic), 244, 245*fig*
- mastery curve, 247*fig*–248
- meaning of, 241–243
- negatively skewed curve, 244, 246*fig*

Standard deviations (continued)  
 normal (mesokurtic), 244, 245*fig*  
 positively skewed curve, 244, 246*fig*  
 problems with mastery distributions and, 247*fig*–248  
 smaller (leptokurtic), 244, 245*fig*  
 Standard error of measurement, 282–284  
 Standard normal curve, 242*fig*  
 Standards  
 documentation, 63–64  
 as objectives components, 108–109  
 testing, 10–11  
*Standards for Educational and Psychological Testing* (AERA/APA/NCME Joint Committee, 1999), 10, 63–64, 75–76  
*Standards for Educational and Psychological Tests* (APA, 1974), 394  
 Statistically based test construction, 33  
 Statistics  
 item, 248–251, 255–257  
 item analysis, 48–49, 66, 251–255*fig*  
 standard deviations, 241–248  
 Stenner, A. J., 220  
 Style elements, 193  
 Sudweeks, R. R., 137  
 Substitutability issue, 269–270  
 Swaminathan, H., 296  
 Swanson, D. B., 136, 159  
 Sylvan, 7–8  
 Synthesis level (Bloom's taxonomy), 93–94

## T

*Taxonomy of Educational Objectives*, 123  
 Taylor, R. G., 255, 294  
*Teal, Connecticut v.*, 384  
 Tessmer, M., 78  
 Test administration  
 accommodating special needs of test-takers, 387–394  
 of initial pilot test, 217–226

security measures during, 234–240  
 single-test techniques for, 293, 294–297  
 special issues of performance tests, 225–226  
 test reliability and conditions of, 316  
 two-test techniques for, 294, 299–313  
*See also* Testing  
 Test construction methods  
 objectives-based, 34  
 practice statements on, 34–35  
 statistically based, 33  
 topic-based, 32–33  
 Test item types  
 essay, 148–149, 163–165  
 fill-in, 147, 162, 163–165  
 matching, 132, 159–160, 163–165  
 overview of, 130  
 short answer, 147–148, 162–165  
 true/false, 131, 159, 163–165  
*See also* Multiple-choice items  
 Test items  
 cookbook for working with SMEs  
 on cognitive test, 172–174  
 guidelines for writing, 158–170  
 newer computer-based, 129–130  
 security regarding preview of, 234–235  
 six most common types of, 130–149  
 test reliability and role of, 314–315  
*See also* Cognitive items  
 Test items guidelines  
 basic issues to consider, 158  
 for most common item types, 159–165  
 practice statements on, 163–165  
 Test length  
 criticality of decisions and, 167–168  
 determining number of items and, 166  
 domain size of objectives and, 168–169

- homogeneity of objectives and, 169–170
- relationship between reliability and, 166–167
- research on, 170
- resources available and relationship to, 168
- summary of determinants of, 170–171
- Test scores
  - cut-off, 49–50, 67, 265–288
  - definition and meaning of, 17–18
  - hierarchy validation using posttest, 101*fig*–102*fig*
  - inflation and reduction factors of, 17–18
  - lack of variance in, 295
  - practice statement on, 18
  - reliability of, 18–19
  - reporting, 52–53, 67–68, 357–363
  - squared-error-loss coefficients on, 296
- Test scores systems
  - computer-based testing, 176–180
  - Computerized Adaptive Testing (CAT), 180–181
  - hand scoring, 175
  - OPSCAN (optical scanning), 175–176
- Test security
  - limiting interaction/collaboration of test-takers, 235
  - organization-wide policies regarding, 236–237
  - protecting preview of test items, 234–235
  - using statistical methods to detect cheating, 237–240
- Test security plan
  - described, 68–69
  - matrix of, 71–73
  - topic areas of, 69–71
- Test-retest reliability
  - calculating kappa coefficient using data from, 311–313
  - described, 51–52
  - example for a CRT, 301*t*
  - phi table for, 303*fig*–304*fig*
  - simulated by  $S_c$ , 296–297
  - as two-test administration technique, 300–301*t*
- Test-takers
  - accommodating special needs of, 387–394
  - agreement coefficient of, 306–308
  - anxiety levels of, 222–224
  - avoiding irrelevant assignments prior to test, 226
  - cheating by, 231–234
  - controlling talking by, 223, 224–225
  - establishing test reliability and sampling, 315–316
  - giving and monitoring, 223–225, 226
  - giving pilot test directions to, 218
  - interviewing pilot test, 215–216
  - item response theory (IRT) predicting responses by, 256–257
  - test security measures and, 234–240
  - when passed regardless of performance, 349–353
- Testing
  - accommodating special needs test-takers, 387–394
  - definition of, 15
  - using an independent testing center for, 227–231
  - practice statement on, 16
  - reasons behind demand for, 1–2
  - See also* Test administration
- Tests
  - certification, 31
  - diagnostic, 31
  - documenting purpose of, 65
  - entry, 30–31
  - equivalency, 31
  - Graduate Record Exam (GRE), 47
  - high stake, 10
  - initial pilot, 47–48, 66, 211–240
  - item writing and length of, 166–171

## Tests (continued)

- legal issues regarding archiving of, 362
- low stakes, 10
- Medical College Aptitude Test (MCAT), 28
- performance, 52, 67, 183–194*fig.*, 225–226
- posttests, 31
- power, 221
- prerequisite, 30
- security measures related to, 234–240
- six purposes in training settings, 30–32
- speeded, 221
- standards for, 10–11
- as unlawful employee selection procedure, 372–373
- See also* Cognitive tests; Criterion-referenced tests (CRTs); Norm-referenced tests (NRTs); Reliability
- Thalheimer, W., 43
- Thigpen v. UPS*, 412–413
- Thomas, S., 8
- Thompson, D. E., 77
- Thompson, T. A., 77
- Threshold-loss technique, 296–297
- Time limits
  - adhering to, 225
  - pilot test, 221–222
- Tinker, T., 180
- Title 29 (Code of Federal Regulations), 371
- Title VII (Civil Rights Act of 1964), 368, 370, 374
- Topic-based test construction, 32–33
- Training tests, 30–32
- True/false items
  - described, 131
  - guidelines for writing, 159, 163–165
- Two-test administration techniques
  - calculating reliability for, 301–313
  - equivalence reliability, 19, 51, 299–300
  - listed, 294

- practice using CRT data, 305–306
- test-retest reliability, 51–52, 296–297, 300–301*t*

## U

- UBS Warburg, Zubulake v.*, 409
- UK Financial Services Authority (FSA), 59
- Undue hardship concept, 388
- The Uniform Guidelines on Employee Selection Procedures*
  - on adverse impact, 380–385, 387
  - on affirmative action, 385–386
  - “bottom line” standard used by, 383–385
  - bringing a legal claim under, 368
  - documentation recommended by, 63
  - “four-fifths” rule of, 375
  - job analyses recommendations by, 76
  - legal challenges to testing and the, 373–376
  - origins and development of, 370–371
  - purpose and scope of, 371–373
  - on reasonable reconsideration, 376
  - on tests as unlawful selection procedure, 372–373
  - validation techniques recommended by, 394–400
- See also* EEOC (Equal Employment Opportunity Commission); Employee Selection Laws; Legal issues
- United States Nuclear Regulatory Commission, 161
- Upper-lower index, 253*fig.*–254
- UPS, Bates v.*, 379–380
- UPS, Thigpen v.*, 412–413
- U.S. Equal Opportunity Commission, 371
- U.S. Medical Licensure Exam, 239
- U.S. Supreme Court cases
  - Albemarle Paper Company v. Moody*, 375, 386
  - Bates v. UPS*, 379–380

- Connecticut v. Teal*, 384  
*Fire Fighters Institute for Racial Equality v. City of St. Louis*, 102  
*Griggs v. Duke Power Company*, 374–375  
*Kirkland v. Department of Correctional Services*, 77  
*Myart v. Motorola*, 374  
*Thigpen v. UPS*, 412–413  
*Walls v. Mississippi State Department of Public Welfare*, 375  
*Zubulake v. UBS Warburg*, 409
- V**
- Validity  
 concurrent, 22, 23–24, 202–207*fig*  
 construct of, 185–186  
 definition of, 20, 195, 289–290  
 face, 22, 23, 185–186, 196–197  
 four types listed, 22, 195  
 importance of, 317  
 practice statements on, 21–22  
 predictive, 22, 24, 208*fig*  
 summary comment about, 209  
*See also* Content validity; Hierarchy validation
- Validity examples  
 both reliable and valid, 21*fig*  
 neither reliability nor valid, 21*fig*  
 reliable but not valid, 20*fig*
- Validity study  
 documentation of, 403  
 issues to consider in, 394–396  
 on ranking of performers on a test, 396  
 step-by-step guide on, 397–400  
*Uniform Guidelines* recommendations on, 394–396
- Validity study guidelines  
 1: obtain professional guidance, 397  
 2: select legally acceptable validation strategy for your particular test, 397–398  
 3: understand and employ standards for content-valid tests, 398–399
- 4: evaluate overall test circumstances to assure equality of opportunity, 399–400
- Variables  
 correlation between, 290–293  
 graphic illustrations of correlation between, 291*fig*  
 Verbal information hierarchy, 97  
 Visual impairments, 390, 393–394  
 Volatile retakes, 238
- W**
- Walls v. Mississippi State Department of Public Welfare*, 375
- Waltz, C. F., 200
- Wang, C., 177
- Wang, N., 63
- Websites  
 Caveon Test Security, 69  
*Cheating in the New* repository, 232  
 Performance Testing Council (PTC), 184  
 Test Security Planning, 69  
 U.S. Medical Licensure Exam, 239
- Weighting tests  
 calculating cut-off score for course, 360*t*  
 calculating individual performance score, 360–361*t*  
 issues to consider when, 358–360
- Witt, E., 63
- Wood, B., 137
- Work behavior sample, 399
- Work product sample, 399
- Work-Learning Research, 43
- Wright, B. D., 220
- Wynd, C. A., 202
- Y**
- Yelon, S., 192–193, 194
- Z**
- Z coefficient, 338–339*t*, 340
- Zicky, M., 272
- Zikey, M. J., 265
- Zubulake v. UBS Warburg*, 409

