Index

Ability tests:
- described, 4
- item difficulty and, 247–248
Abscissa, 49
Absolute scaling, 250
ACT:
- SAT and, 17
- science test, 180
- WorkKeys system, 182
Additivity, 44
Administration, test:
- digital age/Internet, 306–308
- examiner preparation, 304–306
- test environment, 298–299
- test takers, preparing, 299–304
AEE. See Automated essay evaluation (AEE)
AERA. See American Educational Research Association (AERA)
AFQT. See Armed Forces Qualification Test (AFQT)
Age:
- chronological (see Chronological age (CA))
- mental (see Mental age)
- performance, 85
Age differentiation, 198
Age equivalent scores, 85
- See also Mental age scores
Allport-Vernon Study of Values, 20
Alpha:
- coefficient (see Coefficient alpha)
- Cronbach’s (see Cronbach’s alpha)
Alternate-form reliability, 137
Alternate forms, 105–106
Alzheimer’s disease, 204
American Educational Research Association (AERA), 6, 32
American Psychological Association (APA):
- Insurance Trust, 318
- Internet resources, 30–31
- Task Force on Psychological Testing on the Internet, 308
- Task Force on Test User Qualifications, 28
- Testing Standards and, 6
- test user qualifications and, 279–280
- website, 282
American Psychologist, 217, 282
Analysis:
- confirmatory factor (see Confirmatory factor analysis (CFA))
- exploratory factor (see Exploratory factor analysis (EFA))
- path, 71
- profile, 159, 160
Analysis of variance (ANOVA), 149
Anchor items, 250
Anchor tests, 106
ANOVA. See Analysis of variance (ANOVA),
Anxiety:
- math, 38, 266
- test, 38, 185, 301–302
APA. See American Psychological Association (APA)
APA PsycNET, 31
Aptitude(s):
- batteries, multiple, 18–19
- tests for, scholastic, 16–17
- tests of special skills and, 18
Armed Forces Qualification Test (AFQT), 225
Armed Services Vocational Aptitude Battery (ASVAB):
CAT-ASVAB, 225, 272
Project A and, 224
Army Alpha:
described, 15
validity and, 168
Army Beta, 15, 187
Assembly line output per hour, 47, 205, 206, 207, 210–211
Assessment:
criterion-referenced test interpretation and clinical, 122
defined, 281
performance (see Performance assessment)
psychological, 27
steps in process, 25–26
vs. testing, 24–25
therapeutic model of, 24
See also Neuropsychological assessment
Assessment centers, 118
Assessment of Older Adults With Diminished Capacity, 282
Assessment tools, 292–295
biodata, 292–294
interview data, 294
observation, 294
Association of Test Publishers (ATP), 30, 32
ASVAB. See Armed Services Vocational Aptitude Battery (ASVAB)
Atkins v. Virginia, 111
ATP. See Association of Test Publishers (ATP)
Attitudes Toward Women Scale (AWS), 184
Authors, test, 7
Automated essay evaluation (AEE), 277, 309
Automated scoring strategies, 277
AWS. See Attitudes Toward Women Scale (AWS)
Bar graphs, 49
Baseline, 49
Base rates, 289, 291
Battery, 4–5
BC. See Buros Center for Testing (BC)
BDAS. See Boston Diagnostic Aphasia Examination (BDAS)
Beck Depression Inventory-II (BDI-II), 122, 183, 193
Bell curve. See Normal curve model
Beta III, 187–189
BI. See Buros Institute of Mental Measurements (BI)
Bias:
assessor, 306
differential validity and, 213–214
intercept, 214
item, 233, 237, 247, 250, 271, 272, 273–274
moderator variables and, 213
psychometrics and, 234
slope, 214
test, 216–217, 295
The Big Test: The Secret History of the American Meritocracy, 17
Bilaterally symmetrical, 56
Binet, Alfred, 12
Binet-Simon intelligence scales, 37
Biodata, 292–294
Bivariate data:
correlation and, 66, 67
Pearson r and, 72–73, 75
regression toward the mean and, 70
scatterplots and, 68, 69, 73, 74, 75
validity and, 205, 206, 216
Bivariate statistics, 66
Bluetooth, 307
Boston Diagnostic Aphasia Examination (BDAS), 183
Bracken Basic Concept Scale, 198
Brain damage, 21–22
Buros, Oscar, 33
Buros Center for Testing (BC), 33
Buros Institute of Mental Measurements (BI), 33
CA. See Chronological age (CA)
CAB. See Clerical Abilities Battery (CAB)
Calendar time, 43
California Verbal Learning Test (CVLT): Q-interactive system, 307
Second Edition (CVLT-II), 93
Campbell, John P., 224
Carnegie Interest Inventory, 19
Carroll, John, 189
CAT. See Computerized adaptive testing (CAT)
Categorical data, 42
Cattell, James McKeen, 12
Cattell, R. B., 20
CBTA. See Computer-Based Test Administration (CBTA)
CBTIs. See Computer-based test interpretations (CBTIs)
CEEB. See College Entrance Examination Board (CEEB) scores
Ceiling, test, 92, 251
Center(s):
assessment, 118
Buros Center for Testing (BC), 33
Educational Resources Information Center (ERIC), 31–32
Central tendency, measures of, 50–52
Certification, standardized tests and, 16, 280
CFA. See Confirmatory factor analysis (CFA)
Charts, 49
Chauncey, Henry, 17
China, civil service examinations, 9
Chronological age (CA):
Developmental Profile and, 85
ratio IQs and, 45–46
Classical test theory (CTT), 262–265
vs. item response theory, 262–264
score reliability and, 270
shortcomings of, 264–265
Classification, 222–223
Clinical psychology, testing and, 10
Clinical testing:
neuropsychological tests, 21–22
personality inventories, 19–20
projective techniques, 20–21
Coefficient(s):
correlation (see Correlation coefficients)
multiple correlation ($R$), 77–78
validity, 207, 208
Coefficient alpha, 141–143, 144
See also Cronbach’s alpha
Coefficient of determination, 68, 206, 208
Cognitive Assessment System–2, 100
College Board. See SAT
College Entrance Examination Board (CEEB) scores, 98, 101
College-Level Examination Program (CLEP) Introductory Psychology Test, 180
Computer-Based Test Administration (CBTA), 306–307
Computer-based test interpretations (CBTIs), 316
Computerized adaptive testing (CAT):
  - CAT-ASVAB, 225
  - item response theory and, 108–109, 271

Computer programs:
  - LISREL, 192
  - statistical operations and, 94

Conant, James Bryant, 17

Concurrent validation, 202

Confirmatory factor analysis (CFA):
  - cognitive and personality traits, 193
  - described, 187
  - WAIS-IV and BDI-II data, 192–193

Co-norming, 107

Consequential validity, 226

Constant, defined, 38

Construct:
  - defined, 170, 172–173
  - representation, 173
  - validity, 172–176

Constructed-response items,
  - 244–246
  - advantages of, 244–245
  - disadvantages of, 245–246
  - response length, 246

Contamination, criterion, 210

Content heterogeneity, 140, 141

Content-referenced testing, 116

Content sampling error, 136–137
  - reliability coefficients and, 133
  - time sampling error and, 145

Continuous data, 44

Continuous variables, 40

Convenience norms, 90–91

Convergent validity, 196

Coping With Psychiatric and Psychological Testimony, 285

Correcting Fallacies About Educational and Psychological Testing, 23

Correlation:
  - facts, essential, 71
  - linear, 66–68
  - notation for, basic, 73
  - point biserial, 77
  - regression and, 65–70, 68
  - restricted range and, 77
  - scatterplots and, 68, 69

Correlation coefficients, 70–78
  - correlation methods and, 66
  - defined, 70–71
  - Pearson r (see Pearson product-moment correlation coefficient (Pearson r))
  - variance and, 68

See also specific types

Correlation matrix, 187, 188

Crawford Small Parts Dexterity Test (CSPDT), 181

Criterion, 119

Criterion contamination, 210

Criterion measures, 199

Criterion-referenced test interpretation.
  - See Score interpretation, criterion-referenced

Criterion-related validation studies, 209–221

Critical ratios, 61

Cronbach’s alpha, 141, 144
  - See also Coefficient alpha

Cross-validation, 19, 212, 237

CSPDT. See Crawford Small Parts Dexterity Test (CSPDT)

CTT. See Classical test theory (CTT)

CVLT. See California Verbal Learning Test (CVLT)

DAS. See Differential Ability Scales (DAS)

Data:
  - base rate, 289
  - bivariate, 67, 74
categorical, 42
continuous, 44
interview, 294
validity, 289
Data, numerical description of, 49–53
central tendency, measures of, 50, 51–52
variability, measures of, 52–53
Database, PsycINFO, 31
Dean-Woodcock Neuropsychological Battery, 4
Decision making:
clinical, 291–292, 293
psychological tests and, 22–23
Decision theory, 287–288
Delayed alternate-form reliability, 145
Dementia, 202, 203, 204, 210, 223
Depression, 202, 223
   See also Beck Depression Inventory-II (BDI-II)
Derived scores, 87
Descriptive statistics:
   raw data and, 48–49
test scores and, 54
Design, universal, 273
Deviation IQs, 98, 99–100
Diagnostic and Statistical Manual of Mental Disorders (DSM-5), 158
Diagnostic questions, assessment and, 26
Dichotomous variables, 39
DIF. See Differential item functioning (DIF)
Differential Ability Scales (DAS), 271
Differential Ability Scales–II, 100
Differential item functioning (DIF), 275–276
Differential prediction, 213, 214, 215, 220, 221, 225
Differential validity, 222–223
Discrete variable, 39
Discriminant functions, 223
Discriminant validity, 196
Discrimination, 247
Disorders, psychiatric, 202
Dissimulation, test-taker, 304
Distractors, difficulty and, 252–253
Distribution(s):
   frequency, 49, 50
   leptokurtic, 63
   mesokurtic, 63
   nonnormal (see Nonnormal distributions)
   platykurtic, 63
   sampling, 58–59, 61
   standard normal, 57
Domain(s):
   content, 115–117
   developmental, 84
Domain-referenced testing, 116
DSM-5. See Diagnostic and Statistical Manual of Mental Disorders
Ebbinghaus, Hermann, 12
Education:
special, 13
testing and, 170, 180
   Educational Resources Information Center (ERIC), 31–32
Educational testing, standards for, 280–281
Educational Testing Service (ETS), 17, 98
   CATs and, 109
   Test Collection database
test item examples, 240
EFA. See Exploratory factor analysis (EFA)
Effect sizes, 218
Embretson, Susan, 167, 173, 174, 261, 262–263, 264, 267, 271, 276
Empirical criterion keying, 19, 20
Employment:
   job analysis, 177–178
   meta-analyses and, 216, 217, 218
Employment: (Continued)
occupational testing, 179, 181–182
personnel testing and, 18–19
range restriction and, 76, 77
selection, 288
See also Vocational guidance
Equal-unit scales. See Interval scales
Equating procedures, 105–108
alternate forms, 105–106
anchor tests, 106
fixed reference groups, 106–107
simultaneous norming, 107–108
Equations, multiple regression, 71, 211
ERIC. See Educational Resources
Information Center (ERIC)
Error(s):
content sampling (see Content
sampling error)
estimates of, 40
time sampling (see Time sampling
error)
type I/II, 217
Eta, 77
Ethical Principles of Psychologists and
Code of Conduct, 227, 304, 319
Ethical use of tests:
administration and scoring, 298–310
resources, additional, 281–284
rights and responsibilities, test takers, 283
source of information on, 279–284
test score interpretation and scoring, 310–318
test selection. see Test selection
ETS. See Educational Testing Service
(ETS)
Examiner preparation, 304–306
Expectancy tables/charts, 119–120
Experimental results, validity and, 198
Exploratory factor analysis (EFA), 186–190, 224–225
FA. See Factor analysis (FA)
Factor analysis (FA):
conceptual foundation for, 20
exploratory, 186–190
Factor loadings, 187
Factor matrix, 187–189
Fair Game? The Use of Standardized
Admissions Tests in Higher
Education, 303
False negative/false positive findings, 291
“FAQ/Finding Information About
Psychological Tests,” 30
FFM. See Five-factor model (FFM)
Fischer, Constance, 24
Five-factor model (FFM), 190
Fixed-response items, 240
Floor, test, insufficient, 93,
251–252
Fluid intelligence (Gf), 193
Flynn, J. R., 110
Flynn effect, 110–112, 158
fMRI. See Functional magnetic
resonance imaging (fMRI)
Forced-choice items, 241–242
Forensic applications, psychological
tests, 285
Form(s):
alternate, 105–106
parallel, 106–107
qualification, 30
Form boards, 21
Formula(s):
basic, 54
Kuder-Richardson formula 20
(K-R 20), 141–144
Spearman-Brown (S-B), 138–140
Fourfold coefficient, 77
Frames of reference:
item response theory and, 122–123
relationship among, 120–122
France, testing in, 10, 13, 14
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free-response items. See Constructed-response items</td>
<td></td>
</tr>
<tr>
<td>Frequency distributions, 49, 50</td>
<td></td>
</tr>
<tr>
<td>FSIQs. See Full Scale IQs (FSIQs)</td>
<td></td>
</tr>
<tr>
<td>Full Scale IQs (FSIQs), 112, 186, 193</td>
<td></td>
</tr>
<tr>
<td>Function(s):</td>
<td></td>
</tr>
<tr>
<td>discriminant, 223</td>
<td></td>
</tr>
<tr>
<td>item information, 270–271</td>
<td></td>
</tr>
<tr>
<td>Functional magnetic resonance imaging (fMRI), 37</td>
<td></td>
</tr>
<tr>
<td>g. See General intelligence factor (g)</td>
<td></td>
</tr>
<tr>
<td>Galton, Francis, 68–70</td>
<td></td>
</tr>
<tr>
<td>General Aptitude Test Battery (GATB), 215</td>
<td></td>
</tr>
<tr>
<td>General intelligence factor (g), 189, 193</td>
<td></td>
</tr>
<tr>
<td>Generalizability theory (G theory), 149, 151</td>
<td></td>
</tr>
<tr>
<td>Germany, testing in, 10, 11, 12</td>
<td></td>
</tr>
<tr>
<td>Gesell, Arnold, 83</td>
<td></td>
</tr>
<tr>
<td>Gesell Developmental Schedules, 83</td>
<td></td>
</tr>
<tr>
<td>Gf. See Fluid intelligence (Gf)</td>
<td></td>
</tr>
<tr>
<td>Globalization, 6</td>
<td></td>
</tr>
<tr>
<td>Goldstein, Kurt, 21</td>
<td></td>
</tr>
<tr>
<td>Goodness-of-fit statistics, 191</td>
<td></td>
</tr>
<tr>
<td>Grade equivalent scores, 86</td>
<td></td>
</tr>
<tr>
<td>Grade inflation, 295</td>
<td></td>
</tr>
<tr>
<td>Graduate Record Exam (GRE):</td>
<td></td>
</tr>
<tr>
<td>CEEB scores and, 98</td>
<td></td>
</tr>
<tr>
<td>IRT models and, 271</td>
<td></td>
</tr>
<tr>
<td>scholastic aptitude tests, 17</td>
<td></td>
</tr>
<tr>
<td>test item examples, 240</td>
<td></td>
</tr>
<tr>
<td>validity studies and, 219–221</td>
<td></td>
</tr>
<tr>
<td>Graphs, 49</td>
<td></td>
</tr>
<tr>
<td>GRE. See Graduate Record Exam (GRE)</td>
<td></td>
</tr>
<tr>
<td>Greater male variability, hypothesis of, 63, 64</td>
<td></td>
</tr>
<tr>
<td>Group:</td>
<td></td>
</tr>
<tr>
<td>reference (see Reference group) testing, 14–15</td>
<td></td>
</tr>
<tr>
<td>Group Intelligence Scale, 15</td>
<td></td>
</tr>
<tr>
<td>G theory. See Generalizability theory (G theory)</td>
<td></td>
</tr>
<tr>
<td>Guidelines for Quality Control in Scoring, Test Analysis, and Reporting of Test Scores, 6</td>
<td></td>
</tr>
<tr>
<td>Guidelines for Translating and Adapting Tests, 6</td>
<td></td>
</tr>
<tr>
<td>Guilford, J. P., 20</td>
<td></td>
</tr>
<tr>
<td>Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, 318, 319</td>
<td></td>
</tr>
<tr>
<td>Heterogeneity, 140, 141, 142</td>
<td></td>
</tr>
<tr>
<td>Heteroscedasticity, 75</td>
<td></td>
</tr>
<tr>
<td>HIPAA. See Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule</td>
<td></td>
</tr>
<tr>
<td>Histograms, 49, 62</td>
<td></td>
</tr>
<tr>
<td>Hit rates, 200</td>
<td></td>
</tr>
<tr>
<td>Homogeneity, 141</td>
<td></td>
</tr>
<tr>
<td>Homoscedasticity, 75</td>
<td></td>
</tr>
<tr>
<td>Human figure drawings, 21</td>
<td></td>
</tr>
<tr>
<td>Hypothesis, greater male variability, 63, 64</td>
<td></td>
</tr>
<tr>
<td>Hypothesis testing, table of areas and, 340–341</td>
<td></td>
</tr>
<tr>
<td>ICC. See Item characteristic curve (ICC)</td>
<td></td>
</tr>
<tr>
<td>ID. See Intellectual disability (ID)</td>
<td></td>
</tr>
<tr>
<td>IDA. See Infant-Toddler Developmental Assessment (IDA)</td>
<td></td>
</tr>
<tr>
<td>Identity, 42</td>
<td></td>
</tr>
<tr>
<td>Idiographic, 27</td>
<td></td>
</tr>
<tr>
<td>Inconsistency:</td>
<td></td>
</tr>
<tr>
<td>interitem, 133, 140</td>
<td></td>
</tr>
<tr>
<td>internal (see Internal consistency)</td>
<td></td>
</tr>
<tr>
<td>Incremental validity, 289</td>
<td></td>
</tr>
<tr>
<td>Index of discrimination (D), 254–256</td>
<td></td>
</tr>
<tr>
<td>Industrial-organizational psychology, 219</td>
<td></td>
</tr>
<tr>
<td>Infant Psychological Development Scales, 85</td>
<td></td>
</tr>
</tbody>
</table>
Infant-Toddler Developmental Assessment (IDA), 83–84
Inferences, 175
Inferential statistics, 47, 61
Informed consent, 304–305
Intellectual disability (ID):
  Flynn effect and, 111
  measurement error and, 158
Intelligence:
  fluid (see Fluid intelligence (Gf))
  instrument titles and, 100
  scales (see Binet-Simon intelligence scales)
  See also General intelligence factor (g)
Intelligence quotient (IQ):
  deviation, 98, 99–100
  Flynn effect and, 111–112
  Ratio (see Ratio IQs)
  raw scores and, 82
  See also Deviation IQs; Full Scale IQs (FSIQs)
Interests, measures of, 19
Interitem inconsistency, 133, 140
Internal consistency:
  measures, 140–144
  score reliability and, 185
Internal-External Locus of Control Scale, 4
International Association for Computerized Adaptive Testing, 272
International Guidelines on Computer-Based and Internet Delivered Testing, 309
International Test Commission (ITC), 6
Internet:
  resources, 30–33
  testing on, 308–309
  test misuse and, 28
  See also Websites
Interquartile range (IQR), 52
Interscorer differences, 133, 134
Intertest comparisons, 103, 104
Interval scales, 43–44
Interview data, 294
Inventories:
  personality tests and, 3
  State-Trait Anxiety Inventory (STAI), 183–184
  See also Carnegie Interest Inventory;
  Millon Clinical Multiaxial Inventory–III (MCMI-III);
  Minnesota Multiphasic Personality Inventory (MMPI)
Iowa Tests of Basic Skills (ITBS), 121
Ipsative scores, 241
IQ. See Intelligence quotient (IQ)
IQR. See Interquartile range (IQR)
IRT. See Item response theory (IRT)
ITBS. See Iowa Tests of Basic Skills (ITBS)
ITC. See International Test Commission (ITC)
Item analysis, 247–260
  CTT and, 262–265
  IRT and, 265–271
  item difficulty, 247–253, 258–260
  item discrimination, 258–260
  item types (see Item types)
  item validity, 253–258
  methods, classical, 246–260
  qualitative, 233, 273
  quantitative methods, 247, 273–274
  test development and, 235–239
Item bias:
  qualitative analysis, 274
  quantitative analysis, 273–274
Item characteristic curve (ICC), 267–270
  defined, 267
  DIF and, 275
one-, two-, and three-factor models, 268
parameters in models and, 270
Item difficulty, 247–253
ability testing and, 247–248
distractors and, 252
gauging of, 248–251
item discrimination and, 258–260
item parameters and, 267
normal curve units and, 249
normative vs. relative basis for, 248
personality testing and, 253
test difficulty, test purpose and, 251–252
Item discrimination, item difficulty and, 258–260
Item fairness, 247, 272–273
Item generation, 276
Item information function, 270
Item parameters, 267
Item response theory (IRT):
applications of, 270–271
behavior samples and, 266
computerized adaptive testing and, 108–109, 246, 260–261
DIF and, 275
features of, essential, 265–272
frames of reference and, 122–123
item characteristic curve (ICC) and, 267–270
item information function, 270–271
item parameters and, 267
multidimensional models, 271
reliability and, 151
test information function, 270
unidimensional models, 265, 270
See also Item analysis, IRT and
Item selection, 235, 263
See also Reliability
Item-test regression analysis, 258–260
Item types, 239–246
constructed-response (see
Constructed-response items)
examples of, 240
fixed-response, 240
selected-response (see Selected-response items)
Item validity, 253–258
speed and, 257–258
statistics, 254–256
validation criteria, 253–254
Jackson Vocational Interest Survey, 242
Jigsaw puzzles, 21
Jobs. See Employment; Vocational guidance
Joint Committee on Testing Practices (JCTP), 282
Journal of Psychoeducational Assessment, 111
Judgments, evaluative, 26
Kaufman series of tests, 4, 100
Kelley, Truman L., 19
Kohlberg, Lawrence, 85
Kraepelin, Emil, 20
Kuder-Richardson formula 20 (K-R 20), 141–144
Kurtosis, 63
Latent trait models, 108
Latent trait theory. See Item response theory (IRT)
Latent variables, 170, 187, 189, 190–191
Law School Admission Test (LSAT), 17, 272
Lemann, Nicolas, 17
Leptokurtic distributions, 63
Licensing, standardized tests and, 16, 179
Linear correlation, 66–68
Linearity, deviations from, 74–75
Linear transformation, 94
See also Nonlinear transformations
LISREL, 192
Local norms, 90
LSAT, 272

MA. See Mental age
Magnetic resonance imaging (fMRI), functional, 36
Malingering, 304
Mantel-Haenszel (MH) technique, 275
Manual dexterity, 131, 170, 205, 206
Mastery testing, 118–119
Math anxiety, 38, 266
Matrix:
correlation, 187, 188
factor, 187–189
multitrait-multimethod (MTMMM), 195–198
MBTI. See Myers-Briggs Type Indicator (MBTI)
MCAT. See Medical College Admission Test (MCAT)
MCMI-III. See Millon Clinical Multiaxial Inventory–III (MCMI-III)
Mdn. See Median
Mean:
described, 51–52
normal curve model and, 56
z scores and, 94
Meaning of numbers, 40–47
interval scales, 43–44
nominal scales, 41–42
ordinal scales, 42–43
psychological measurement and, 47
psychological testing and, 44–46
ratio scales, 44
Measurement:
about, 38
inexact processes of, 40
levels of, 41
Measurement error:
defined, 127
and effect on people’s lives, 158
Median:
described, 51
normal curve model and, 56
Medical College Admission Test (MCAT), 17
Mental ability. See General intelligence factor (g); Intelligence quotient (IQ)
Mental age (MA):
ratio IQs and, 45–46
See also Age equivalent scores; Grade equivalent scores
Mental age scores, 85
Mental Measurements Yearbook (MMY) series, 33
Mental retardation, 111, 158
Mesokurtic distributions, 63
Messick, Samuel, 166, 167, 174, 226
Meta-analyses, 216–221
MH. See Mantel-Haenszel technique
Military occupational specialties (MOSs), 224, 226
Millon Clinical Multiaxial Inventory–III (MCMI-III), 195
Mini-Mental State Examination, 2nd Edition (MMSE-2):
concurrent validation, 202–204
as criterion referenced, 122
Minnesota Multiphasic Personality Inventory (MMPI):
Depression scale of, 4
Flynn effect and, 112
MMPI-2, 109, 195
MMPI-A, 90
original test, 20
raw scores and, 81
$T$ scores and, 98
validity and, 166, 185–186
Misuse of tests, 283, 285
MMPI. See Minnesota Multiphasic Personality Inventory (MMPI)
MMSE. See Mini-Mental State Examination (MMSE)
Mode:
described, 51
normal curve model and, 56
Model:
five-factor, 190
latent trait, 108
normal curve. see Normal curve model
Rasch, 267
Moderator variables, 213
MTMM. See Multitrait-multimethod matrix (MTMMM)
Multiple aptitude batteries, 18–19
Multiple-choice items:
carelessly written, 243
distractors and, 252
perceived weakness of, 241
traditional vs. complex, 273
Multiple correlation coefficient, 77–78
Multiple regression equations, 71, 211
Multitrait-multimethod matrix (MTMMM), 195–198
Multivariate statistics, 66, 190
Murray, Henry, 20
Myers-Briggs Type Indicator (MBTI), 241–242
NAEP. See National Assessment of Educational Progress (NAEP)
National Assessment of Educational Progress (NAEP), 123, 180, 240
National Council on Measurement in Education (NCME), 6, 32
NCME. See National Council on Measurement in Education (NCME)
NEO Personality Inventory—3 (NEO-PI-3), 197
NEO PI-3. See NEO Personality Inventory—3 (NEO-PI-3)
Neuropsychology:
assessment, 183
tests, 21–22
Nominal scales, 41–42
Nomological net, 173
Nomothetic span, 173
Nonlinear transformations, 101–102
See also Linear transformations
Nonnormal distributions, 62–65
kurtosis, 63
skewness (Sk), 63–64, 65
Normal curve model:
areas and ordinates, table of, 59, 249, 329–341
defined, 55
descriptive uses, 58
hypothesis testing and, 340–341
inferential uses, 58–59
item difficulty and, 249
nonnormal distributions and (see Nonnormal distributions)
percentiles, standard scores and, 99
population parameters, estimating, 59–62
properties of, 55–57
psychological testing and, 62
websites, 57–58
$z$ scores and, 94
Normalization, 62
Normative sample:
applicability of, 89–90
described, 87–88, 89
See also Reference group; Standardization sample
Norming:
  co-, 107
  simultaneous, 107–108
Norm-referenced testing, social
considerations and, 123–124
Norm-referenced test interpretation.
  See Score interpretation, norm-
  referenced
Norms:
  convenience, 90–91
  developmental, 83–85
  within-group (see Within-group
  norms)
  local, 90
  longitudinal changes in, 110–112
  subgroup, 90
Notation, basic, 54
Numbers, meaning of. See Meaning of
numbers
Observation, 294
Observed scores, 128
Occupational testing, 179, 181–182
OLSAT. See Otis-Lennon School Ability
Test (OLSAT)
One-tailed test, 341
Order, rank, 42
Ordinal scales, 42–43
  behavioral sequences and, 83
  limitations of, 45
  percentile rank (PR) scores and, 43
  rank order and, 42
  theory-based, 85
Ordinal Scales of Psychological
  Development, 85
Ordinate, 49
Organicity, 21
OSS. See United States Office of
  Strategic Services (OSS)
Otis, Arthur, 15, 37
Otis Group Intelligence Scale, 98
Otis-Lennon School Ability Indices, 98, 99
Otis-Lennon School Ability Test
  (OLSAT), 98
Overpredict, 214, 215, 216
Parallel forms, 106–107
Parameters, 59–60
  defined, 47–48
  estimating population, 59–60
  item
Path analysis, 71
Pearson, Karl, 70, 72
Pearson product-moment correlation
  coefficient (Pearson r), 72–78
  conditions necessary for use of,
  72–74
  deviations from linearity, 74–75
  other correlation methods and,
  76–78
  range restriction and correlation,
  75–76
  z scores and, 96
Percentage passing (p), 249
Percentage scores, 91
Percentile rank (PR) scores (percentiles):
  within-group norms and, 91–92
  normal curve and, 99
  rank order and, 42
Performance age, 85
Performance assessment:
  evaluation and scoring in, 117–118
  mastery testing, 118–119
  of personality, 21
Performance criteria, 83
Personal development, 24
Personality:
  assessment, 183–184
  inventories, 19–20
  performance-based measurements,
  244
  tests, 4
Personality Assessment Inventory (PAI), 98
Personality tests:
   ability tests vs., 4
   item difficulty and, 253
Personnel testing. See Employment, personnel testing and
PET. See Positron emission tomography (PET)
Phi, 77
Piaget, Jean, 85
Pie charts, 49
Placement, 221–222
Platykurtic distributions, 63
Point biserial correlation, 77
Polygons, frequency, 49, 51, 62
Polytomous variable. See Variables
Polytomous variables, 39
Population parameters, estimating, 59–60
Positron emission tomography (PET), 37
Power tests, pure, 257
PPE. See Pruebas Publicadas en Español (PPE)
Practice effects, 106, 145
Predictions:
   anchor point for, 293
   assessment and, 26
   correlations and, 71
   differential (see Differential prediction)
   level of probability and, 208
   overprediction and, 214
   questions, decisions and, 203
   regression analyses and, 70
   standard error of estimate and, 207
   test scores and, 199
   validity data and, 289
   validity evidence and, 223
Predictive validation, 202–204
   example of, 204–209
Predictors:
   ASVAB and, 224
   classification decisions and, 222
   criterion measures and, 199
   differential validity and, 213–214
   discriminant functions and, 223
   GRE scores and, 219, 220
   multiple, 211, 212
   nonetest, 221
   performance criteria and, 224
   personality dimensions and, 224
   placement decisions and, 222
   synthetic validation functions and, 223, 226
   validity generalization and, 226
Preparation, examiner, 304–306
Print resources, 33–34
   published tests, 33
   unpublished tests, 33–34
Probability theory, 47, 48, 55
PRO-ED, 33
Professional associations, 284
   See also specific association
Profile analysis, 159, 160
Program for International Student Assessment (PISA), 123
Project A, 224–226
Projection, 21
Projective tasks, 253
Projective techniques, 20–21, 244
Proportion passing (p), 249
Provence Birth-to-Three Developmental Profile, 83–85
Pruebas Publicadas en Español (PPE), 33
Psychiatric disorders, 202
Psychological assessment, 27
Psychological research, 23–24
Psychological testing:
   vs. assessment, 24–26
   continuous variables and, 40
developments in, 16–22
efficiency and, 286–287
Psychological testing: (Continued)
- error, estimates of, 40
- error in, sources of, 132–145
- evaluation of, 320
- forensic applications, 285
- history of, 8–15, 167
- meaning of numbers and, 44–47
- normal curve and, 62
- objectivity and, 287
- participants in the process, 7, 8
  as products (see specific assessment instrument)
- psychological assessment vs., 27
- qualifications for users of tests, 280
- shape of distributions and, 64–65
- standards for, 280–281
- usage, 22–24, 284–287
- usefulness of, 129
- validity and, 170–171, 172

Psychology, industrial-organizational, 219

*The Psychology of Assessment Centers*, 118

Psychometrics, 38, 234, 287

PsyclINFO database, 31

PsycTESTS, 30–31, 34

Published tests, 7, 33

Pure power tests, 257

Pure speed tests, 257

Purpose of test, 251–253

Q-interactive testing system, 307–308, 310

Qualification(s):
- forms, 30
  for users of psychological tests, 280

Qualitative item analysis, 233, 274

Quality of Life Inventory (QOLI), 198

Quantitative item analysis, 233, 273–274

Questionnaires, 3, 19, 21, 190, 195, 221, 242, 244, 253, 292, 320

Range, 52

Range restriction, correlation and, 75–76

Rank order, 42

Rapport, 241

Rasch model, 267

Rating scales, 117

Ratio(s):
  - critical, 61
  - selection, 289–290

Ratio IQs:
  - computations, attendant problems, 46
  - deviation IQs and, 99–100
  - MA/CA ratio and, 45–46

Ratio scales, 44

Raw scores, 81–82, 95
  - into stanine scores, 104
  - into z scores, transforming, 97–98
  - z scores and, formula for, 95

Ream, M. J., 19

Reasoning ability, 193

Recentering, 106–107

Reference group:
  - described, 88, 89
  - fixed, 106
  - test scores and, 105
  See also Normative sample; Standardization sample

Regression:
  - discovery of, 68, 69–70
  - toward the mean, 70
  See also Multiple regression equations

Regression analysis, item-test, 258–260

Regression line, 70, 206

Reliability:
  - alternate-form, 137
  - delayed alternate-form, 145
  - item response theory and, 151
  - relativity of, 130–131
sources of error, 132–145
split-half, 137–140
test interpretation and, 152–161
test–retest, 135
in test use, 127
validity and, 160–161
Reliability coefficient, 129
Representativeness, 178, 183
Response length, 246
Revisions, test. See Test revisions
Rice, Joseph, 16
Rights and Responsibilities of Test Takers: Guidelines and Expectations
Rorschach, Hermann, 20
Rorschach Comprehensive System, 310
Rorschach Performance Assessment System (R-PAS), 21
Rorschach technique, 20–21
Rubrics, 117, 245

Safeguarding tests. See Security, test
Sample, standardization, 87, 88, 89
Sampling:
distributions, 58–59, 60, 61
educational testing and, 179
problems, 63
See also Standardization sample
Sampling distributions, 58–59, 61
Sampling error:
content, 136–140, 185, 246
reliability coefficients and, 133
time, 134–135
SAS, 94
SAT:
creation of, 16–17
differential prediction and, 216
fixed reference groups and, 106
former name for, 239
IRT and, 270, 271
item analysis and, 239
recent changes to, 17
recentering and, 106–107
test item examples, 240
S-B. See Stanford-Binet Intelligence Scale (S-B)
Scale(s):
Attitudes Toward Women Scale (AWS), 184
Binet-Simon intelligence scales, 37
Bracken Basic Concept Scale, 198
Differential Ability Scales (DAS), 271
Differential Ability Scales-II, 100
Group Intelligence Scale, 15
Infant-Toddler Developmental Assessment (IDA), 83–84
interval, 43–44
nominal, 41–42
ordinal (see Ordinal scales)
Ordinal Scales of Psychological Development, 85
Otis Group Intelligence Scale, 98
rating, 117
ratio, 44
Stanford-Binet (see Stanford-Binet Intelligence Scale (S-B))
usage of term, 4
Wechsler (see Wechsler scales)
Scales, Norms, and Equivalent Scores, 105
Scaling, absolute, 250
Scatter, 52
Scatter diagram, 206
Scatterplots:
bivariate data and, 68, 69
item groups and, 250
Pearson r and, 73, 74, 75
Scholastic aptitude tests, 16–17
Schuhfried Company, 240
Scientific psychology, testing and, 11–12
Score(s):
age equivalent, 85
derived, 87
grade equivalent, 86
Score(s): (Continued)
within-group norms and, 91–103
interpretation of (see Score interpretation)
ipsative, 241
mental age (MA), 85–86
normalizing, 58
observed, 128
percentage, 91
percentile rank (see Percentile rank (PR) scores)
raw (see Raw scores)
reliability and, 130–131
standard (see Standard scores)
T, 98
t true (see True score(s))
Wechsler scale subtest, 98
z (see z scores)
Score interpretation:
norm-referenced, 82
raw scores (see Raw scores)
scoring and, 310–318
Score interpretation, criterion-referenced, 113–124
clinical assessment and, 122
content domains and, 115–117
frames of reference, relationship among, 120–122
item response theory and, 122–123
norm vs., 114
performance, predicting, 119–120
performance assessment and, 117–119
performance criteria and, 83
readings, selected, 115
social considerations and, 123–124
varieties of, 113–114
Score Reliability: Contemporary Thinking on Reliability Issues, 130
Scorer reliability, 134
Scoring:
administration and, 298–310
automated, 276
Scoring rubrics, 117
Screening, 221
SDMT. See Stanford Diagnostic Mathematics Test (SDMT)
SDRT. See Stanford Diagnostic Reading Test (SDRT)
SE. See Standard error (SE)
Seattle Longitudinal Study, 198
Security, test, 30, 109, 299, 304, 308, 309, 318, 319
Selected-response items, 240–244, 241–244
advantages of, 242
disadvantages of, 242–244
forced-choice, 241–244
Selection, test. See Test selection
Selection ratios, 289–290
Self-understanding, 24
SEM. See Structural equation modeling (SEM)
Semi-interquartile range, 52–53
Sentence completion tests, 21
Shrinkage, 212
Significance tests. See Statistical significance levels
Simon, Theodore, 13
SIOP. See Society for Industrial and Organizational Psychology (SIOP)
Situational tests, 118
Skewness (Sk): applied, 64
defined, 64
negative/positive distributions, 65
Society for Industrial and Organizational Psychology (SIOP), 282
Sophistication, test, 301
Sources of information about tests, 279–284, 295
Spearman, Charles, 18, 189
Spearman-Brown (S-B) formula, 138–140
Spearman’s rho ($r_s$) correlation coefficient:
  ordinal data and, 43
  ordinal scales and, 76–77
  See also Correlation coefficients

Specimen sets, 297

Speed of performance, 47, 81, 117, 189, 198, 257–258

Speed tests, pure, 257

Spiral-omnibus format, 138

SPSS, 94

SS. See Sum of squares (SS)

Standard deviation:
  described, 53
  normal curve model and, 55, 57

Standard error (SE):
  described, 59
  sampling distributions and, 61
  significance of, 61–62

Standard error of estimate (SEest), 207–209

Standard error of measurement (SEM), 152–155, 161
  CTT and, 265, 270, 271
  score levels/score accuracy and, 252

Standard error of the difference (SEdiff), 155–157, 160, 161, 340

Standard error of the mean (SEm), 60

Standardization sample, 87, 88, 89
  See also Normative sample; Reference group

Standardized achievement tests, 16

Standard normal distribution, 57

Standard scores, 102–103
  about, 93–96
  normal curve and, 99
  normalized, 102–103
  systems for deriving, additional, 96–100
  transforming $z$ scores into, 100–101
  $z$ scores and, formula for, 95
  See also $z$ scores

Standards for Educational and Psychological Testing. See Testing Standards

Stanford-Binet Intelligence Scale (S-B):
  early editions of, 98
  publication of, 110
  ratio IQs and, 45–46, 85
  validity and, 168

Stanford Diagnostic Mathematics Test (SDMT), 121

Stanford Diagnostic Reading Test (SDRT), 121

Stanines, 103, 104

States, 135

State-Trait Anxiety Inventory (STAI), 183–184

Statistical significance levels, 217, 218

Statistics, 47–55
  advice on, 39
  bivariate, 66
  descriptive (see Descriptive statistics)
  goodness-of-fit, 191
  inferential, 47, 61
  item validity, 254–256
  meanings of, 48
  multivariate, 66, 190
  univariate, 65

Stereotype threat, 214

Stern, William, 13

Stevens, S. S., 40

Strong Interest Inventory, 19

Strong Vocational Interest Blank (SVIB), 19, 20

Structural equation modeling (SEM), 190–192

Subcortical dementia, 204

Subgroup norms, 90
Sum of squares (SS), 53
SVIB. See Strong Vocational Interest Blank (SVIB)
Symmetrical, bilaterally, 56
Symptom Checklist-90-Revised (SCL-90-R), 184
Synthetic validation, 223
System of needs, 20

Table of specifications, 116–117
Task decomposition, 173
Taylor-Russell tables, 290

Technical Recommendations for Psychological Tests and Diagnostic Techniques, 171

Technique(s):
   Mantel-Haenszel, 275
   projective, 20–21, 244
   Rorschach, 20–21
Terman, Lewis, 14

Test(s):
   ability, 4, 247–248
   anchor, 106
   commercially available, 323–325
   design, universal, 273
   ethical use of (see Ethical use of tests) items (see Item analysis)
   Kaufman series of, 100
   misuse of, 283
   one-tailed, 341
   personality (see Personality tests) published, 7, 33
   pure power, 257
   reliability and, 130–131
   revisions (see Test revisions) sentence completion, 21
   situational, 118
   standard achievement, 16
   two-tailed, 340
   usage, 7

See also specific test

Test administration. See Administration, test
Test-age equivalents. See Age equivalent scores
Test ages. See Age equivalent scores
Test anxiety, 38, 185, 301–302
Test ceiling/floor, 92–93, 251

Test Critiques series, 33
Test difficulty levels, 251–252
Test information function, 270
Testing:
   clinical (see Clinical testing)
   computerized adaptive (see Computerized adaptive testing (CAT))
   content-referenced, 116
   domain-referenced, 116
   group, 14–15
   mastery, 118–119
   occupational, 179, 181–182
   psychological (see Psychological testing)

Testing service. See Educational Testing Service (ETS)

Testing Standards:
   assessment tools and, 292
   fairness and, 272
   guidelines and, 6
   major parts of, 280–281
   security of test materials, 319
   specimen sets and, 297
   test takers and, 282
   test user qualifications, 26, 27
   validity and, 165–166, 175, 227–228

Testing system, Q-interactive, 307–308, 310

Test items:
   constructed-response, 244–246
   developing and writing, 234–235
   selected-response, 240–244

See also Item analysis
Test of English as a Foreign Language (TOEFL), 180, 240
Test of Variables of Attention (TOVA)
Test purpose, 251–252
Test-retest reliability, 135
Test revisions, 109–112
Flynn effect and, 110–112
longitudinal changes in test norms, 110–112
norm-referenced tests and, 110
Tests, 33
Test score interpretation. See Score interpretation
Test selection, 284–298
decision to use/not use tests, 284–287
Tests in Print (TIP), 32, 33
Test taker preparation, 299–304
Test takers, IRT and, 265
Test user qualifications, 26–30, 279–280
levels of, 29
psychological testing vs. assessment, 27
Test utility, 287–292
in clinical decision making, 291–292, 293
Thematic Apperception Test (TAT), 21
Theory:
classical test (see Classical test theory (CTT))
decision, 287–288
generalizability (G theory), 149, 151
item response (see Item response theory (IRT))
probability, 47, 48, 55
Therapeutic model of assessment, 24
The Thirteenth Mental Measurements Yearbook, 121
Thorndike, E. L., 15, 16
Time sampling error, 134–135
content sampling error and, 145
reliability coefficients and, 133
TOEFL. See Test of English as a Foreign Language (TOEFL)
TOVA. See Test of Variables of Attention (TOVA)
Traits, 135
True score(s):
in group data, 129
as hypothetical entity, 128
in individual data, 128
Trustworthiness. See Reliability; Validity
T-scores, 98
Two-tailed test, 340
Type I/II errors, 217
Underpredict, 214, 215, 216
Unidimensional IRT models, 265
Unimodal, 56
United States Employment Service (USES), 215
United States Office of Strategic Services (OSS), 118
Univariate statistics, 65
Universal design, 273
Unpublished tests, print resources, 33–34
Usage, test, 7
U.S. Civil Service Examination, 9
U.S. Department of Defense, 109
U.S. Department of Education, 32, 282
User qualifications, 26–30
USES. See United States Employment Service (USES)
Utility, 228
Validation:
cross-, 19, 212, 237
predictive (see Predictive validation)
Validation studies, criterion-related, 209–221
Validity:
concurrent and predictive, 202–216
consequences, 226–228
Validity: (Continued)

consequential, 226
convergent, 196
criterion-related, 199–201
current perspectives on, 172–176
defined, 165, 168–172
differential, 222–223
discriminant, 196
evidence, sources of, 176–228
generalization (VG), 216, 217, 226 historial perspectives on, 167–172
incremental, 289
item (see Item validity)
judgments and, 166
patterns of convergence and differentiation, 194–198
references on, 167
reliability and, 160–161
synthetic, 223
tripartite view of, 171–172
utility and, 228
Validity coefficients, 207, 208
Validity data, 289
Validity generalization (VG), 216, 217, 226
Validity Generalization: A Critical Review, 217
Variability:
importance of, 53–54
measures of, 52–53
Variables:
continuous, 40
defined, 38
dichotomous, 39
discrete, 39
latent, 170, 187, 189, 190–191
moderator, 213
polytomous, 39
Variance, 53
Verbal Comprehension Index (VCI) score, 340, 341

Versant Spoken English Test (VSET), 181
VG. See Validity generalization (VG)
Vienna Test System (VTS), 240
Vocational counseling, 222
Vocational guidance, 18
VSET. See Versant Spoken English Test (VSET)
VTS. See Vienna Test System (VTS)

WAIS. See Wechsler Adult Intelligence Scale (WAIS)
WAIS-IV:
confirmatory factor analysis and, 192–193
percentile rank and, 93
standardization sample, 138–150, 186
WAIS-III and, 88
WAIS-IV: Administration and Scoring Manual, 341
WAIS-IV Technical and Interpretive Manual, 156, 157

Websites:
APA, 282
normal curve, 57–58
on statistics, 39
test publishers/distributors, 327
See also Internet
Wechsler, David, 99
Wechsler Adult Intelligence Scale (WAIS) deviation IQs and, 99
WAIS-III, 192
WAIS-IV (see WAIS-IV)
WAIS-R, 89, 223
Wechsler-Bellevue I, 99

Wechsler Intelligence Scale for Children (WISC):
Flynn effect and, 111–112
publication of, 99
revisions, 109
scale term and, 4
Wechsler intelligence scales, intervals between versions of, 239
Wechsler Memory Scale–Fourth Edition (WMS-IV), 93
Wechsler Preschool and Primary Scale of Intelligence (WPPSI), 239
Wechsler scales:
  deviation IQs and, 98
  IRT models and, 271
  validity and, 166, 185–186
  See also Full Scale IQs (FSIQs)
Wechsler scale subtest scores, 98
Wide Range Achievement Test, 271
Wiesen Test of Mechanical Aptitude (WTMA), 181
WISC. See Wechsler Intelligence Scale for Children
WISC-IV Integrated Manual, 112
Within-group norms, 86–91
  convenience, 90–91
  local, 90
  normative sample and, 87–88
  scores for expressing, 91–103
  subgroup, 90
WMS-IV. See Wechsler Memory Scale–Fourth Edition (WMS-IV)
Wonderlic Personnel Test, 138
Woodcock-Johnson tests, 107
Woodcock tests, 271
Woodworth Personal Data Sheet (P-D Sheet), 19
Working Memory Index (WMI) score, 340, 341
WorkKeys system, 182
WPPSI. See Wechsler Preschool and Primary Scale of Intelligence (WPPSI)
WTMA. See Wiesen Test of Mechanical Aptitude (WTMA)
Wundt, Wilhelm, 11
X-axis, 49
Y-axis, 49
Yerkes, Robert, 14
z scores:
  areas of curve and, 338–340
  derived standard scores and, 95
  linear transformation and, 94
  Pearson $r$ and, 96
  raw scores into, formula for, 95
  transforming into standard scores, 100–101
  transforming raw scores into, 97–98