

Name Index

- Abrahams, H., 136
 Ackerman, T., 150
 Adams, R. J., 74
 Albanese, M. A., 124
 Albers, C. J., 151
 Allen, N. L., 125
 Almond, R. G., 130, 136
 Angoff, W. H., 21

 Barksdale, J., 138
 Beaton, A. E., 125
 Bell, R., 148
 Bertin, J., 104
 Biancarosa, G., 135
 Billings, J., 87
 Bingham, W. V., 10
 Boeve, A. J., 151
 Bosker, R. J., 151
 Boughton, K. A., 122
 Brennan, R. L., 43, 55, 59, 106
 Brown, G. T. L., 39
 Brown, W., 48, 79
 Bulut, O., 64, 66

 Carvajal, J., 120, 121
 Case, S. M., 151
 Casella, G., 114
 Cepeda, N. J., 124
 Chang, Y.-F., 109
 Choi, I., 130, 152
 Clauser, A. L., 124
 Clauser, B. E., 106
 Cronbach, L. J., 48, 70–73, 76, 79, 113

 Dai, S., 71, 120–122, 152
 Davenport, E. C., 109

 Davison, M. L., 82, 109
 Deming, W. E., 88, 148, 149
 DiBello, L. V., 131
 Dillon, G. F., 106
 Dorans, N. J., 55, 57
 Draper, N. R., 126, 128
 DuBois, P., 6, 8
 Dwyer, A., 122

 Edwards, M. C., 120
 Everitt, B., 131

 Feinberg, R. A., xi, 21, 33, 41, 59, 106, 107,
 124, 127, 128, 150, 151
 Flanagan, J., 11, 12
 Fleiss, J. L., 77
 Friendly, M., 41

 George, A. C., 133
 Gessaroli, M., 21
 Gierl, M. J., 131, 132, 140
 Glutting, J. J., 117
 Gong, B., 22, 23
 Goodman, D. P., 19, 110, 122
 Goodman, L. A., 130
 Gross, J., 133
 Gulliksen, H., 12

 Haberman, S. J., x, xi, 45, 52, 57, 59–62, 65,
 67, 71–74, 82, 83, 85, 109, 111, 113–115,
 120, 121, 126–128, 133, 144, 150, 151
 Haladyna, T. M., 45, 81, 99, 102, 106–109,
 150, 153
 Hambleton, R. K., 19, 40, 122
 Hanson, B. A., 99, 150
 Harris, D. J., 99, 150

- Harrison, A. G., 19
 Hattie, J. C. A., 39
 Hegarty, M., 41
 Henson, R. A., 131
 Holland, P. W., 110, 142
 Hubbard, H. D., 3
 Huff, K., 110
 Hunka, S. M., 132
- Jenckes, T., 7
 Johnson, P., 17
 Jones, J. N., 117
 Junker, B. W., 132, 133
 Jurich, D. P., 59, 106, 151
- Kelley, T. L., 45, 46, 53–56, 58, 59, 62, 85, 114, 116
 Kibby, M. W., 125
 Kiefer, T., 133
 Kolstad, A., 126
 Kramer, G. A., 45, 81, 99, 102, 106–109, 150, 153
 Kruskal, W. H., 130
 Kush, J., 117
- LaFlair, G. T., 32
 Lane, S., 40, 151
 Larkin, K., 150
 Lee, W., 152
 Leighton, J. P., 131, 132
 Lewis, D., 122
 Liang, L., 44
 Lim, E., 152
 Ling, G., 150
 Link, H., 9
 Liu, Y., 120, 152
 Livingston, S. A., 40, 126
 Longabach, T., 150, 153
 Longford, N. T., x, 83, 128
 Loomis, S. C., 125
 Lord, F. M., 99, 127
 Lovett, B. J., 19
 Luecht, R., 130, 135
 Lumsden, J., 148
 Lyren, P., 150
- Manna, V., 120, 152
 Margolis, M. J., 106
 Marion, S., 22, 23
 McDermott, P. A., 117
 McFadden, D., 104
- Meijer, R. R., 151, 152
 Menard, S., 104
 Mertler, C. A., 22
 Mislevy, R. J., 130, 136
 Morey, L. C., 37
 Morrison, T., xii
 Muraki, E., 74
 Mussabini, S., 136, 149
- Novick, M. R., 99
 Nungester, R. J., 151
- O'Leary, T. M., 39
- Paolino, J., 104
 Papageorgiou, S., 130, 152
 Pashler, H., 124
 Pearson, K., 143
 Pendleton, G. H., 7
 Perie, M., 22, 23, 126
 Peyton, V. A., 150, 153
 Puhan, G., 44, 113, 120, 150
- Ramsay, J. O., 21
 Rasch, G., 127, 130
 Raymond, M. R., 106
 Reckase, M. D., 131
 Ree, M., 13
 Reynolds, C. R., 117
 Roberts, M. R., 140
 Robin, F., 120, 152
 Robinson, D. H., 81
 Robitzsch, A., 133
 Rohrer, D., 124
 Roussos, L. A., 131, 132
 Rupp, A. A., 131, 133
- Sands, W. A., 14
 Sawaki, Y., 108
 Senturk, D., 125
 Sheehan, K., ix, x
 Shu, Z., 150
 Sijtsma, K., 132, 133
 Silver, M., 40
 Sinharay, S., x, xi, 24, 43, 53, 59, 61, 66, 71, 73, 108, 113, 120, 121, 133, 135, 144, 150, 151
 Skorupski, W. P., 120, 121
 Slater, S., 40
 Smith, H., 126, 128
 Spearman, C., 45, 48, 63, 79

Spencer, B. D., 81	von Davier, M. J., 21, 107, 127, 128, 132, 133
Stanton, H. C., 117	Vue, K., 109
Steffen, M., 122	
Steinberg, L. S., 130	Wainer, H., ix, x, xi, 21, 33, 41–43, 48, 59, 68,
Stone, C. A., 121, 151	81, 107, 114, 120, 121, 140, 141, 150
Stout, W. F., 131	Walker, M. E., 55, 57
Su, S., 109	Wang, J., 125
Sumner, C., 7	Wang, X., x, 120–122, 152
Svetina, D., 120–122, 152	Watkins, M. W., 117
Swanson, D. B., 106, 107, 151	Wilson, K. M., 107, 150, 153
	Wingersky, M., 127
Tanaka, V., 134	Winward, M., 106
Tatsuoka, K. K., 132	Wixted, J. T., 124
Templin, J., 131, 133	
Tendeiro, J. N., 151	Xu, J. R., 131
Theil, H., 104	
Thissen, D., x, 120	Yao, L., 71, 72, 122
Thorndike, E. L., 9	Ye, F., 151
Thurstone, L. L., 10	Yen, W. M., 121
Tufte, E. R., 18, 41	Yerkes, R. M., 8, 9
Twain, M., 145	Yoo, H., 120, 152
Uenlue, A., 133	Zapata-Rivera, D., 40
	Zenisky, A. L., 40
VanWinkle, W., 40	Zhu, X., 151
Verdi, M., 21	Zieky, M. J., 126
Vevea, J. L., 120	Zwick, R., 40, 125

Subject Index

- χ^2 -type statistic, 121, 122
 36 data set survey, 88–91
- ACT report, 27
 actual subscores, 117, 120–122
 added value subscores. *See* value-added subscores
 adequate yearly progress (AYP), 14–15
 AFQT. *See* Armed Forces Qualification Test (AFQT)
 AGCT. *See* Army General Classification Test (AGCT)
 aggregate-level subscores, 49, 83–85
Alpha test, 10
 American Board of Internal Medicine Maintenance of Certification Examination (ABIM MOC), 27–30
 American Community Survey, 4
 Armed Forces Qualification Test (AFQT), 10
 score, 10–11
 Armed Services Vocational Aptitude Battery (ASVAB), 9, 10, 30
 compared with Census, 13–14
 performance of, 11–12
 reliability of, 12
 and scores, 10–11
 subtests of, 10–11
Army Alpha test, 9–10
Army Beta test, 9
 Army General Classification Test (AGCT), 10,
 See also *Army Alpha* test
 assessment
 critical thinking, 36–37
 engineering, 130
 formative, 22–23
 large-scale, 126
 subscores on educational, ix
 summative, 23, 43
 TIMSS, 152
 types, 22–23
 Assessment Tools for Teaching and Learning (asTTle), 39–40
 asTTle. *See* Assessment Tools for Teaching and Learning (asTTle)
 ASVAB. *See* Armed Services Vocational Aptitude Battery (ASVAB)
 augmented subscores, x, 46, 60, 61, 111
 correlations of, 63
 examinees, 123
 scores, 114
 language test and, 61–62
 PRMSEs of, 112
 rational argument in, 62
 reporting, 111–112, 114–121
 subscores vs. total scores, correlations, 62–63
 use of, 111
 validity of, 121
 VAR of, 61
 vs. objective performance index, 122–123
 average proportional reduction of mean square error, 64
 AYP. *See* adequate yearly progress (AYP)
- best linear predictor, 47, 62
 between-items multidimensional item-response models, 74
 Brennan utility index, 59
- canonical analysis, 47–48
 reliability coefficient, 68, 69
 true subscores

Subject Index

173

- correlations of, 68–69
 - prediction of, 67–69
- card-sorting test, 9
- category performance indicators, 128–130
- CCSE. *See* Comprehensive Clinical Science Examination (CCSE)
- ceteris paribus*, 1
- cognitive diagnostic model, 131–133
- Comprehensive Clinical Science Examination (CCSE), 33–34
- conditional subscores, 126–130
- construct definition, 81, 106
- contour lines, 103, 104
- correlation coefficient, 61, 62
- critical thinking assessment, 36–37
- Cronbach's alpha, 48, 70
 - approximation to reliability coefficient, 70–71
 - estimation, 70–71
 - standard errors, reliability coefficients, 71
- cut scores, 77–79
- diagnostic classification model, 131–133
- difference score, 51–52
- disattenuated correlation, 98–99, 105, 107, 108
- displaying data, principles, 41
- Duolingo English Test, 30–32
- Educational Testing Service (ETS), ix
- efficacy, test scores, 11
- eigenvalue, 66
- eigenvector, 66
- Elementary Education: Content Knowledge test, 113–116, 121, 122
- ETS. *See* Educational Testing Service (ETS)
- Every Student Succeeds Act, 110
- evidence-centered design, 130, 136
- Examination a*, 9, *See also* Army Alpha test
- extrapolations, 3–4
- fitted quadratic function, 2–3
- formative assessments, 22–23
 - asTTle report, 39–40
 - focus of, 23
 - SC Ready, 32–33
 - types of, 22–23
- generalized partial credit model, 74–75
- group subscore mean, 84–85
- group test, intelligence, 8
 - half test, 72–73
 - heterogeneity, 144
- incremental validity, 47–49, 81–83
- intelligence, group test, 8
- isarithms, 104
- item mapping, 125–126
- item-level feedback, 124–125
- item-response theory, 73–74, 86
 - cut scores, 77–79
 - latent variable and true subscore by, 75
 - multidimensional models, 74–76, 86
 - scaled subscores, 76–77
 - true transformed subscore by, 76
- job analysis, 106
- Kelley estimation, 53–55
 - application of, 53
 - improvement from, 54
 - of total score, 55, 56
 - of writing score, 55, 56
- language test
 - augmented subscores and, 61–63
 - correlations
 - subscores vs. total scores, 62–63
 - true subscores vs. total scores, 63–64
 - data, 49–50
 - eigenvalue in, 66
 - estimation, 70
 - half-tests in, 72
 - PRMSE
 - canonical score results, 68–69
 - and partial regression coefficients, 62
 - and VAR, 59
 - PRMSE_{res} in, 65
 - reliability of cut scores, 78
 - standard errors, reliability coefficients, 71
 - structure of, 50
- large-scale assessments, 126
- latent variables prediction, 74–76
- linear predictor, 68
- linear regression, 126
- logical evidence, 106
- logistic regression model, 103–105
- maximum-likelihood estimation, 75
- mean square error (MSE), 46, 47
- meaningful subscores, 5
- measurement error (S_e), 51–52, 72, 73
- mirt* program, 73, 75, 77

- MSE. *See* mean square error (MSE)
- multidimensional item-response models, 74–76, 86, 133–134
- NAEP. *See* National Assessment of Educational Progress (NAEP)
- National Adult Literacy Survey, 126
- National Assessment of Educational Progress (NAEP), 23, 126
- National Board of Medical Examiners (NBME), x, 151
- National Council on Measurement in Education (NCME), xi
- NBME. *See* National Board of Medical Examiners (NBME)
- NCME. *See* National Council on Measurement in Education (NCME)
- New York State Testing Program (NYSTP), 41
- No Child Left Behind Act, ix, 14
- non-overlapping subscores, 108
- normal approximation case, 76, 77
- norming group, 22
- NYSTP. *See* New York State Testing Program (NYSTP)
- objective performance index, 121–124
- orthogonality, 6
- PAI. *See* Personality Assessment Inventory (PAI)
- parallel-form subscore, 50, 51
- PRMSE estimation of, 58
- parallel-form total score, 50, 51
- partial correlation coefficient, 61
- partial regression coefficient, 61–62
- Pearson Test of English Academic*, 131
- Personality Assessment Inventory (PAI), 37–38
- practice analysis, 106
- practitioners recommendations, 138–140
- Praxis test, 24–25
- principal-components analysis, 47, 65–67
- true subscores with, 67
- PRMSE. *See* proportional reduction in mean square error (PRMSE)
- PRMSE_{aug}, 47, 61, 62
- PRMSE_{irt}, 75–77
- PRMSE_{psub}, 54, 57
- PRMSE_{punb}, 54
- PRMSE_{res}, 65
- PRMSE_{sub}, 46–47, 79–80
- PRMSE_{tot}, 46–47, 59, 64, 113
- PRMSE_{unb}, 54
- PRMSE_{waug}, 48, 68
- procedural evidence, 106
- profile, 47–48
- profile analysis, ix, 117
- profile bands, 20, 21
- prophecy formula, 79
- proportional reduction in mean square error (PRMSE), 46–47, 108, 139
- of augmented subscores, 112
- available data for, 150–153
- language test
- and partial regression coefficients, 62
- and VAR, 59
- for parallel-form scores estimation, 58
- strength of, 150
- from true subscores, 53
- conditionally unbiased property/estimation, 54–55
- unavailable data for, 153
- psychometric quality, 19, 24, 40
- psychometric-based approach, 131
- reading tests, value-added subscores, 107
- receptive subscore, 52
- reliability, 1, 6
- ASVAB scores, 12
- coefficients, 45–46, 50–53, 79
- canonical analysis, 68, 69
- Cronbach's approximation to, 70–71
- half-sample estimation, 73
- incremental validity in, 48–49
- for numbers of items, 79–81
- of residual subscores, 83
- satisfactory, 46
- standard errors estimation, 71
- variances and, 52
- and correlation requirements, 99–103
- cut scores in language test, 78
- estimation of, 48
- on National Board of Medical Examiners, 151
- of residual subscores, 82
- subscore, 12–13, 86
- report, 45–46
- test scores, 12
- repeated test, 72–73
- residual subscores, 47, 63–65
- reliability, 82
- residual true score, 64
- residual-based indicators, 128–129
- response percent/probability (RP), 125

- rhythm test, 10
- rudimentary proficiency testing, 6
- SAT report, 25–27
- scale anchoring process, 125
- scaled subscores, 76–77
- scatter analysis, ix
- score report, 140–141, *See also* subscores
 - ABIM MOC, 27–30
 - ACT, 27
 - assessment types, 22–23
 - asTTle report, 39–40
 - ASVAB, 30
 - categorical performance indicators, 20, 21
 - CCSE, 33–34
 - complexity of, 18–20
 - comprehensive framework in, 40
 - critical thinking assessment, 36–37
 - Duolingo English Test, 30–32
 - empathy in communication of results, 41–42
 - error representation, 142–143
 - ordering, 141–142
 - PAI Law Enforcement, Corrections, and Public Safety report, 37–38
 - percent correct, 20–21
 - percentile score, 20, 21
 - practices on, 40–42
 - Praxis test, 24–25
 - profile bands, 20, 21
 - raw score, 20, 44
 - SAT, 25–27
 - SC Ready, 32–33
 - user types, 23–24
 - USMLE, 34–36
- scores, history of, 6–8
- signal-to-noise (SNR), 55
- SNR. *See* signal-to-noise (SNR)
- South Carolina College-and Career-Ready Assessments (SC Ready), 32–33
- Spearman–Brown function, 2, 147, 148
 - prophecy formula, 79
- specific subgroups, 144
- specificity, 6, 13–14
- standard deviations (SD), 115
- Stanford–Binet intelligence test, 8
- standardized value, 58
- step-by-step process, 138
- stratified alpha, 48, *See also* Cronbach’s alpha
- subscores
 - aggregate-level, 49, 83–85
 - augmented, x, 46, 60, 61
 - average reliability, 95–98
 - cases of, 5–6
 - chance of adding value, 103–105
 - characteristics, 6
 - conditional, 126–130
 - on educational assessments, ix
 - examinees, 123
 - score, 115–116
 - gain value, 145–146
 - history of, 6–8
 - incremental validity in, 48–49
 - information. *See* score report
 - insisting examinee pass on, 146–147
 - involvement with, ix–xi
 - limited resources, 147–148
 - meaningful, 5
 - non-overlapping, 108
 - open-source software packages, 71
 - orthogonality, 6
 - outliers, reporting, 126–130
 - parallel-form, 50, 51
 - psychometric-based approach, 131
 - receptive, 52
 - reliability, 6, 12–13, 86
 - report complexity, 18–20
 - report on practices, 40–42
 - residual, 47, 63–65
 - scaled, 76–77
 - specificity, 6, 13–14
 - testing programs by, 14–15
 - unreporting, 124
 - unusual, 144
 - used as prod, 146
 - values, 20–22
 - variance of, 51
 - vs. total score prediction, 60–63
 - Wainer-augmented, 68
 - weak information, 144–145
- subtests
 - of ASVAB, 10–11
 - history of, 6–8
 - length of, 92–95
 - mean score and reliability, 113
- summative assessments, 23, 43
 - ACT, 27
 - ASVAB report, 30
 - Duolingo English Test, 30–32
 - SAT, 25–27
- Swedish Scholastic Aptitude Test, 105
- TerraNova test, 122
- test developer, 106–107

- test development process, 106
- test length, 81
 - effect on subscore, 79–81
 - modification, 48
- Test of English for International
 - Communication (TOEIC) test, 153
- test publishers, 23, 24, 44, 111
- test scores, 1, *See also* reliability, coefficients
 - efficacy, 11
 - reliability, 12
- test specifications, 40
- test takers, 72, 83–84, 86
- testing organizations, ix, 48, 49, 83, 85–86
- testing programs. *See also specific test programs*
 - subscores by, 14–15
- tests, history of, 6–8
- TIMSS assessment. *See* Trends in
 - International Mathematics and Science Study (TIMSS) assessment
- TOEFL, 83, 108
- TOEFL Primary® test, 130, 152
- TOEIC test. *See* Test of English for International Communication (TOEIC) test
- total scores
 - alternatives approach, 65
 - canonical analysis, 67–69
 - principal-components analysis, 65–67
 - vs. subscore prediction, 60–63
- Trends in International Mathematics and
 - Science Study (TIMSS) assessment, 152
- true score, 45–46, 48, 52, 63, 69, 72, *See also*
 - true subscores
- predictors of, 74, 86
 - principal-components with, 67
 - profiles, 64
 - residual, 64
 - variances of, 73
- true subscores
 - canonical analysis prediction, 67–69
 - conditionally unbiased property/estimation, 54
 - PRMSE for, 54–55
 - by item-response theory, 75
 - Kelley estimation, 53–55
 - prediction, 53–58
 - options for, 46–47
 - residual, 47
 - total score vs. subscore, 60–63
 - predictors of, 74
 - PRMSE from, 53
 - scree plot of, 67
 - vs. total scores, 47–48
 - correlations of, 64
 - prediction, 58–60
- true-score theory, 48
- two-category transformation case, 76
- US, mental testing in, 8–10
- unidimensional item response model, 121–122
- United States Medical Licensing Examination (USMLE), 34–36, 151
- unreliable subscore, 86, *See also* reliability, subscore
- unusual subscores, 144
- US census
 - decennial census, 2–5
 - extrapolations, 3–4
 - lessons, 4
 - population, factors of, 4
- user types, subscore, 23–24
- validity
 - augmented subscores, 121
 - evidence, 106, 109
 - external criterion, 82
 - incremental, 81–83
- value-added ratio (VAR), 46, 59, 60, 108, 139
 - augmented subscores of, 61
 - available data for, 150–153
 - on National Board of Medical Examiners, 151
- value-added subscores
 - in practice, 91–92
 - reading tests, 107
 - uniqueness of, 105–107
- VAR. *See* value-added ratio (VAR)
- variances
 - half-sample estimation, 73
 - of measurement, 73
 - and reliability coefficients, 52
 - of subscores, 51
- Wainer-augmented subscore, 68
- weighting, 21