

Cambridge University Press

978-0-521-68126-1 - Testlet Response Theory and Its Applications

Howard Wainer, Eric T. Bradlow and Xiaohui Wang

Index

[More information](#)


---

## Author Index

---

- Aitkin, M., 40, 84, 99  
 Albert, J. H., 113  
 Algina, J., 208  
 Anastasi, A., 105  
 Andersen, E. B., 27, 120  
 Angoff, W. H., 207  
  
 Baker, E. L., 144  
 Berstrom, B. A., 190  
 Birnbaum, A., 26, 77, 134, 239  
 Bock, R. D., 7, 32, 40, 64, 65, 78, 80, 84, 99,  
   112, 120, 183, 194, 242  
 Boulet, J. R., 145, 211  
 Bradlow, E. T., 8, 113, 116, 118, 121, 123,  
   128, 131, 137, 140, 145, 149, 190, 207, 235  
 Brennan, R. L., 208  
 Breslow, N. E., 231  
 Brown, W., 18, 75  
  
 Carlin, J. B., 234, 238  
 Carr, A., 50  
 Chang, H.-H., 91  
 Chib, S., 113, 242  
 Clauser, B. E., 221  
 Cohen, J., 208  
 Congdon, P., 237  
 Cook, L. L., 47  
 Coughlan, A. T., 20  
 Cronbach, L. J., 19  
 Crumm, W. L., 75  
  
 Dempster, A. P., 112, 198, 242  
 Donoghue, J. R., 221  
 Dorans, N. J., 47, 78, 96  
 Douglas, G. A., 27  
 Du, Z., 137, 140  
 Duhachek, A., 20  
 Dunbar, S. B., 144  
  
 Earles, J. A., 9  
  
 Efron, B., 98, 122  
 Eignor, D. R., 47  
 Embretson, S. E., 183  
  
 Ferguson, T. S., 237  
 Fisher, R. A., 210  
 Folske, J. C., 110  
  
 Gelfand, A. E., 112, 235  
 Gelman, A., 112, 123, 125, 126, 127, 137,  
   155, 216, 234, 238, 243, 244  
 Gessaroli, M. E., 110  
 Geyer, C. J., 126  
 Ghosh, B. K., 209  
 Gibbons, R. D., 65, 111, 112, 130  
 Gilks, W., 127, 234, 243  
 Green, B. F., 51, 76  
 Greenberg, E., 242  
 Greenland, S., 231  
 Guilford, J. P., 75  
 Gulliksen, H. O., 7, 73  
 Gustafsson, J.-E., 39, 121  
  
 Haenszel, W., 220  
 Hambleton, R. K., 50, 208, 221  
 Hanson, B. A., 208  
 Hastings, R., 112  
 Hedeker, D. R., 111, 130  
 Holland, P. W., 25, 27, 78, 81, 83, 154, 220,  
   221, 225, 231  
 Holzinger, K. J., 75, 111  
 Humphreys, L. G., 79  
 Huynh, H., 208  
  
 Iacobucci, D., 20  
  
 Jöreskog, K. G., 111  
  
 Kelley, T. L., 20, 75  
 Kendall, M. G., 80

Cambridge University Press

978-0-521-68126-1 - Testlet Response Theory and Its Applications

Howard Wainer, Eric T. Bradlow and Xiaohui Wang

Index

[More information](#)

262

*Author Index*

- Kiely, G. L., 75, 106  
 Kingsbury, G. G., 46  
 Kingston, N. M., 47, 62  
 Kolakowski, R., 7, 183  
 Kolen, M. J., 208
- Laird, N. M., 112, 198, 242  
 Lee, W.-C., 208, 210  
 Levine, M., 40  
 Lewis, C., 67, 74, 75, 80, 208  
 Lichten, W., 9  
 Linn, R. E., 144  
 Little, R. J. A., 39, 188, 189, 193, 197, 198  
 Livingston, S. A., 208  
 Loevinger, J., 24  
 Lord, F. M., 7, 15, 18, 21, 26, 27, 30, 44, 46, 76, 77, 223  
 Lukhele, R., 7, 144  
 Lunz, M. E., 190
- MacNicol, K., 50  
 Mantel, N., 220  
 Mazor, K. M., 221  
 Mazzeo, J., 91  
 McBride, J. R., 53  
 McDonald, R. P., 24, 111  
 Merrill, P. F., 50  
 Messick, S., 15  
 Metropolis, N., 112  
 Mislevy, R. J., 31, 32, 40, 41, 194, 242  
 Mollenkopf, W. G., 50  
 Monk, J. J., 50  
 Mooney, J. A., 34, 64, 74, 75, 77  
 Morgan, A., 39, 121  
 Morris, C. N., 98, 122, 237  
 Muraki, E., 65, 112
- Neyman, J., 81  
 Novick, M. R., 7, 15, 18, 21, 26, 30, 44, 76, 77, 208
- Pearson, E. S., 81  
 Phillips, A., 231
- Ramsay, J. O., 216  
 Rasch, G., 26, 27, 239  
 Reckase, M. D., 24  
 Ree, M. J., 9  
 Richardson, S., 234  
 Roberts, G., 127, 243  
 Robins, J., 231  
 Rogers, H. J., 78, 221  
 Rosenbaum, P. R., 34, 64  
 Rowley, G. L., 76  
 Rubin, D. B., 39, 112, 126, 137, 188, 189, 193, 197, 198, 234, 238, 242, 244
- Samejima, F., 64, 82, 114, 146, 156, 211  
 Sands, W. A., 53  
 Sax, G., 50  
 Segall, D. O., 24  
 Shealy, R., 78  
 Sheehan, K., 80  
 Silverman, B. W., 216  
 Sinharay, S., 126, 244  
 Sireci, S. G., 74, 75, 79, 82, 144  
 Smith, A. F. M., 112, 124, 235  
 Sörbom, D., 111  
 Spearman, C., 18  
 Spiegelhalter, D. J., 234  
 Stallings, W. M., 50  
 Stanley, J. C., 87  
 Stein, C., 21  
 Steinberg, L., 34, 35, 64, 74, 75, 77, 81, 223  
 Stern, H. S., 234, 238  
 Stocking, M. L., 41  
 Stout, W. F., 78, 130  
 Stuart, A., 80  
 Subkoviak, M., 208  
 Swaminathan, H., 78, 208, 221  
 Swineford, F., 111
- Tanner, M. A., 112  
 Thayer, D. T., 78, 83, 220, 221  
 Thissen, D., 5, 7, 34, 35, 64, 74, 75, 77, 78, 79, 81, 82, 106, 144, 190, 194, 223  
 Thomas, N., 190  
 Thomson, G. H., 75  
 Thorndike, R. L., 75  
 Thurstone, L. L., 110  
 Towle, N. J., 50  
 Traub, R. E., 76
- Ulan, S., 112
- Wainer, H., 5, 7, 8, 9, 10, 11, 31, 35, 39, 46, 48, 51, 53, 74, 75, 78, 79, 81, 82, 95, 106, 107, 113, 116, 118, 121, 123, 128, 131, 137, 140, 144, 145, 149, 154, 190, 207, 216, 220, 223, 225, 235  
 Wald, A., 209  
 Wang, M. D., 87  
 Wang, X., 113, 116, 118, 123, 128, 131, 145, 149, 235  
 Waters, B. K., 53  
 Weiss, D. J., 46  
 Wong, W. H., 112  
 Wright, B. D., 27, 190
- Yen, W. M., 73, 106
- Zenisky, A. L., 144  
 Zhang, J., 130  
 Zwick, R., 83

---

## Subject Index

---

- 1-PL (one-parameter logistic) testlet model, 28, 29, 189, 190, 240
  - and item characteristic curves (ICCs), 28
- 2-PL testlet model, 30, 119, 126, 190, 197, 222, 239, 240
  - 2-PL Bayesian item response theory (IRT) model parameters, 118, 125
  - computational feature of, 137
  - dichotomous form of, 123
  - ICCs of, 135
  - prior distributions of, 161
  - testlet response theory (TRT) model, 120, 160
  - TRT likelihood approach, 123
- 3-PL testlet model, 30, 31, 36, 39, 40, 114, 115, 135, 142, 189, 190, 239, 240
  - 2-PL testlet model as special case of, 135
  - 3-PL Bayesian testlet model, 134
  - analogous version of, 136
  - and 3-PL TRT model, 222
  - and binary items, 145
  - for binary items, 156
  - item characteristic curves (ICCs) of, 135
  - Metropolis-Hastings algorithm, 137
  - testlet variance for, 139
  - weak identifiability of, 155
- algebra testlet, 68
  - algorithm for, 107
  - effects of context in, 106
- ALL DIFFERENTIAL item functioning (DIF) model, 92, 93
- Appreciation of Life testlet, 176
- Armed Services Vocational Aptitude Battery (ASVAB), 11, 53
  - Computerized Adaptive Test version of (CAT-ASVAB), 11, 53, 63
  - and traditional item response theory (IRT), 53
- asymptotic standard errors/theory, 122, 124, 142, 191, 231, 234
- Bayes factors, 237
- Bayes modal estimation, 32, 34, 35, 36, 38, 40
  - empirical methods of, 122
- Bayes theorem, 161
  - application of, 191
- Bayesian 2-PL testlet model, 127, 138
  - and absolute error prediction, 129
  - and coverage probabilities, 129
  - and Markov chain Monte Carlo (MCMC) methods, 130
  - efficacy of, 127
  - item characteristic curve (ICC), 134
  - limitations of, 130
  - parameters in, 126
  - simulation design within
    - and testlet response theory (TRT) model, 138
  - MML estimation approaches in, 120
- Bayesian 3-PL testlet model
  - and application on Graduate Record Exam (GRE), 140
  - and its use in mixed data, 146
  - computation aspect of
    - Markov chain Monte Carlo (MCMC) method, 137
  - parameters with true values for, 139
  - Scholastic Assessment Test (SAT), 140
  - simulation design for, 137, 138
- Bayesian estimation, 123, 166, 191, 192, 236.
  - See also* Bayes modal estimation
  - and mean squared errors, 199
  - ignorable missing pattern in, 193
  - Markov chain Monte Carlo method, 123
  - posterior distribution in, 193
  - underlying aspects of, 235
- Bayesian framework, 113, 115, 119, 169, 184
  - uses of, 161

- Bayesian hierarchical specification, 119, 222, 238
- Bayesian item response theory (IRT) model, 236, 241, 242, 243, 245  
 computation for, 241  
 convergence issues for, 244  
 uses of, 244
- Bayesian mixed testlet model, 151. *See also* Bayesian 2-PL testlet model; 3-PL testlet model  
 and testlet design variables, 148  
 application of, 147, 148  
 parameters from, 149  
 simulation design for, 148  
 simulation results for, 148
- Bayesian testlet model, 98, 145  
 covariate augmented analysis result in, 162  
 inferential goal in, 241  
 limitation of, 155  
 parameters in, 155  
 specification of, 136
- Bayesian testlet response theory (TRT) model*, 105, 119, 219, 223, 226  
 covariates insertion/utilization in, 155  
 extended version of, 155
- Bayesian theorem, 113, 121, 207  
 advantage of, 198  
 consequence of, 113
- Bernoulli likelihood approach, 121, 189
- beta-binomial test scoring models, 208
- bi-factor model, 110, 111, 112  
 goal of, 110
- BILOG software/computer program, 32, 129, 137, 194, 226
- binary 3-PL testlet model, 147
- binary/polytomous data  
 Bayesian testlet model for, 144, 145  
 psychometric scoring model for, 144
- Bock's model, 64, 82  
 item factor analysis in, 64  
 latent proficiency in, 64
- breast cancer patients, survey of, 169  
 and testlet response theory, 183  
 Bayesian analysis in, 178, 181  
 Bayesian framework, 173  
 covariates associated with, 176  
 demographic/background questions in, 171  
 drug intake (Tamoxifen), 172  
 included survey items, 171  
 item response theory (IRT) model, 173, 181  
 local dependence effect of, 181  
 post-MCMC regression analyses, 177  
 probability curves, 173. *See* traelines  
 regression analysis in, 177  
 stem-and-leaf diagram, 176  
 test-scoring model in, 173
- classification consistency, test scores, 207  
 and use of Rasch model, 208  
 assessment of, 208  
 psychometric scoring model for, 208
- Clinical Skills Assessment (CSA)  
 Integrated Clinical Encounter (ICE) score of, 211
- Cohen's coefficient (k), 208
- computerized adaptive test (CAT), 10, 44, 46, 47, 107, 190  
 Ad Hoc testlet construction, 107  
 and context effects, 51  
 and instructional diagnosis, 10  
 and true score theory, 10  
 item selection algorithm, 51, 52  
 reading comprehension items, 63  
 scoring models in, 10  
 standard error of estimates in, 52  
 testing algorithm for, 10
- conditional maximum likelihood (CML), 27, 39
- construct validity concept, 15
- context effect, 47, 48  
 definition of, 47
- Cronbach's coefficient ( $\alpha$ ), 18
- Cronbach's statistic, 19, 20
- cumulative density function (CDF), 115
- data augmentation approaches, 124, 137
- DIFferential item functioning (DIF)  
 analysis/study, 79, 219  
 and Mantel-Haenszel (MH) statistics, 87  
 and simulations interpretation, 231  
 and true score curves, 87, 95  
 approach for, 220
- DIFferential item functioning (*cont.*)  
 Bayesian approach for, 219, 221, 224, 231  
 detection of, 87  
 testlet response theory based, 221  
 Holland's Rule for, 225  
 practical procedure for, 225  
 results of, 225  
 testlet response theory (TRT) model in  
 characteristics of, 222  
 use of Mantel-Haenszel (MH) statistics in, 220
- differential test functioning (DTF), 154
- Educational Testing Service (ETS), 11, 61
- empirical Bayes methods, 98, 99, 122
- error variance, 16, 21
- examinee's proficiency  
 estimation of, 37, 38
- Fisher information function/matrix, 37, 122
- Gaussian error distribution, 22, 122, 216, 219

- parameters drawn from, 139
- Gibbs sampling, 234
- Graduate Management Aptitude Test (GMAT), 46
- Graduate Record Exam (GRE), 46, 137
  - verbal section of (GRE-V), 141
    - non-testlet based methods for, 141
    - parameter estimates for, 141
    - structure of, 140
    - testlet-based methods for, 141
- Hebrew exam, 9
- Hessian matrix calculations, 234
- Holland's Rule, 225
- instructional diagnosis, 10
- Integrated Clinical Encounter (ICE) score, 211
- item characteristic curves (ICCs), 28, 29, 134
- item difficulty parameter (b), 159, 221
  - covariates of, 162
  - kernel density estimate of, 163
- item discrimination parameter (a), 159, 189, 194
  - covariates of, 162
  - kernel density estimate of, 163
- item response functions (IRFs), 28, 78
- item response theory (IRT), 10, 22, 24, 49, 53, 57, 70
  - and index of reliability, 77
  - development of, 105
  - key feature of, 105
  - models supporting IRT theory, 27, 30, 70, 83, 118, 126, 135, 189, 197, 210, 211, 221, 235
  - 3-PL testlet model as, 31
  - characteristics of, 72
  - conditional independence of, 34
  - conditional maximum likelihood (CML), 27
  - dichotomous form of, 81, 90
  - joint maximum likelihood (JML), 27
  - likelihood-based, 234
  - parameters for, 27, 192, 239
  - parametric form of, 234
  - polytomous form of, 80, 84, 100, 107, 211
  - probability-in-the-soul idea of, 26
  - Rasch model as, 24
  - statistics of, 74
  - unconditional maximum likelihood (UML), 27
  - use of Bayesian framework, 211
  - use of MCMC methods, 124
- item response theory (IRT) model, 182
- item vs. question, 5
- joint maximum likelihood (JML), 27, 39, 120
- judge-to-judge, score variance, 8
- kernel density estimation, 164, 165, 166, 167, 168
  - results of, 196
- Law School Admissions Test (LSAT), 73, 74, 79, 85, 88, 97, 98
  - forms of, 80
  - parallel forms of, 84
  - sections in, 73
  - statistical characteristics of, 98
  - structure of, 74
  - subgroups on, 84
  - testlet structure of, 86
- logistic model*, 30
- Mantel-Haenszel (MH) statistics, 87, 88, 219
- marginal reliability, test scoring, 77
- Markov chain Monte Carlo method, 112, 113, 123, 211, 235, 241
- maximum a posteriori (MAP) estimation, 123
- maximum likelihood estimation, 32, 34, 191, 192, 194, 198, 236
  - error variance of, 37
  - goal of, 192
  - programs for, 32
  - standard errors for estimates by, 124
- maximum marginal likelihood method (MML), 27, 39, 40, 84, 99, 120
  - estimation of, 40
- MCMC methods, 123. *See* Markov chain Monte Carlo method
  - absolute errors for, 129
  - efficacy of, 148
- Metropolis-Hastings algorithm, 124, 126, 127, 137, 147, 242, 243
- Missing At Random (MAR), 188
  - definition of, 190
  - description of, 188
  - missing data element, 197, 198
- Missing Completely At Random (MCAR), 188, 189, 194, 197
  - definition of, 190
  - description of, 188
- missing data set
  - and SCORIGHT analysis, 197, 199, 200. *See also* SCORIGHT software/computer program
- missingness indicator (M), test data set
  - distribution of, 188, 189
- MML estimation/method, 27, 39, 40, 84, 99, 120, 122. *See also* maximum marginal likelihood method
  - asymptotic standard errors/theory in, 120, 122
  - second stage of, 121
- MULTILOG software, 32, 194
- multivariate normal random variable, 156

- National Council of Architectural Registration Boards (NCARB), 61, 63  
 decision making in, 70
- National Merit Scholars  
 screening for, 6
- Newton-Raphson technique, 36, 121, 123
- Neyman-Pearson lemma  
 test by, 81
- NO DIF model, 92, 93. *See also* Differential item functioning
- no testlet effect, 128, 129, 134, 138
- Nobel Prize winners  
 algorithm used to choose, 6
- North Carolina Test of Computer Skills, 148, 150  
 Bayesian mixed testlet model to application of, 150  
 results from, 150  
 performance sections of, 151
- Not Missing At Random (NMAR), 188, 189, 190
- options word count, covariate, 158, 159, 163
- ordinal response model, 114
- passing scores  
 description of, 207  
 posterior distribution evaluation of, 207
- polytomous IRT model, 72  
 graphical description of, 146  
 parameter for, 157
- Posterior Probability of Passing (PPoP) curve, 209  
 and application to Clinical Skills Assessment, 209  
 and maximum likelihood theory, 209  
 Bayesian approach to construct, 216  
 construction of, 213  
 determination of, 210  
 idea behind, 209  
 Integrated Clinical Encounter (ICE) score from, 212  
 parallelism in, 215  
 posterior standard deviation, 197  
 kernel density estimation for, 197
- probability-in-the-soul approach, 25, 26, 27
- psychometric scoring model, 57, 144
- Rasch model, 83, 107
- reliability coefficient, 21
- reliability, test scoring, 18, 20  
 and error of estimation, 77  
 calculation of, 76  
 definition of, 76, 77
- Rosenbaum's theorem, 54
- Scholastic Assessment Test (SAT), 136, 140  
 summary results from, 9  
 verbal section of (SAT-V), 140, 142  
 non-testlet based methods for, 141  
 structure of, 141  
 testlet-based methods for, 141
- score variance, 8, 16
- SCORIGHT software/computer program, 116, 123, 127, 131, 150, 161, 211, 221, 223, 235, 241  
 and missingness patterns, 193  
 MCMC methods in, 149  
 use of Bayesian theorem in, 194
- Simulation Studies, Differential item functioning (DIF) detection, 228
- Spearman-Brown Formula, 18, 75, 87
- split-half reliability, 75. *See also* reliability, test scoring
- Standards for Educational and Psychological Tests, 75
- stem word count, covariate, 158, 159, 163
- stem-and-leaf diagram, 176
- test data set  
 and data matrix for examinee's response, 186  
 and missingness indicator (M), 187  
 and SCORIGHT analysis, 187, 195, 197  
 fractional design in, 189
- Missing At Random (MAR) in, 187
- Missing Completely At Random (MCAR) in, 187
- missingness patterns of, 187, 188, 189  
 model for, 199
- Not Missing At Random (NMAR) in, 188
- test scores/scoring, 14  
 accuracy measurement, 14  
 and reliability, 15  
 Bayesian methods, 216  
 coherent psychometric scoring model, 106  
 conditional reliability of, 208  
 educational system efficacy and, 7  
 error variance in, 76  
 fineness in, 6  
 models for, 186  
 multiple choice item for, 7  
 split-half reliability, 21  
 statistical models for, 15
- testing, 3  
 allocation of points in, 5  
 and education, 9  
 and item response theory, 10  
 and licensing, 9  
 and performance of candidates, 3  
 and question of primary interest, 4  
 and test score, 3  
 and theory of testlets, 3  
 computerized form of, 9

Cambridge University Press

978-0-521-68126-1 - Testlet Response Theory and Its Applications

Howard Wainer, Eric T. Bradlow and Xiaohui Wang

Index

[More information](#)*Subject Index*

267

- content balancing in, 48
- cross-information in, 48
- error in, 7
- first step in preparing test, 4
- fundamental principle of, 3
- key questions related to, 4
- nonadaptive form of
  - context effects in, 51
  - item location in, 47
  - ordering of items in, 50
- predictive accuracy in, 5
- question of robustness in, 5
- scoring rule in, 5
- use of computer algorithm in, 49
- use of multiple choice item in, 5
- testlet, 106, 110, 111, 115
  - advantages of, 53
  - alternative approach to, 52
  - and polytomous IRT models, 72
  - and psychometric scoring model, 60
  - and test scoring model, 56
  - and true score curves, 94
  - and use of factor analytic vocabulary, 110
  - and White-African-American performance, 91
- as constructing tests method, 44
- assessment accuracy by, 65
- broad categories of, 60
- case study for, 91
- concept of, 53, 60
- conditional independence in
  - context effects in, 55
- covariates effect, 163
- data matrix score pattern, 82
- DIF detection, 80
- differential alternative functioning (DAF), 81
- DIFferential item functioning (DIF), 78, 82, 91, 97
- error of equating in, 57
- expected true score in, 95
- hierarchical form of, 65
- information extraction from, 107
- IRT model for, 77
- item presentation algorithm in, 62
- item selection algorithm in, 107
- lack of fit in, 55
- latent trait approach to, 72
- linear administration of, 65
- local dependence of item in, 73, 136
- local independence in, 73
- nonresponse theory
  - and missing test data set, 186
- parameters equality constraints, 91
- polytomous model, 211
- psychometric theory of, 108
- reasons for using, 106
- reliability of tests, 74
- robustness in, 56
- Run Out of Time (ROT), 187
- scoring model for, 61, 108
- traditional reading comprehension in, 53
- unidimensional item response model for, 54
- unidimensional model in, 55
- Validity-Based Scoring (VBS) approach in, 67
- White/African-American (DIF)
  - performance, 91, 92
- testlet response theory (TRT), 57, 108, 113
  - and bi-factor model, 110
  - and polytomous scoring of tests, 111
  - and United States Medical Licensing Examination (USMLE), 158
- binary portion of model for, 222
- description of, 110
- goal of, 118
- model item parameter estimate, 226
- model parameters
  - and covariates of interest, 184
- origin of, 60
- utilized for screening survey of (breast cancer patients), 169
- version of, 114
- testlet response theory model, 182
- testlet variance, 128, 129, 130, 134, 138, 140
  - log-normal prior for, 157
- test-scoring algorithm, 82
- total impact concept, 96
- tracelines, 173
- true score theory, 10, 14, 15, 22, 57, 114
  - and reliability, 15, 76, 114
  - and score estimation, 20, 21
  - error variance, 114
  - fundamental idea of, 15
  - Spearman-Brown Formula, 15
  - variance in, 16
- unconditional maximum likelihood (UML), 27
- United States Medical Licensing Examination (USMLE), 158, 221, 224
- Validity-Based Scoring approach, 67, 69
  - and algebra testlet, 69
- vignette word count, covariate, 158, 159, 163
- writing ability
  - multiple choice item/test to measure, 7