

# References

- Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and non-compensatory multidimensional items. *Applied Psychological Measurement, 13*, 113–127. <https://doi.org/10.1177/014662168901300201>
- Adams, R. J., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1–24. <https://doi.org/10.1177/0146621697211001>
- Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood equations. *Journal of the Royal Statistical Society, Series B, 32*, 283–301.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573. <https://doi.org/10.1007/BF02293814>
- Angoff, W. H., & Huddleston, E. M. (1958). *The multi-level experiment: A study of a two-level test system for the college board scholastic aptitude test* (ETS Statistical Report No. SR-58-21). Princeton, NJ: Educational Testing Service.
- Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York: Marcel Dekker.
- Barrada, J. R., Abad, F. J., & Veldkamp, B. P. (2009). Comparison of methods for controlling maximum exposure rates in computerized adaptive testing. *Psicothema, 21*, 313–320.
- Barrada, J. R., Mazuela, P., & Olea, J. (2006). Maximum information stratification method for controlling item exposure in computerized adaptive testing. *Psicothema, 18*, 156–159.
- Barrada, J. R., Olea, J., Ponsoda, V., & Abad, F. J. (2008). Incorporating randomness to the fisher information for improving item exposure control in cats. *British Journal of Mathematical and Statistical Psychology, 61*, 493–513. <https://doi.org/10.1348/000711007X230937>
- Barrada, J. R., Olea, J., Ponsoda, V., & Abad, F. J. (2009). Item selection rules in computerized adaptive testing: Accuracy and security. *Methodology, 5*, 7–17. <https://doi.org/10.1027/1614-2241.5.1.7>
- Barrada, J. R., Olea, J., Ponsoda, V., & Abad, F. J. (2010). A method for the comparison of item selection rules in computerized adaptive testing. *Applied Psychological Measurement, 34*, 438–452. <https://doi.org/10.1177/0146621610370152>
- Barton, M. A., & Lord, F. M. (1981). *An upper asymptote for the three-parameter logistic item-response model* (Research Bulletin No. 81-20). Princeton, NJ: Educational Testing Service.
- Bejar, I. I., Lawless, R. R., Morley, M. E., Wagner, M. E., Bennett, R. E., & Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning, and Assessment, 2*(3), 3–29.

- Belov, D. I., & Armstrong, R. D. (2009). Direct and inverse problems of item pool design for computerized adaptive testing. *Educational and Psychological Measurement, 69*, 533–547. <https://doi.org/10.1177/0013164409332224>
- Betz, N. E., & Weiss, D. J. (1974). *Simulation studies of two-stage ability testing* (Research Report 74-4). Minneapolis: University of Minnesota.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Birnbaum, A. (1969). Statistical theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology, 6*, 258–276. [https://doi.org/10.1016/0022-2496\(69\)90005-4](https://doi.org/10.1016/0022-2496(69)90005-4)
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29–51. <https://doi.org/10.1007/BF02291411>
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443–459. <https://doi.org/10.1007/BF02293801>
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for  $n$  dichotomously scored items. *Psychometrika, 35*, 179–197. <https://doi.org/10.1007/BF02291262>
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6*, 431–444. <https://doi.org/10.1177/014662168200600405>
- Bradlow, E. T. (1996). Negative information and the three-parameter logistic model. *Journal of Educational and Behavioral Statistics, 21*, 179–185. <https://doi.org/10.2307/1165216>
- Braeken, J., Tuerlinckx, F., & De Boeck, P. (2007). Copulas for residual dependencies. *Psychometrika, 72*, 393–411. <https://doi.org/10.1007/s11336-007-9005-4>
- Breiman, L., Friedman, L., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. New York: CRC Press.
- Chalmers, R. P. (2016). Generating adaptive and non-adaptive test interfaces for multidimensional item response theory applications. *Journal of Statistical Software, 71*(5), 1–39. <https://doi.org/10.18637/jss.v071.i05>
- Chang, H.-H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement, 20*, 213–229. <https://doi.org/10.1177/014662169602000303>
- Chang, S., & Ansley, T. (2003). A comparative study of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement, 40*, 71–103. <https://doi.org/10.1111/j.1745-3984.2003.tb01097.x>
- Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265–289. <https://doi.org/10.3102/10769986022003265>
- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika, 74*, 619–632. <https://doi.org/10.1007/s11336-009-9123-2>
- Cheng, Y., & Chang, H.-H. (2009). The maximum priority index method for severely constrained item selection in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology, 62*, 369–383. <https://doi.org/10.1348/000711008X304376>
- Cheng, Y., Chang, H.-H., Douglas, J., & Guo, F. (2009). Constraint-weighted a-stratification for computerized adaptive testing with nonstatistical constraints: Balancing measurement efficiency and exposure control. *Educational and Psychological Measurement, 69*, 35–49. <https://doi.org/10.1177/0013164408322030>
- Choi, S. W. (2009). FIRESTAR: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement, 33*, 644–645. <https://doi.org/10.1177/0146621608329892>
- Choi, S. W., & Swartz, R. J. (2009). Comparison of CAT item selection criteria for polytomous items. *Applied Psychological Measurement, 32*, 419–440. <https://doi.org/10.1177/0013164408322030>

- Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Davey, T., & Parshall, C. G. (1999). *New algorithms for item selection and exposure control with computerized adaptive testing*. Paper presented at the annual meeting of the American Educational Research Association.
- De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.
- DeMars, C. (2010). *Item response theory*. Oxford: Oxford University Press.
- Dodd, B. G., De Ayala, R. J., & Koch, W. R. (1995). Computerized adaptive testing with polytomous items. *Applied Psychological Measurement, 19*, 5–22. <https://doi.org/10.1177/014662169501900103>
- Dragow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polytomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology, 38*, 67–86. <https://doi.org/10.1111/j.2044-8317.1985.tb00817.x>
- Eggen, T. J. H. M. (2010). Three-category adaptive classification testing. In W. V. der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 373–387). New York: Springer.
- Eggen, T. J. H. M., & Verhelst, N. D. (2011). Item calibration in incomplete testing designs. *Psicologica, 32*, 107–132.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Finch, H., & Habing, B. (2007). Performance of DIMTEST- and NOHARM-based statistics for testing unidimensionality. *Applied Psychological Measurement, 31*, 292–307. <https://doi.org/10.1177/0146621606294490>
- Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika, 46*, 59–77. <https://doi.org/10.1007/BF02293919>
- Fraser, C., & McDonald, R. P. (2003). *Noharm 3.0 [Computer software manual]*. <http://people.niagaracollege.ca/cfraser/download/>
- Geerlings, H., Glas, C., & van der Linden, W. J. (2011). Modeling rule-based item generation. *Psychometrika, 76*, 337–359. <https://doi.org/10.1007/s11336-011-9204-x>
- Gessaroli, M. E., & De Champlain, A. F. (1996). Using an approximate chi-square statistic to test the number of dimensions underlying the responses to a set of items. *Journal of Educational Measurement, 33*, 157–179. <https://doi.org/10.1111/j.1745-3984.1996.tb00487.x>
- Glas, C. A. W., & van der Linden, W. J. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement, 27*, 247–261. <https://doi.org/10.1177/0146621603027004001>
- Glas, C. A. W., & Vos, H. J. (2010). Adaptive mastery testing using a multidimensional IRT model. In W. V. der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 409–431). New York: Springer.
- Green, B. F. J. (1950). *A general solution for the latent class model of latent structure analysis* (ETS Research Bulletin Series No. RB-50-38). Princeton, NJ: Educational Testing Service.
- Haberman, S. J., & von Davier, A. A. (2014). Considerations on parameter estimation, scoring, and linking in multistage testing. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 229–248). New York: CRC Press.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research, 22*, 144–149.
- Haley, D. (1952). *Estimation of the dosage mortality relationship when the dose is subject to error* (Technical report No. 15). Palo Alto, CA: Applied Mathematics and Statistics Laboratory, Stanford University.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
- Hambleton, R. K., & Zenisky, A. L. (2013). Reporting test scores in more meaningful ways: A research-based approach to score report design. In K. F. Geisinger, et al. (Eds.), *APA handbook of testing and assessment in psychology, Vol. 3. Testing and assessment in school psychology and education* (pp. 479–494). Washington, DC: American Psychological Association. <https://doi.org/10.1037/14049-023>

- Hattie, J. (1984). An empirical study of various indices for determining unidimensionality. *Multivariate Behavioral Research*, *19*, 49–78. [https://doi.org/10.1207/s15327906mbr1901\\_3](https://doi.org/10.1207/s15327906mbr1901_3)
- Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measure: Issues and Practice*, *26*, 44–52. <https://doi.org/10.1111/j.1745-3992.2007.00093.x>
- Hetter, R. D., & Sympson, J. B. (1997). Item exposure control in CAT-ASVAB. In J. R. McBride (Ed.), *Computerized adaptive testing: From inquiry to operation* (pp. 141–144). Washington, D.C.: American Psychological Association.
- Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, *55*, 577–602. <https://doi.org/10.1007/BF02294609>
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel–Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Erlbaum.
- Hsu, C.-L., Wang, W.-C., & Chen, S.-Y. (2013). Variable-length computerized adaptive testing based on cognitive diagnosis models. *Applied Psychological Measurement*, *37*, 563–582. <https://doi.org/10.1177/0146621613488642>
- Huitzing, H. A., Veldkamp, B. P., & Verschoor, A. J. (2005). Infeasibility in automated test assembly models: A comparison study of different methods. *Journal of Educational Measurement*, *42*, 223–243. <https://doi.org/10.1111/j.1745-3984.2005.00012.x>
- Irvine, S., & Kyllonen, P. (2002). *Item generation for test development*. Mahwah, NJ: Lawrence Erlbaum
- Jeffreys, H. (1939). *Theory of probability*. Oxford, UK: Oxford University Press.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *186*, 453–461.
- Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, *39*, 167–188. <https://doi.org/10.1177/0146621614554650>
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty six person-fit statistics. *Applied Measurement in Education*, *16*, 277–298. [https://doi.org/10.1207/S15324818AME1604\\_2](https://doi.org/10.1207/S15324818AME1604_2)
- Kelderman, H., & Rijkes, C. P. M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika*, *59*, 149–176. <https://doi.org/10.1007/BF02295181>
- Kingsbury, G. G., & Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, *2*, 359–375. [https://doi.org/10.1207/s15324818ame0204\\_6](https://doi.org/10.1207/s15324818ame0204_6)
- Kingsbury, G. G., & Zara, A. R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied Measurement in Education*, *4*, 241–261. [https://doi.org/10.1207/s15324818ame0403\\_4](https://doi.org/10.1207/s15324818ame0403_4)
- Klein Entink, R. H., Fox, J.-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, *74*, 21–48. <https://doi.org/10.1007/s11336-008-9075-y>
- Kosinski, M., Lis, P., Mahalingam, V., Kielczewski, B., Okubo, T., Stillwell, D., et al. (2013). *Concerto - Open-source online R-based adaptive testing platform (Version 4.0) [Computer software manual]*. Cambridge, UK: The Psychometrics Centre.
- Leung, C. K., Chang, H.-H., & Hau, K. T. (2003). Computerized adaptive testing: A comparison of three content balancing methods. *The Journal of Technology, Learning and Assessment*, *2*, 1–15.
- Lewis, C., & Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement*, *14*, 367–386. <https://doi.org/10.1177/014662169001400404>
- Linn, R. L., Rock, D. A., & Cleary, T. A. (1968). *The development and evaluation of several programmed testing methods* (ETS Research Bulletin Series No. i-29). Princeton, NJ: Educational Testing Service.
- Lord, F. M. (1951). *A theory of test scores and their relation to the trait measured* (ETS Research Bulletin Series No. RB-51-13). Princeton, NJ: Educational Testing Service.

- Lord, F. M. (1971a). The self-scoring flexilevel test. *Journal of Educational Measurement*, 8, 147–151. <https://doi.org/10.1111/j.1745-3984.1971.tb00918.x>
- Lord, F. M. (1971b). A theoretical study of two-stage testing. *Psychometrika*, 36, 227–242. <https://doi.org/10.1007/BF02297844>
- Lord, F. M. (1974). *Practical methods for redesigning a homogeneous test, also for designing a multilevel test* (Research Bulletin No. RB-74-30). Princeton, NJ: Educational Testing Service.
- Lord, F. M. (1977). A broad-range tailored test of verbal ability. *Applied Psychological Measurement*, 1, 95–100. <https://doi.org/10.1177/014662167700100115>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lord, F. M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika*, 48, 233–245. <https://doi.org/10.1007/BF02294018>
- Lord, F. M. (1986). Maximum likelihood and bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, 23, 157–162. <https://doi.org/10.1111/j.1745-3984.1986.tb00241.x>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Luecht, R. M. (1998). Computer-assisted test assembly using optimization heuristics. *Applied Psychological Measurement*, 22, 224–236. <https://doi.org/10.1177/01466216980223003>
- Luecht, R. M. (2014). Systems, design and implementation of large-scale multistage testing. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 69–83). New York: CRC Press.
- Luecht, R. M., & Nungester, R. J. (1998). Some practical examples of computer-adaptive sequential testing. *Journal of Educational Measurement*, 35, 229–249. <https://doi.org/10.1111/j.1745-3984.1998.tb00537.x>
- Magis, D. (2013). A note on the item information function of the four-parameter logistic model. *Applied Psychological Measurement*, 37, 304–315. <https://doi.org/10.1177/0146621613475471>
- Magis, D. (2014). On the asymptotic standard error of a class of robust estimators of ability in dichotomous item response models. *British Journal of Mathematical and Statistical Psychology*, 67, 430–450. <https://doi.org/10.1111/bmsp.12027>
- Magis, D. (2015a). *Empirical comparison of scoring rules at early stages of CAT*. Paper presented at the Conference of the International Association for Computerized Adaptive Testing, Cambridge, UK.
- Magis, D. (2015b). A note on the equivalence between observed and expected information functions with polytomous IRT models. *Journal of Educational and Behavioral Statistics*, 40, 96–105. <https://doi.org/10.3102/1076998614558122>
- Magis, D. (2015c). A note on weighted likelihood and Jeffreys modal estimation of proficiency levels in polytomous item response models. *Psychometrika*, 80, 200–204. <https://doi.org/10.1007/S11336-013-9378-5>
- Magis, D. (2016). Efficient standard error formulas of ability estimators with dichotomous item response models. *Psychometrika*, 81, 184–200. <https://doi.org/10.1007/s11336-015-9443-3>
- Magis, D., & Barrada, J. R. (2017). Computerized adaptive testing with R: Recent updates of the package catR. *Journal of Statistical Software, Code Snippets*, 76(1), 1–19. <https://doi.org/10.18637/jss.v076.c01>
- Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42, 847–862. <https://doi.org/10.3758/BRM.42.3.847>
- Magis, D., & Raïche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software*, 48(8), 1–31. <https://doi.org/10.18637/jss.v048.i08>
- Magis, D., & Verhelst, N. (2017). On the finiteness of the weighted likelihood estimator of ability. *Psychometrika*. <https://doi.org/10.1007/s11336-016-9518-9>

- Maris, E. (1995). Psychometric latent response models. *Psychometrika*, *60*, 523–547. <https://doi.org/10.1007/BF02294327>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174. <https://doi.org/10.1007/BF02296272>
- McBride, J. R., & Martin, J. T. (1983). Reliability and validity of adaptive ability tests in a military setting. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 224–236). New York: Academic Press.
- McClarty, K. L., Sperling, R. A., & Dodd, B. G. (2006). *A variant of the progressive-restricted item exposure control procedure in computerized adaptive testing systems based on the 3PL and partial credit models*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- McKinley, R. L., & Reckase, M. D. (1982). *The use of the general Rasch model with multidimensional response data* (Research Report No. ONR 82-1). Iowa City, IA: American College testing.
- Meyer, P. (2014). *Applied measurement with Jmetrik*. New York: Routledge.
- Meyer, P. (2015). *Jmetrik (version 4.0.3) [Computer software manual]*. Charlottesville, VA: Psychomeasurement Systems, LLC.
- Mills, C. N., Potenza, M. T., Fremer, J. J., & Ward, W. C. (2002). *Computer-based testing: Building the foundation for future assessments*. Mahwah, NJ: Lawrence Erlbaum.
- Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, *17*, 297–334. <https://doi.org/10.1177/014662169301700401>
- Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, *49*, 359–381. <https://doi.org/10.1007/BF02306026>
- Mislevy, R. J. (1986). Bayesian modal estimation in item response models. *Psychometrika*, *51*, 177–195. <https://doi.org/10.1007/BF02293979>
- Mislevy, R. J., & Bock, R. D. (1982). Biweight estimates of latent ability. *Educational and Psychological Measurement*, *42*, 725–737. <https://doi.org/10.1177/001316448204200302>
- Mislevy, R. J., & Chang, H.-H. (2000). Does adaptive testing violate local independence? *Psychometrika*, *65*, 149–156. <https://doi.org/10.1007/BF02294370>
- Mosteller, F., & Tukey, J. (1977). *Exploratory data analysis and regression*. Reading, MA: Addison-Wesley.
- Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, *14*, 59–71. <https://doi.org/10.1177/014662169001400106>
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 19–176. <https://doi.org/10.1177/014662169201600206>
- Muraki, E., & Bock, R. D. (2003). *PARSCALE 4.0 [Computer software manual]*. Lincolnwood, IL: Scientific Software International.
- Muraki, E., & Carlson, J. E. (1993). *Full-information factor analysis for polytomous item responses*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.
- Nydicke, S. W. (2014). *catirt: An r package for simulating IRT-based computerized adaptive tests [Computer software manual]*. <https://CRAN.R-project.org/package=catIrt> (R package version 0.5-0).
- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (2nd ed.). Thousand Oaks, CA: Sage.
- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. Thousand Oaks, CA: Sage.
- Parshall, C. G., Davey, T., & Nering, M. L. (1998). *Test development exposure control for adaptive testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, Vol. 26. Psychometrics* (pp. 125–167). Amsterdam: Elsevier.

- Rao, C. R., & Sinharay, S. (2007). *Handbook of statistics, Vol. 26. Psychometrics*. Amsterdam: Elsevier.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*, 207–230. <https://doi.org/10.2307/1164671>
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement, 35*, 311–327. <https://doi.org/10.1111/j.1745-3984.1998.tb00541.x>
- Riley, B. B., Dennis, M. L., & Conrad, K. J. (2010). A comparison of content-balancing procedures for estimating multiple clinical domains in computerized adaptive testing: Relative precision, validity, and detection of persons with misfitting responses. *Applied Psychological Measurement, 34*, 410–423. <https://doi.org/10.1177/0146621609349802>
- Roskam, E. E. (1987). Toward a psychometric theory of intelligence. In E. E. Roskam & R. Suck (Eds.), *Progress in mathematical psychology* (pp. 151–171). Amsterdam: North-Holland.
- Roskam, E. E. (1997). Models for speed and time-limit tests. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 187–208). New York: Springer.
- Rudner, L. M. (1998). Item banking. *Practical Assessment, Research, and Evaluation, 6*, 41.
- Rulison, K., & Loken, E. (2009). I've fallen and I can't get up: Can high-ability students recover from early mistakes in CAT? *Applied Psychological Measurement, 33*, 83–101. <https://doi.org/10.1177/0146621608324023>
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores. Psychometrika monograph supplement, Vol. 34* (Monograph no. 17). Richmond: Byrd Press.
- Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional space. *Psychometrika, 39*, 111–121. <https://doi.org/10.1007/BF02291580>
- Samejima, F. (1977). A use of the information function in tailored testing. *Applied Psychological Measurement, 1*, 233–247. <https://doi.org/10.1177/014662167700100209>
- Samejima, F. (1994). Some critical observations of the test information function as a measure of local accuracy in ability estimation. *Psychometrika, 59*, 307–329. <https://doi.org/10.1007/BF02296127>
- Samejima, F. (1998). *Expansion of Warm's weighted likelihood estimator of ability for the three-parameter logistic model to general discrete responses*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Santos, V. D. O. (2017). A computer-adaptive test of productive and contextualized academic vocabulary breadth in English (CAT-PAV): Development and validation. *Graduate Theses and Dissertations, 16292*. Ames, IA: Iowa State University. <http://lib.dr.iastate.edu/etd/16292>
- Schuster, C., & Yuan, K.-H. (2011). Robust estimation of latent ability in item response models. *Journal of Educational and Behavioral Statistics, 36*, 720–735. <https://doi.org/10.3102/1076998610396890>
- Segall, D. O. (2004). A sharing item response theory model for computerized adaptive testing. *Journal of Educational and Behavioral Statistics, 29*, 439–460. <https://doi.org/10.3102/10769986029004439>
- Segall, D. O. (2010). Principles of multidimensional adaptive testing. In W. V. der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 57–75). New York: Springer.
- Smith, R., & Lewis, C. (2014). Multistage testing for categorical decisions. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 189–203). New York: CRC Press.
- Snijders, T. A. B. (2001). Asymptotic null distribution of person fit statistics with estimated person parameter. *Psychometrika, 66*, 331–342. <https://doi.org/10.1007/BF02294437>
- Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. *Journal of Educational and Behavioral Statistics, 23*, 57–75. <https://doi.org/10.3102/10769986023001057>

- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201–210. <https://doi.org/10.1177/014662168300700208>
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, 589–617. <https://doi.org/10.1007/BF02294821>
- Stout, W. (2005). *DIMTEST (Version 2.0) [Computer software manual]*. Champaign, IL: The William Stout Institute for Measurement.
- Swaminathan, H., Hambleton, R. K., & Rogers, H. J. (2007). Assessing the fit of item response theory models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, Vol. 26. psychometrics* (pp. 683–718). Amsterdam: Elsevier.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361–370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Swanson, L., & Stocking, M. (1993). A model and heuristic for solving very large item selection problem. *Applied Psychological Measurement*, 17, 151–166. <https://doi.org/10.1177/014662169301700205>
- Sympson, J. B. (1978). A model with testing for multidimensional items. In D. J. Weiss (Ed.), *Proceedings of the 1977 computerized adaptive testing conference*. Minneapolis, MN: University of Minnesota.
- Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement*, 27, 159–203. <https://doi.org/10.1177/0146621603027003001>
- Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51, 567–577. <https://doi.org/10.1007/BF02295596>
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group difference in trace lines. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 147–170). Hillsdale, NJ: Erlbaum.
- Thompson, N. A. (2009). Item selection in computerized classification testing. *Educational and Psychological Measurement*, 69, 778–793. <https://doi.org/10.1177/0013164408324460>
- Urry, V. W. (1970). *A Monte Carlo investigation of logistic test models*. Unpublished doctoral dissertation, Purdue University, West Lafayette, IN.
- van der Linden, W. J. (1998a). Bayesian item selection criteria for adaptive testing. *Psychometrika*, 63, 201–216. <https://doi.org/10.1007/BF02294775>
- van der Linden, W. J. (1998b). Optimal test assembly of psychological and educational tests. *Applied Psychological Measurement*, 22, 195–211. <https://doi.org/10.1177/01466216980223001>
- van der Linden, W. J. (2000). Constrained adaptive testing with shadow tests. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 27–52). Norwell, MA: Kluwer.
- van der Linden, W. J. (2005). A comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement*, 42, 283–302. <https://doi.org/10.1111/j.1745-3984.2005.00015.x>
- van der Linden, W. J., Ariel, A., & Veldkamp, B. P. (2006). Assembling a computerized adaptive testing item pool as a set of linear tests. *Journal of Educational and Behavioral Statistics*, 31, 81–99. <https://doi.org/10.3102/10769986031001081>
- van der Linden, W. J., & Diao, Q. (2014). Using a universal shadow test assembler with multistage testing. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 101–118). New York: CRC Press.
- van der Linden, W. J., & Glas, C. A. W. (2010). *Elements of adaptive testing*. New York: Springer.
- van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika*, 73, 365–384. <https://doi.org/10.1007/s11336-007-9046-8>
- van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.



- van der Linden, W. J., & Hambleton, R. K. (2017). *Handbook of modern item response theory*. New York: Springer.
- van der Linden, W. J., Klein Entink, R., & Fox, J.-P. (2010). IRT parameter estimation with response times as collateral information. *Applied Psychological Measurement*, *34*, 327–347. <https://doi.org/10.1177/0146621609349800>
- van der Linden, W. J., & Pashley, P. J. (2000). Item selection and ability estimation in adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 1–25). Norwell, MA: Kluwer.
- van der Linden, W. J., & Pashley, P. J. (2010). Item selection and ability estimation in adaptive testing. In W. V. der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 3–30). New York: Springer.
- van der Linden, W. J., & Veldkamp, B. P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational and Behavioral Statistics*, *29*, 273–291. <https://doi.org/10.3102/10769986029003273>
- van der Linden, W. J., Veldkamp, B. P., & Reese, L. M. (2000). An integer programming approach to item bank design. *Applied Psychological Measurement*, *24*, 139–150. <https://doi.org/10.1177/01466210022031570>
- Veerkamp, W. J. J., & Berger, M. P. F. (1997). Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, *22*, 203–226. <https://doi.org/10.3102/10769986022002203>
- Veldkamp, B. P. (2014). Item pool design and maintenance for multistage testing. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 39–54). New York: CRC Press.
- Veldkamp, B. P., & van der Linden, W. J. (2010). Designing item pools for adaptive testing. In W. V. der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 231–245). New York: Springer.
- Verhelst, N. D., Verstralen, H. H. F. M., & Jansen, M. G. (1997). A logistic model for time limit tests. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 169–185). New York: Springer.
- von Davier, M., & Haberman, S. J. (2014). Hierarchical diagnostic classification models morphing into unidimensional ‘diagnostic’ classification models—a commentary. *Psychometrika*, *79*, 340–346. <https://doi.org/10.1007/s11336-013-9363-z>
- Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). New York: Routledge/Taylor and Francis.
- Wainer, H., Bradlow, E. T., & Du, Z. (2000). Testlet response theory: An analog for the 3-PL useful in testlet-based adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 246–270). Boston, MA: Kluwer-Nijhoff.
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge: Cambridge University Press.
- Wainer, H., Kaplan, B., & Lewis, C. (1992). A comparison of the performance of simulated hierarchical and linear tests. *Journal of Educational Measurement*, *29*, 243–251. <https://doi.org/10.1111/j.1745-3984.1992.tb00376.x>
- Wang, C. (2013). Mutual information item selection method in cognitive diagnostic computerized adaptive testing with short test length. *Educational and Psychological Measurement*, *73*, 1017–1035. <https://doi.org/10.1177/0013164413498256>
- Wang, C., Chang, H.-H., & Huebner, A. (2011). Restrictive stochastic item selection methods in cognitive diagnostic computerized adaptive testing. *Journal of Educational Measurement*, *48*, 255–273. <https://doi.org/10.1111/j.1745-3984.2011.00145.x>
- Wang, T., & Hanson, B. A. (2005). Development and calibration of an item response model that incorporates response time. *Applied Psychological Measurement*, *29*, 323–339. <https://doi.org/10.1177/0146621605275984>
- Warm, T. (1989). Weighted likelihood estimation of ability in item response models. *Psychometrika*, *54*, 427–450. <https://doi.org/10.1007/BF02294627>

- Weiss, D. J. (1983). *New horizons in testing: Latent trait theory and computerized adaptive testing*. New York: Academic Press.
- Weissman, A. (2014). IRT-based multistage testing. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 153–168). New York: CRC Press.
- Wright, B. O., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.
- Wright, B. O., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.
- Xu, X., Chang, H.-H., & Douglas, J. (2003). *A simulation study to compare CAT strategies for cognitive diagnosis*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Yan, D., Lewis, C., & Stocking, M. L. (2004). Adaptive testing with regression trees in the presence of multidimensionality. *Journal of Educational and Behavioral Statistics*, 29, 293–316. <https://doi.org/10.3102/10769986029003293>
- Yan, D., Lewis, C., & von Davier, A. A. (2014a). Multistage test design and scoring with small samples. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 303–324). New York: CRC Press.
- Yan, D., Lewis, C., & von Davier, A. A. (2014b). A tree-based approach for multistage testing. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 169–188). New York: CRC Press.
- Yan, D., von Davier, A. A., & Lewis, C. (2014). *Computerized multistage testing: Theory and applications*. New York: CRC Press.
- Yao, L., & Schwarz, R. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement*, 30, 469–492. <https://doi.org/10.1177/0146621605284537>
- Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, 5, 245–262. <https://doi.org/10.1177/014662168100500212>
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8, 125–145. <https://doi.org/10.1177/014662168400800201>
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187–213. <https://doi.org/10.1111/j.1745-3984.1993.tb00423.x>
- Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 111–153). Westport, CT: Praeger.
- Zheng, Y., Wang, C., Culbertson, M., & Chang, H.-H. (2014). Overview of test assembly methods in multistage testing. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 87–99). New York: CRC Press.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432–442. <https://doi.org/10.1037/0033-2909.99.3.432>

# Index

## C

### catR functions

- EPV(), 67
- GDI(), 67
- KL(), 67
- MEI(), 67
- MWI(), 67
- nextItem(), 67, 70–72, 128
- OII(), 58
- plot.cat(), 80, 81, 140
- plot.catResult(), 84, 85
- print.cat(), 80
- print.catResult(), 84
- randomCAT(), 73–77, 79–84, 90, 91, 95, 99, 104, 106, 136–138, 147, 158
- simulateRespondents(), 73–75, 81–84, 105, 106, 109, 110
- startItems(), 67–70, 75, 128

## G

### General functions

- dataGen(), 124
- genDichoMatrix(), 59–61, 124, 126
- genPattern(), 62, 64, 75, 97, 102, 103, 157
- genPolyMatrix(), 61, 124
- Ii(), 58, 59, 69, 88, 89, 124
- Ji(), 59, 124
- Pi(), 57, 58, 124
- semTheta(), 63, 65, 66, 98, 124
- thetaEst(), 63–66, 98, 104, 124, 131

## M

### mstR functions

- nextModule(), 128, 130–135, 138
- plot.mst(), 139, 140
- print.mst(), 139
- randomMST(), 123, 136–139, 144, 149, 152, 154, 158
- startModule(), 128–130