

## **Tim Moses**

Lincoln, NE 68508

[Tmoses2@unl.edu](mailto:Tmoses2@unl.edu); [tmoses2@buros.org](mailto:tmoses2@buros.org)

### **EXPERIENCE**

**Buros Center for Testing**, Lincoln, NE  
*Research Professor*

December, 2024 – Present

- Lead psychometric consulting work provided by the Buros Center.

**College Board**, New York NY

**May, 2014 – Sept, 2024**

*Robert L. Brennan Chair of Psychometric Research/Chief Psychometrician*  
*Senior Psychometrician (2014-2015)*

- Led and provided oversight and consultation for teams of psychometricians, statisticians, assessment developers and applied researchers in support of the development, administration, analyses and score reporting for College Board assessments, including SAT, PSAT/NMSQT/10, PSAT8/9, AP, CLEP and ACCUPLACER.
- Led research and consultation for Automated Essay scoring and for measurement aspects of generative AI implementations (2023-2024)
- Led special studies to develop and implement updates of CSEM and Classification results for CLEP (2021-2022)
- Led the development and implementation of psychometric linking procedures for the AP Digital exams (2020 and 2021)
- Led College Board's Psychometric Research Advisory Meetings (2017-2019)
- Led the development and implementation of IRT-based form assembly and evaluation procedures for paper-based SAT and PSAT-related test forms (2018-2019)
- Led the most recent ACT/SAT concordance (2018) and SAT Suite growth reports (2018)

**Educational Testing Service (ETS)**, Princeton, NJ  
*Senior Psychometrician (2008-2014)*  
*Psychometrician (2005-2008)*  
*Associate Psychometrician (2003-2005)*

**Sept, 2003 - May, 2014**

- AP statistical coordinator, trainer and advisor for AP statistical coordinators for PIA, DIF, Equating, Linking, Gradesetting and special studies.
- Worked and provided psychometric consultation for the research, operational work and statistical programming needs of other ETS testing programs, including Praxis, state K-12 programs, GRE, TOEIC, SAT, PSAT/NMSQT and HiSET, and in internal audits of other programs.

## **EDUCATION**

<b>Ph.D. Educational Psychology</b> <b>Measurement, Statistics and Research Design</b> University of Washington, Seattle, WA	2003
<b>M.S. Organization Development</b> Central Washington University, Ellensburg, WA	1998
<b>B.S. Psychology</b> Eastern Oregon University, La Grande, OR	1996

## **HONORS AND SERVICE**

Vice-Chair of the Executive Committee for the 2025-206 Scaling, Linking, and Equating (SLE) SIGIMIE, National Council on Measurement in Education (NCME).

Selected as an Outstanding Reviewer, 2011 for the American Educational Research Association and the Journal of Educational and Behavioral Statistics.

Winner of the 2006 Alicia Cascallar Early Career Scholar award for an outstanding paper given by a young scholar, National Council for Measurement in Education.

## **RESEARCH IN PROGRESS**

Dorans, N. J., Moses, T., & Walker, M. E. (in progress). The evolving role of conditions in procedures used to assess the fairness of measurements and their uses. Chapter for Geisinger, K., & Beaujean, A. (Eds) upcoming volume on methodology for Division 5 of the American Psychological Association. Guilford.

## PUBLISHED ARTICLES

- Moses, T., & Kim, Y. K. (2025). IRT scoring and recursion for estimating reliability and other accuracy indices. *Journal of Educational Measurement*, 1-22.  
<https://doi.org/10.1111/jedm.70008>
- Moses, T., & Puhan, G. (2022). Introduction to the special issue maintaining score comparability: Recent challenges and some possible solutions. *Journal of Educational Measurement*, 59(2), 137–139.
- Moses, T. (2022). Linking and comparability across conditions of measurement: Established frameworks and proposed updates. *Journal of Educational Measurement*, 59(2), 231-250.
- Moses, T., & Dorans, N. J. (2021). Aggregate-level test-scale linking: A new solution for an old problem? *Journal of Educational and Behavioral Statistics*, 46, 198-202.
- Moses, T., & Kim, Y. (2017). Stabilizing conditional standard errors of measurement in scale score transformations. *Journal of Educational Measurement*, 54, 184-199.
- Moses, T., & Kim, S. (2015). Methods for evaluating composite reliability, classification consistency, and classification accuracy for mixed-format licensure tests. *Applied Psychological Measurement*, 39, 314-329.
- Kim, S., Moses, T., & Yoo, H. (2015). A comparison of IRT proficiency estimation methods under adaptive multistage testing. *Journal of Educational Measurement*, 52, 70-79.
- Moses, T. (2014). Alternative smoothing and scaling strategies for weighted composite scores. *Educational and Psychological Measurement*, 74, 516-536.
- Moses, T. (2013). Adjoined piecewise linear approximations (APLAs) for equating: Accuracy evaluations of a postsmoothing equating method. *Journal of Educational Measurement*, 50, 427-446.

- Moses, T. (2014). Quantifying error and uncertainty reductions in scaling functions: An ITEMS module. *Educational Measurement: Issues and Practice*, 33, 29-40.
- Kim, S., & Moses, T. (2013). Determining when single scoring for constructed-response items is as effective as double scoring in mixed-format licensure tests. *International Journal of Testing*, 13, 314-328.
- Oh, H. J., & Moses, T. (2012). Comparison of the one- and bi-direction chained equipercentile equating. *Journal of Educational Measurement*, 49, 399-418.
- Moses, T. (2012). Relationships of measurement error and prediction error in observed score regression. *Journal of Educational Measurement*, 49, 380-398.
- Moses, T. (2012). Underlying distributions in loglinear models of discrete data. *Journal of Modern Applied Statistical Methods*, 11, 2-23.
- Moses, T. P., & Klockars, A. J. (2012). Traditional and proposed tests of slope homogeneity for non-normal and heteroscedastic data. *British Journal of Mathematical & Statistical Psychology*, 65, pp. 402-426.
- Moses, T., & Kim, S. (2012). Evaluating ranking strategies in assessing change when the measures differ across time. *Educational and Psychological Measurement*, 71, 78-98.
- Moses, T., & Zhang, W. (2011). Standard errors of equating differences: Prior developments, extensions, and simulations. *Journal of Educational and Behavioral Statistics*, 36, 779-803.
- Moses, T., Deng, W., & Zhang, Y. (2011). Two approaches for using multiple anchors in NEAT equating: A description and demonstration. *Applied Psychological Measurement*, 35, 362-379.
- Moses, T., & von Davier, A. (2011). A SAS IML macro for loglinear smoothing. *Applied Psychological Measurement*, 35, 250.

- Moses, T., & Klockars, A. J. (2010). The influence of data generation on simulation study results: Tests of mean differences. *Journal of Modern and Applied Statistical Methods*, 9, 15-27.
- Moses, T., & Holland, P. W. (2010). A comparison of statistical selection strategies for univariate and bivariate loglinear models. *British Journal of Mathematical and Statistical Psychology*, 63, 557-574.
- Moses, T., & Holland, P. W. (2010). The effects of selection strategies for bivariate loglinear smoothing models on NEAT equating functions. *Journal of Educational Measurement*, 47, 76-91.
- Moses, T., Miao, J., & Dorans, N. J. (2010). A comparison of strategies for estimating conditional DIF. *Journal of Educational and Behavioral Statistics*, 6, 726-743.
- Moses, T., Liu, J., & Dorans, N. (2010). Systemized SEA in SAS. *Applied Psychological Measurement*, 34, 552-553.
- Puhan, G., Moses, T., Yu, L., & Dorans, N. (2009). Using log-linear smoothing to improve small sample DIF estimation. *Journal of Educational Measurement*, 46, 59-83.
- Puhan, G., Moses, T., Grant, M., & McHale, F. (2009). Small-sample equating using a single group nearly equivalent test (SiGNET) Design. *Journal of Educational Measurement*, 46, 344-362.
- Moses, T. & Holland, P. W. (2009). Selection strategies for univariate loglinear models and their effect on equating function accuracy. *Journal of Educational Measurement*, 46, 159-176.
- Moses, T. (2009). A comparison of statistical significance testing strategies for selecting equating functions. *Applied Psychological Measurement*, 33, 285-306.
- Moses, T. & Klockars, A. J. (2008). An evaluation of standard, alternative and robust slope test strategies. *Journal of Modern Applied Statistical Methods*, 7, 77-93.

- Moses, T. (2008). Using the kernel method of test equating for estimating the standard errors of population invariance measures. *Journal of Educational and Behavioral Statistics*, 33, 137-157.
- Moses, T., Yang, W., & Wilson, C. (2007). Using kernel equating to check the statistical equivalence of nearly identical test editions. *Journal of Educational Measurement*, 44, 157-178.
- Moses, T. P., & Klockars, A. J. (2005). Testing for aptitude-treatment interactions in analysis of covariance and randomized block designs under assumption violations. *Journal of Modern Applied Statistical Methods*, 4, 460-472.
- Klockars, A. J. & Moses, T. (2002). Type I error rates for rank-based tests of homogeneity of slopes. *Journal of Modern Applied Statistical Methods*, 1, 452-460.
- Moses, T. P., Stahelski, A. J., & Knapp, G. R. (2000). Effects of attribute control charts on organizational performance. *Journal of Organizational Behavior Management*, 20, 69-90.
- Moses, T. P. & Stahelski, A. J. (1999). A productivity evaluation of teamwork at an aluminum manufacturing plant. *Group and Organization Management*, 24, 391-412.

## CHAPTERS

- Moses, T. (2025). Scaling, equating, and linking. In L. L. Cook & M. J. Pitoniak (Eds.), *Educational measurement* (5th ed., pp. 735–820). National Council on Measurement in Education. DOI: 10.1093/oso/9780197654965.003.0011
- Moses, T., & Dorans, N. J. (2025). Test score linking and equating. *Reference Module in Social Sciences*. Elsevier. <https://doi.org/10.1016/B978-0-443-26629-4.00064-2>
- Moses, T. (2025). Linking Responsive Assessments: Challenges and Possibilities. Chapter in R. E. Bennett, L. Darling-Hammond, & A. Badrinarayan (Eds.) *Socioculturally Responsive Assessment: Implications for Theory, Measurement, and Systems-Level Policy*. (pp. 180-198). Routledge.
- Kim, S., & Moses, T. (2025). Item Response Theory (IRT) proficiency estimation methods under multistage testing. Chapter in D. Yan & A. A. von Davier (Eds.). *Research for practical Issues and Solutions in Computerized Multistage Testing (Volume 1)*. (pp. 185-199). ETS.
- Dorans, N. J., & Moses, T. (2023). Score equating: an aspirational form of score linking. *International Encyclopedia of Education (Fourth Edition)*, (pp. 236-248). Elsevier.
- Moses, T. (2017). A review of developments and applications in item analysis. *Advancing Human Assessment: The methodological, psychological and policy contributions of ETS* (Bennett, R. E., & von Davier, M., Eds.). (pp. 19-46). Springer.
- Moses, T. (2017). Psychometric contributions: Focs on test scores. *Advancing Human Assessment: The methodological, psychological and policy contributions of ETS* (Bennett, R. E., & von Davier, M., Eds.). (pp. 47-78). Springer.



Moses, T. (2016). Estimating observed score distributions with loglinear models. *Handbook of Item Response Theory, Second Edition* (van der Linden, W. J. Eds.) (pp. 71-85). Boca Raton: Chapman and Hall.

Moses, T. (2011). Loglinear models as smooth operators: Holland's statistical applications and their practical uses. In N. J. Dorans and S. Sinharay (Eds.), *Looking back: Proceedings of a conference in honor of Paul W. Holland* (pp. 185-202). New York: NY: Springer.

Dorans, N. J., Moses, T. P. & Eignor, D. R. (2011). Equating test scores: Toward best practices. In A. A. von Davier (Ed.), *Statistical models for test equating, scaling, and linking* (pp. 21-42). New York: NY: Springer.

## RESEARCH REPORTS

Moses, T. & Wheeler, J. M. (2025). *Florida State Assessment 2025 Audit II Report: Calibration and Scoring*. Technical Report. Buros Center for Testing.

Römhild, A., Moses, T., Jonson, J., & Geisinger, K. F. (2025). *2025 Audit Report: Handscoring of the B.E.S.T. Writing 2025 Spring administration*. Technical Report. Buros Center for Testing.

Römhild, A., Moses, T., & Geisinger, K. F. (2025). *Review of handscoring specifications for Florida B.E.S.T. Writing Spring 2025*. Technical Report. Buros Center for Testing.

Kim, S. Y., Kim, Y. & Moses, T. (2020). *Impact of degrees of postsmoothing on long-term equated scale score accuracy* (CASMA Research Report No. 54). Iowa city, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.  
(Available on <http://www.education.uiowa.edu/casma>).

Kim, Y., Moses, T., & Zhang, X. (2018). *Student-Level Growth Estimates for the SAT® Suite of Assessments*. College Board Report.  
<https://collegereadiness.collegeboard.org/pdf/student-level-sat-suite-growth-estimates.pdf>

Kim, Y., Moses, T., & Zhang, X. (2018). *School-Level Growth Estimates for the SAT® Suite of Assessments*. College Board Report. <https://collegereadiness.collegeboard.org/pdf/sat-suite-growth-estimates.pdf>

Kim, S., & Moses, T. (2018). *The impact of aberrant responses and detection in forced-choice noncognitive assessment MUPP*. (ETS Research Report RR-18-32). Princeton, NJ: Educational Testing Service.

Kim, Y., Moses, T., Kolen, M. J., & Hendrickson, A. (2016). *Scaling for the SAT Suite of Assessments*. New York, NY: College Board.

- Kim, S., & Moses, T. (2016). *Investigating robustness of Item Response Theory proficiency estimators to atypical response behaviors under two-stage multistage testing*. ETS Research Report ETS GRE-16-03). Princeton, NJ: Educational Testing Service.
- Kim, S., Moses, T., & Yoo, H. (2015). *Effectiveness of Item Response Theory (IRT) proficiency estimation methods under adaptive multistage testing IRT MST*. (ETS Research Report-15-11). Princeton, NJ: Educational Testing Service.
- Kim, S., & Moses, T. (2014). *An investigation of the impact of misrouting under two-stage multistage testing: A simulation study*. (ETS Research Report 14-01). Princeton, NJ: Educational Testing Service.
- Moses, T. (2013). Constructed-Response (CR) Differential Item Functioning (DIF) evaluations for TOEIC® Speaking and Writing tests. *The Research Foundation for the TOEIC Tests: A compendium of studies: Volume II* (Powers, D. E., eds.). (pp, 10.1-10.19). Princeton, NJ: Educational Testing Service.
- Moses, T., Liu, J., Tan, A., Deng, W., & Dorans, N. (2013). *Constructed-Response DIF evaluations for mixed format tests*. (ETS Research Report 13-33). Princeton, NJ: Educational Testing Service.
- Golub-Smith, M., & Moses, T. (2014). How the scales for the GRE Revised general test were defined. *The Research Foundation for the GRE revised General Test: A Compendium of Studies*. (Wendler, C., & Bridgeman, B., eds.). (pp. 2.2.1-2.2.9). Princeton, NJ: Educational Testing Service.
- Moses, T. (2013). *ETS psychometric contributions: Focus on test scores* (Research Report 13-15). Princeton, NJ: Educational Testing Service.

- Moses, T., & Liu, J. (2011). *Smoothing and equating methods applied to different types of test score distributions and evaluated with respect to multiple equating criteria* (Research Report 11-20). Princeton, NJ: Educational Testing Service.
- Yang, W., Bontya, A. M., & Moses, T. P. (2011). *Repeater effects on score equating for a graduate admissions exam* (Research Report 11-17). Princeton, NJ: Educational Testing Service.
- Moses, T. (2011). *Evaluating empirical relationships among prediction, measurement, and scaling invariance* (Research Report 11-06). Princeton, NJ: Educational Testing Service.
- Moses, T., & Golub-Smith, M. (2011). *A scaling method that produces scale score distributions with specific skewness and kurtosis* (Research Memorandum 11-04). Princeton, NJ: Educational Testing Service.
- Dorans, N.J., Moses, T. P., & Eignor, D. R. (2010). *Principles and practices of test score equating* (Research Report 10-29). Princeton, NJ: Educational Testing Service.
- Moses, T., & Zhang, W. (2010). *Standard errors of equating differences* (Research Report 10-25). Princeton, NJ: Educational Testing Service.
- Moses, T., Deng, W., & Zhang, Y. (2010). *The use of two anchors in nonequivalent groups with anchor test (NEAT) equating* (Research Report 10-23). Princeton, NJ: Educational Testing Service.
- Moses, T., Miao, J., & Dorans, N. (2010). *A comparison of methods for estimating conditional item score differences in DIF assessments* (Research Report 10-15). Princeton, NJ: Educational Testing Service.
- Miao, J., Dorans, N. J., Moses, T. P. & Ruan, C. (2010). *Population invariance in linking AP foreign language exams*. (ETS Statistical Report No. SR-10-23). Princeton, NJ: Educational Testing Service.

- Moses, T., & Holland, P. W. (2009). *Alternative loglinear smoothing models and their effect on equating function accuracy* (Research Report 09-48). Princeton, NJ: Educational Testing Service.
- Moses, T., & Oh, H. (2009). *Pseudo Bayes estimates for test score distributions and chained equipercentile equating* (Research Report 09-47). Princeton, NJ: Educational Testing Service.
- Moses, T. & Klockars, A. J. (2009). *Strategies for testing slope differences* (Research Report 09-32). Princeton, NJ: Educational Testing Service.
- Liu, J., Moses, T., & Low, A. (2009). *Evaluation of the effects of loglinear smoothing models on equating functions in the presence of structured data irregularities* (Research Report 09-22). Princeton, NJ: Educational Testing Service.
- Moses, T., Liu, J., & Dorans, N. J. (2009). *Systemized score equity assessment (SEA) in SAS* (Research Memorandum 09-08). Princeton, NJ: Educational Testing Service.
- Moses, T., & Holland, P. W. (2009). *Selection strategies for bivariate loglinear smoothing models and their effects on NEAT equating functions* (Research Report 09-04). Princeton, NJ: Educational Testing Service.
- Moses, T. (2008). *A comparison of statistical strategies for making equating function selections* (Research Report 08-60). Princeton, NJ: Educational Testing Service.
- Moses, T., & Holland, P. W. (2008). *Notes on a general framework for observed score equating* (Research Report 08-59). Princeton, NJ: Educational Testing Service.
- Yu, L, Moses, T, Puhon, G, & Dorans, N. (2008). *DIF detection with small samples: Applying smoothing techniques to frequency distributions in the Mantel-Haenszel procedure* (Research Report 08-44). Princeton, NJ: Educational Testing Service.

- Moses, T. & Holland, P. W. (2008). *The influence of strategies for selecting loglinear smoothing models on equating functions* (Research Report 08-25). Princeton, NJ: Educational Testing Service.
- Puhan, G., Moses, T., Grant, M., & McHale, F. (2008). *An alternative data collection design for equating with very small samples* (Research Report 08-11). Princeton, NJ: Educational Testing Service.
- Dorans, N. J., Moses, T. P., & Liu, J. (2008). *Sensitivity of ACT/SAT concordances to gender and ethnicity*. (ETS Statistical Report No. SR-08-94). Princeton, NJ: Educational Testing Service.
- Moses, T. P. & Dorans, N. J. (2008). *Assessing choice of concordance sample on ACT-SAT concordances using counterbalanced design linking analysis*. (ETS Statistical Report No. SR-08-93). Princeton, NJ: Educational Testing Service.
- Dorans, N. J., & Moses, T. P. (2008). *SAT/ACT concordance 2008: Which score pair to concord?* (ETS Statistical Report No. SR-08-92). Princeton, NJ: Educational Testing Service.
- Moses, T. P. & Dorans, N. J. (2008). *ACT-SAT concordances 2008*. (ETS Statistical Report No. SR-08-91). Princeton, NJ: Educational Testing Service.
- Puhan, G., Moses, T., Yu, L., & Dorans, N. (2007). *Small-sample DIF estimation using loglinear smoothing: A SIBTEST application* (Research Report 07-10). Princeton, NJ: Educational Testing Service.
- Moses, T., & Kim, S. (2007). *Reliability and the nonequivalent groups with anchor test design* (Research Report 07-16). Princeton, NJ: Educational Testing Service.
- Moses, T., & Holland, P. W. (2007). *Kernel and traditional equipercentile equating with degrees of presmoothing* (Research Report 07-15). Princeton, NJ: Educational Testing Service.

- Moses, T. (2006). *Using the kernel method of test equating for estimating the standard errors of population invariance measures* (Research Report 06-20). Princeton, NJ: Educational Testing Service.
- Moses, T., & von Davier, A. A. (2006). *A SAS macro for loglinear smoothing: Applications and implications* (Research Report 06-05). Princeton, NJ: Educational Testing Service.
- Moses, T. P., Dorans, N. J., Miao, J., & Yon, H. (2006). Weighting strategies for Advanced Placement exams (ETS Statistical Report No. SR-06-31). Princeton, NJ: Educational Testing Service.
- Moses, T., von Davier, A. A., & Casabianca, J. (2004). *Loglinear smoothing: An alternative numerical approach using SAS* (Research Report 04-27). Princeton, NJ: ETS.

## PRESENTATIONS

- Moses, T., & Liu, J. (2025). Principles and practices in establishing score scales for medical achievement tests. *Presentation for the Scaling, Linking & Equating Special Interest Group in Measurement in Education, November, 2025.*
- Moses, T., Liu, J., & Barnum, S. (2025). Perspectives on intervals in scaling: Ability measures vs. test performance indicators. *Accepted paper and presentation for the annual meeting of the National Council for Measurement in Education, Denver, CO, April, 2025.*
- Kim, Y., & Moses, T. (2025). Incorporating NLP features into transformer-based AES models. *Accepted paper and presentation for the annual meeting of the National Council for Measurement in Education, Denver, CO, April, 2025.*
- Moses, T. (2025). Test scoring, language learners, and performance evaluations with medical students. *Accepted paper and presentation for the annual meeting of the National Council for Measurement in Education, Denver, CO, April, 2025.*
- Moses, T. (2025). Linking responsive assessments: Challenges and possibilities. *Accepted paper and presentation for the annual meeting of the National Council for Measurement in Education, Denver, CO, April, 2025.*
- Oh, H., Yoo, H., & Moses, T. (2025). Impact of different IRT calibration and scaling methods on vertical scale. *Accepted paper and presentation for the annual meeting of the National Council for Measurement in Education, Denver, CO, April, 2025.*
- Moses, T. (2025). The fifth edition of Educational Measurement (EM5): A discussion. *Presentation and session discussion for the annual meeting of the National Council for Measurement in Education, Denver, CO, April, 2025.*



- Moses, T., & Dorans, N. J. (2024). Linking Responsive Assessments: Tradeoffs and Implications for Inferences. *Presentation for the Scaling, Linking & Equating Special Interest Group in Measurement in Education, November, 2024.*
- Kim, Y. & Moses, T. (2024). Modeling AI generated essay detection using ChatGPT generated Essays. *Presentation at the annual meeting of the International Association for Computerized Adaptive Testing, Seoul, South Korea, September, 2024.*
- Moses, T. & Kim, Y. (2024). Recursive IRT-based reliability estimation. *Paper presentation at the annual meeting of the National Council on Measurement in Education, Philadelphia PA, April, 2024.*
- Oh, H., Moses, T., & Yoo, H. (2024). Vertical Scaling: Hands-on practice and evaluation of IRT linear and non-linear methods training session. *Training session at the annual meeting of the National Council on Measurement in Education, Philadelphia PA, April, 2024.*
- Moses, T. (2024). Scaling, Linking, and Equating. *Discussant at the annual meeting of the National Council on Measurement in Education, Philadelphia PA, April, 2024.*
- Moses, T. (2024). Scaling: Approaches and Interpretations. *Presentation for the Scaling, Linking & Equating Special Interest Group in Measurement in Education, January, 2024.*
- Packer, T., Moses, T., & Tabony, N. (2023). *Seminar on the Advanced Placement Program Prepared for the Fall, 2023 Gordon Commission Study Group.*
- Moses, T., & Kim, Y. (2023). A comparison of methods for adjusting samples and test score distributions for group differences. *Presentation for the International Meeting of the Psychometric Society, Baltimore, MD., July, 2023.*
- Kim, Y., Moses, T., & Bay, L. (2023). Impact of scaling on Automated Essay Scoring. *Electronic Board Session at the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2023.*

- Moses, T. (2023). Comparability Around the World: Methodological Challenges and Solutions. *Discussant for a session at the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2023.*
- Lee, D., Kim, Y., Moses, T. (2023). simpleIPD: An R Tool for Evaluating Common Item Parameter Drift. *Software and Training Module demonstration for a session at the annual meeting of the National Council on Measurement in Education, April, 2023.*
- Moses, T., & Kim, Y. (2023). Comparing SAT cohorts before and after the COVID-19 pandemic. *Paper presentation at the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2023.*
- Moses, T. (2023). Linking Analyses for the Current and Digital SAT. *Paper prepared for the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2023*
- Moses, T., & Dorans, N. J. (2022). Scaling, Linking, & Equating Beyond Our Comfort Zones. *Co-discussant at the annual meeting of the National Council on Measurement in Education, April, 2022.*
- Kim, Y., & Moses, T. (2022). Automated Essay Scoring using neural network model with hybrid features. *Electronic Board Session at the annual meeting of the National Council on Measurement in Education, San Diego, CA, April, 2022.*
- Moses, T. (2022). Discussion of Dorans' Revisiting Lingering Linking Issues: How; Never-ending? *Discussant at the annual meeting of the American Educational Research Association, San Diego, CA, April, 2022.*
- Moses, T., & Puhon, G. (2022). Recent Challenges to ensuring score comparability. *Co-organizer for a session at the annual meeting of the National Council on Measurement in Education, San Diego, CA, April, 2022.*

- Moses, T. (2022). Linking by definition and presumed invariance across conditions of measurement. *Paper presentation at the annual meeting of the National Council on Measurement in Education, San Diego, CA, April, 2022.*
- Moses, T. P. & Dorans, N. J. (2021, June). Aggregate-level test-scale linking: A new solution for an old problem? *Paper presented at the 2021 annual meeting of the National Council on Measurement in Education, Virtual Meeting.*
- Moses, T. (2021). Scaling, Linking, & Equating Du Jour: A Discussion with Experts. Co-panelist at the annual meeting of the National Council on Measurement in Education, April, 2021.
- Moses, T. (2021). Procedures for establishing and evaluating linkages between scores collected in different modes. *Discussant at the annual meeting of the National Council on Measurement in Education, April, 2021.*
- Kim, S., & Moses, T. (2019). Equating precision of two data collection designs: Random groups versus common items. *Paper presentation at the annual meeting of the National Council on Measurement in Education, British Columbia, CA, April, 2019.*
- Cui, W., Kim, Y., & Moses, T. (2019). Transition to Item Response Theory for an Existing CTT-based Assessment. Paper presented at the 2019 annual meeting of the National Council on Measurement in Education, *Paper prepared for the annual meeting of the National Council on Measurement in Education, British Columbia, CA, April, 2019.*
- Allen, J., Handel, S. J., Moses, T., & Pauls, K. (2018). ACT/SAT Concordance: What you need to know. *Presentation for the National Association of College Admission Counselors and the American Association of Collegiate Registrars and Admissions Officers, November, 2018.*

- Kim, Y., & Moses, T. (2018). Computing test score distributions with the Hierarchical Rater Model. *Presentation for the annual meeting of the National Council on Measurement in Education, New York, NY, April, 2018.*
- Kim, S., Kim, Y., Moses, T., & Zhang, C. (2018). Impact of degrees of postsMOOTHING on long-term equated scale score accuracy. *Presentation for the annual meeting of the National Council on Measurement in Education, New York, NY, April, 2018.*
- Walker, M., Kim, S., & Moses, T. (2018). Evaluating group equivalence in a random groups design. *Presentation for the annual meeting of the National Council on Measurement in Education, New York, NY, April, 2018.*
- Moses, T. (2018). Advances in IRT equating: Old methods and new tricks. *Chair of a session for the annual meeting of the National Council on Measurement in Education, New York, NY, April, 2018.*
- Moses, T. (2018). Introduction to Advances in IRT equating: Old methods and new tricks. *Paper and presentation prepared session for the annual meeting of the National Council on Measurement in Education, New York, NY, April, 2018.*
- Kim, S., & Moses, T. (2018). Comparison of three possible approaches to Composite-score equating. *Presentation prepared session for the annual meeting of the National Council on Measurement in Education, New York, NY, April, 2018.*
- Kim, Y., & Moses, T. (2018). IRT true score equating 2.0. *Presentation prepared session for the annual meeting of the National Council on Measurement in Education, New York, NY, April, 2018.*
- Kim, S., & Moses, T. (2017). Impact of aberrant responding and its detection in forced-choice noncognitive assessments. *Paper and presentation prepared session for the annual*

*meeting of the National Council on Measurement in Education, San Antonio, TX, April, 2017.*

Kim, S., & Moses, T. (2017). Impact of aberrant responding and its detection in forced-choice noncognitive assessments. *Paper and presentation prepared session for the annual meeting of the National Council on Measurement in Education, San Antonio, TX, April, 2017.*

Barry, C., Kim, Y., Cui, W., & Moses, T. (2017). Development of a noncognitive assessment for supporting College Readiness and Success. *Paper and presentation prepared session for the annual meeting of the National Council on Measurement in Education, San Antonio, TX, April, 2017.*

Kim, S., & Moses, T. (2017). Impact of aberrant responding and its detection in forced-choice noncognitive assessments. *Paper and presentation prepared session for the annual meeting of the National Council on Measurement in Education, San Antonio, TX, April, 2017.*

Moses, T. (2017). Applications of regression concepts to test linking. *Organizer of a session for the annual meeting of the National Council on Measurement in Education, San Antonio, TX, April, 2017.*

Moses, T. (2017). Prediction error implications of using concord scores in regression functions. *Paper and presentation prepared for the annual meeting of the National Council on Measurement in Education, San Antonio, TX, April, 2017.*

Chen, L., Moses, T., Kim, Y., & Cui, W. (2017). A comparative study on methods for estimating scale score reliability. *Presentation prepared for the annual meeting of the National Council on Measurement in Education, San Antonio, TX, April, 2017.*

- Kim, Y., Moses, T., & Antal, J. (2017). Comparing propensity score matching and weighting for random groups equating accuracy. *Electronic board presentation prepared for the annual meeting of the National Council on Measurement in Education, San Antonio, TX, April, 2017.*
- Liu, J., & Moses, T. (2016). DIF related to test takers' culture background and language proficiency. *Electronic board presentation prepared for the annual meeting of the National Council on Measurement in Education, Washington, DC, April, 2016.*
- Kaliski, P., Reshetar, R., Moses, T., Deng, H., & Rawls, A. (2016). Producing concordance tables for the transition to the Redesigned SAT. *Presentation prepared for the annual meeting of the National Council on Measurement in Education, Washington, DC, April, 2016.*
- Kim, Y., Moses, T., & Lee, W. (2016). Classification accuracy and consistency using the loglinear model based methods. *Presentation prepared for the International Meeting of the Psychometric Society, Asheville, NC., July, 2016.*
- Moses, T., Kim, Y., & Barry, C. (2016). Implications of pairing strategies and scoring approaches in forced-choice assessments. *Presentation prepared for the International Meeting of the Psychometric Society, Asheville, NC, July, 2016.*
- Zhang, X., Rawls, A., Hendrickson, A., & Moses, T. (2016). Evaluation of small sample DIF estimation methods. *Poster prepared for the International Meeting of the Psychometric Society, Asheville, NC, July, 2016.*
- Kim, Y., & Moses, T. (2015). A method of estimating test score distributions by item sampling. *Presentation for the International Meeting of the Psychometric Society, Beijing, China., July, 2015.*

- Kim, S., & Moses, T. (2014). Comparing IRT proficiency estimators for aberrant responding and multistage testing. *Paper and presentation prepared for the annual meeting of the National Council on Measurement in Education, Philadelphia, PA, April, 2014.*
- Cao, Y., Moses, T. & Bontya, A. (2014). Pragmatic versus statistical strategies for log-linear smoothing model selection. *Presentation prepared for the annual meeting of the National Council on Measurement in Education, Philadelphia, PA, April, 2014.*
- Kim, J.P., Kramer, L., & Moses, T., & Oh, H. (2013). An overview of psychometric work in real world. *Training session and presentation prepared for the annual meeting of the National Council on Measurement in Education, San Francisco, CA, April, 2013.*
- Kim, S., & Moses, T. (2012). Determining when single scoring for constructed-response items is as effective as double scoring in licensure tests. *Paper and presentation prepared for the annual meeting of the National Council on Measurement in Education, Vancouver, BC, April, 2012.*
- Moses, T. (2010). Updating the ACT/SAT concordances- Coordinated session. *Organizer of a session prepared for the annual meeting of the National Council on Measurement in Education, Denver, CO, April, 2010.*
- Moses, T., Crouse, J., & Sconing, J. (2010). Selection decisions for the sample data used in the ACT-SAT concordances. *Paper and presentation prepared for the annual meeting of the National Council on Measurement in Education, Denver, CO, April, 2010.*
- Liu, J., Dorans, N., & Moses, T. (2010). Evaluating the subpopulation sensitivity of the ACT-SAT concordances. *Paper and presentation prepared for the annual meeting of the National Council on Measurement in Education, Denver, CO, April, 2010.*

- Moses, T. (2010). Invariance and equating – Paper session. *Discussion for a session prepared for the annual meeting of the National Council on Measurement in Education, Denver, CO, April, 2010.*
- Oh, H. & Moses, T. (2010). Comparison of Angoff and Kolen-Brennan approaches in conducting chained equipercentile equating. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, Denver, CO, April, 2010.*
- Albano, A., Moses, T., & Liu, J. (2010). The impact of loglinear smoothing on equating with structured data irregularities. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, Denver, CO, April, 2010.*
- Moses, T., & Zhang, W. (2009). Standard errors of equating differences. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, San Deigo, CA, April, 2009.*
- Moses, T., & von Davier, A. A. (2009). Research directions in equating-Coordinated session. *Co-organizer for a session prepared for the annual meeting of the National Council on Measurement in Education, San Deigo, CA, April, 2009.*
- Moses, T., Deng, W., & Zhang, L (2009). The use of two anchors in NEAT equating. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, San Deigo, CA, April, 2009.*
- Moses, T., & Holland, P. W. (2009). Alternative loglinear smoothing models and their effect on equating function accuracy. *Paper and presentation for a session prepared for the annual meeting of the American Educational Research Association, San Deigo, CA, April, 2009.*



- Moses, T., & Klockars, A. J. (2009). The influence of data generation on simulation study results. *Paper and presentation for a session prepared for the annual meeting of the American Educational Research Association, San Deigo, CA, April, 2009.*
- Dorans, N. J., Moses, T., & Liu, J. (2009, May). ACT/SAT concordances: What are concordances, why concord, which scores, how to concord, what results? *Presentation to the College Board, New York, NY.*
- Dorans, N. J., Moses, T., & Liu, J. (2009, May). ACT/SAT concordances: Selected results. *Presentation to the NCAA Data Analysis Research Network, Indianapolis, IN.*
- Dorans, N. J., Moses, T. P., & Liu, J. (2009, February). The variability of ACT/SAT concordances. *Presentation at the Ledyard R Tucker Psychometric Workshop, Educational Testing Service. Princeton, NJ.*
- Moses, T., Miao, J., & Dorans, N. (2008). A comparison of methods for estimating conditional item score differences in DIF assessments. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, New York, NY, April, 2008.*
- Dorans, N. J. & Moses, T. P. (2008, September). Holland's blood test and contest distinction: An application to differential item functioning. *Paper presented at the ETS Conference, Looking Back: A Conference in Honor of Paul W. Holland, Princeton, NJ.*
- Moses, T. (2008, September). Loglinear models as smooth operators. *Paper presented at the ETS Conference, Looking Back: A Conference in Honor of Paul W. Holland, Princeton, NJ.*
- Dorans, N. J., & Moses, T. P. (2008, July). Score equating: Attempting to ensure a fair contest. *Paper presented at the annual meeting of the Psychometric Society, Durham, NH.*
- Dorans, N. J., & Moses, T. P. (2008, June). ACT/SAT concordance. *Paper presented at the NCAA Data Analysis Research Network, Washington, DC.*

- Moses, T., & Holland, P. W. (2007). Selection strategies for loglinear smoothing models and their effect on equating function accuracy. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2007.*
- Yu, L., Moses, T., Puhon, G., & Dorans, N. (2007). A practical investigation of detecting DIF with small samples. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2007.*
- Puhon, G., Moses, T., Yu, L., & Dorans, N. (2007). Small-sample DIF estimation using log-linear smoothing: A SIBTEST application. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2007.*
- Moses, T., & Miao, J. (2007). An incisive look at a suspect regression correction. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2007.*
- Moses, T. (2007). Using the kernel method for estimating the standard errors of population invariance measures. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2007.*
- Moses, T., & Klockars, A. J. (2007). An evaluation of standard, alternative and robust analysis strategies for aptitude-treatment interactions. *Paper and presentation prepared for the annual meeting of the American Educational Research Association, Chicago, IL, April, 2007.*
- Moses, T., & Holland, P. W. (2006). Kernel equating without presmoothing. *Paper and presentation prepared for the annual meeting of the American Educational Research Association, San Francisco, CA, April, 2006.*

- Moses, T., & von Davier, A. A. (2006). Using PROC GENMOD for loglinear smoothing. *Paper and presentation for the annual meeting of the Southeast SAS Users Group, Portsmouth, VA, October, 2006.*
- Moses, T., Yang, W., & Wilson, C. (2005). Using kernel equating to check the statistical equivalence of nearly identical test editions. *Paper and presentation for a session prepared for the annual meeting of the National Council on Measurement in Education, Montreal, Canada, April, 2005.*
- Moses, T., & Klockars, A. J. (2005). Testing for aptitude-treatment interactions in analysis of covariance and randomized block designs under assumption violations. *Paper and presentation prepared for the annual meeting of the American Educational Research Association, Montreal, Canada, April, 2005.*
- Moses, T., & von Davier, A. A. (2005). A SAS macro for loglinear smoothing: Applications and implications. *Paper and presentation for a session prepared for the annual meeting of the American Educational Research Association, Montreal, Canada, April, 2005.*
- Moses, T., & von Davier, A. A. (2004). Using PROC GENMOD for loglinear smoothing. *Paper and presentation for the annual meeting of the Northeast SAS Users Group, Baltimore, MD, November, 2004.*
- Klockars, A. J., & Moses, T. (2001). Dealing with violations of the assumptions for aptitude x treatment interactions in analysis of covariance. *Paper and presentation prepared for the annual meeting of the American Educational Research Association, Seattle, WA, April, 2001.*
- Schepman, S., Moses, T., & Stahelski, A. J. (1999). An analysis of the efficacy of attribute control charts as a quality improvement tool. *Presentation prepared for the fifth international conference of the Decision Science Institute.*

Moses, T., Stahelski, A. J., & Schepman, S. (1998). The effects of the introduction of problem solving teams on productivity in an aluminum plant. *Presentation prepared for the Nevada Conference on Organization Change.*

## **CONSULTING**

National Board of Osteopathic Medical Examiners (NBOME): Standards and Assurances

Committee (2025)

Scholastic: October, 2024-June, 2025

National Board of Osteopathic Medical Examiners (NBOME): March-August, 2024

Enrollment Management Association (formerly the Secondary School Admission Test Board):

2013 - 2019

University of Cape Town, South Africa and the Alternative Admissions Research Project: 2008

## **REVIEWER WORK**

Applied Measurement in Education (\*Currently on the Editorial Board)

ETS Research Report Series

Journal of Educational and Behavioral Statistics

Journal of Educational Measurement

Practical Assessment, Research, and Evaluation

NCME Conference Proposals