

STEPHEN G. SIRECI, PhD

College of Education—Center for Educational Assessment
University of Massachusetts
Amherst, MA 01003-4140
413-545-0564

sireci@acad.umass.edu

Google Scholar: <https://scholar.google.com/citations?user=qrGroqMAAAAJ&hl=en>

Education

Ph.D. in Psychology (Psychometrics), Fordham University, Bronx, NY

Dissertation: *Evaluating test content using cluster analysis and multidimensional scaling*

Master of Arts in Psychology, Loyola College, Baltimore, MD

(Counseling, with a concentration in Employee Assistance Programs)

Thesis: *The effects of aerobic exercise on select psychological variables among the chronic mentally ill.*

Bachelor of Arts in Psychology, Loyola College, Baltimore, MD

Professional Experience

September, 1995 to Present:

Distinguished University Professor¹, College of Education, University of Massachusetts Amherst
Executive Director, Center for Educational Assessment, University of Massachusetts Amherst
Teach graduate courses in statistics, scaling methods, test development, educational assessment, educational assessment policy, validity theory, and research methods. Supervise and mentor doctoral students in ongoing research. Acquire, direct, and coordinate research grants and contracts for the Center for Educational Assessment (over \$20 million to date). Supervise Center staff. Current research activities include personalized, designed-in-real-time assessments; culturally sustaining assessment; cross-lingual assessment; assessing test dimensionality; implementing innovative scaling and standard setting methodologies; appraising test validity, designing digitally-based tests; implementing and evaluating test accommodations, ensuring historically minoritized people are considered in all aspects of test development and evaluation; and refining validity theory.

June, 1992 to August, 1995:

Senior Psychometrician, American Council on Education, Washington, D.C.
Directed, supervised, and coordinated research and test development activities related to the Tests of General Educational Development (GED Tests). Psychometric responsibilities included test construction, investigations of score reliability and validity, item analyses, standard setting, IRT research, and equating. Management responsibilities included coordinating norming and equating projects, training professional staff, supervising support staff, and mentoring psychometric fellows. Principal author of GED technical manual. Initiated and directed research linking English and Spanish language versions of the GED Tests. Initiated sensitivity (fairness) review. Directed and coordinated client research projects including statewide norming and GED/high school exit test comparability studies. Provided guidance on psychometric issues related to testing persons with disabilities. Author/co-author of numerous policy and technical reports.

¹Promoted from Assistant Professor to Associate Professor, May 2000. Promoted from Associate Professor to Professor September 1, 2004. Promoted from Professor to Distinguished University Professor June 20, 2017. Also appointed Adjunct Associate Professor, Psychology Department at UMass Amherst (November 2002).

August, 1990 to July, 1992:

Psychometrician, American Institute of Certified Public Accountants, New York, NY

Responsible for research and psychometric activities related to the Uniform CPA Examination and the Accredited Personal Financial Specialist Examination. Conducted item analyses (including applications of item response theory) and reliability and validity investigations. Trained item writers and Examination Preparation Subcommittees in item development procedures. Provided psychometric expertise to the Board of Examiners, Examination Change Implementation Task Force, and the Grading Subcommittee. Developed computerized item banking system and item analysis reporting package. Project manager for the Grading Methodology and the Standard Setting Task Forces.

June, 1990 to August, 1990:

Predoctoral Fellow, Educational Testing Service, Princeton, NJ

Evaluated reading passages associated with the new version of the SAT. Assessed the differential item functioning (bias) of these passages and determined score reliability. All analyses were conducted using item response theory within a testlet framework.

May, 1989 to June, 1990:

Research Supervisor of Testing, Newark Board of Education, Newark, NJ

(Promoted from Senior Research Assistant in January, 1990). Coordinated district-wide testing programs for over 50,000 students, including proficiency, achievement, and bilingual testing. Aided in the selection and placement of students into instructional programs. Analyzed and reported district-wide test results. Ran workshops for high school test administrators. Performed program evaluations of remedial instructional programs. Established longitudinal student data base. Reported and disseminated test results and results of Chapter I program evaluations. Conducted equating studies, established cutoff scores, and performed item analyses for locally-developed tests. Supervised professional and support staff.

April, 1986 to September, 1986:

Residential Counselor, Omni House, Glen Burnie, MD

Residential counselor in psychosocial rehabilitation center for adults with chronic psychiatric disabilities. Organized activities and provided group and individual counseling.

Selected Commissions, Blue-Ribbon Panels, and Advisory Committees

2004-present	Puerto Rico Technical Advisory Committee (Chair, since 2010)
2004-present	Texas Technical Advisory Committee
2014-present	Educational Records Bureau, Technical Advisory Committee
2017-present	ELPA21 Technical Advisory Committee
2018-present	Duolingo English Test Technical Advisory Committee
2025	BBVA Foundation Frontiers of Knowledge Awards, Jury Chair, Social Sciences
2023-present	SweSAT, International Advisory Committee (Umea, Sweden)
2018-2024	HumRRO Board of Trustees
2019-2024	Maryland Technical Advisory Committee
2022-2023	New York State Blue Ribbon Commission on Graduation Measures
2018-2023	Centre for Educational Measurement, University of Oslo
2021-2022	Technical Work Group, Ministry of Education, Chile
2010-2017	Florida Alternate Assessment Technical Advisory Committee
2005-2011	National Center on Educational Outcomes, Research-to-Practice Panel
2006-2011	National Alternate Assessment Center, Expert Panel
2006-2010	Psychometric Oversight Committee, American Institute of CPAs
2006-2009	Assessing multiple sources reading comprehension, Advisory Board

Selected National Commissions, Blue-Ribbon Panels, and Advisory Committees (continued)

2007-2009	Massachusetts Teacher Educator Licensure Pass Rate Study Group
2004-2009	Designing Accessible Reading Assessments Technical Advisory Committee
2004-2009	Partnership for Accessible Reading Assessment Technical Advisory Committee
2003-2009	Graduate Management Admissions Council Technical Advisory Committee
2003-2009	Federation of State Boards of Physical Therapy Technical Advisory Committee
2004-2008	New Hampshire Assessment Technical Advisory Committee
2004-2008	New Hampshire Enhanced Assessment Initiative Tech. Advisory Committee
2005-2007 & 2012-2016	National Board of Professional Teaching Standards Assessment Certification Advisory Panel (Chair)
2003-2007	Senior Scientist, The Gallup Organization
2003-2006	Montana Comprehensive Assessment System Technical Advisory Committee
2003-2006	Graduate Records Exam Technical Advisory Committee
2005-2006	Adult ESL Assessment Design Team, Center for Applied Linguistics
2005-2006	Technical Adequacy of Assessments for Alternate Student Populations, WestEd
2002-2004	National Assessment of Educational Progress Quality Assurance Panel
2002-2003	Maine Comprehensive Assessment System Technical Advisory Committee
2003	Committee on Diagnostic Methodology (The College Board)
2001-2002	College Board's Blue Ribbon Panel on the Flagging of Test Scores
2001-2002	Commission on Instructionally Supportive Assessment
2001-2002	Massachusetts Comprehensive Assessment System Blue Ribbon Panel

Selected Awards/Honors

Member, National Academy of Education, elected 2025
 Fellow, International Test Commission, 2024
 Professor Extraordinary, Dept. of Industrial Psych., Stellenbosch University, South Africa 2022-2024
 Distinguished University Professor, University of Massachusetts, 2017
 Samuel J. Messick Memorial Lecture Award, TOEFL Program, Educational Testing Service, 2017
 Public Engagement Fellow, University of Massachusetts, 2016
 Leo D. Doherty Award for Outstanding Leadership & Service, NE Educ. Research Assoc., 2014
 Samuel F. Conti Faculty Fellowship Award, University of Massachusetts Amherst, 2012
 Thomas Donlon Award for Distinguished Mentoring (Northeastern Educ. Research Assoc.), 2010
 Fellow, American Educational Research Association, 2009
 Outstanding Accomplishments in Research and Creative Activity, UMass Amherst, 2009
 Fellow, Div. of Evaluation, Measurement, and Statistics, American Psychological Association, 2007
 Chancellor's Award, University of Massachusetts Amherst, 2007
 Outstanding Teacher Award, School of Education, University of Massachusetts, 2002-2003

Selected Funded Research

(PI unless otherwise indicated)

2023	Massachusetts Department of Education: <i>Developing and Validating Assessments for Adult Learners in Massachusetts</i> (4 years, approximately \$1,700,000)
2023	Massachusetts Consortium for Innovative Education Assessment, \$10,000
2022	Massachusetts Department of Education: <i>The English Test for Adults: Assessing and Supporting Adult Learners' Acquisition of English Proficiency</i> , 5 years, \$1,950,000
2021	Institute of Educational Sciences, U.S. Department of Education, <i>Adult Skills Assessment Program</i> (Co-PIs April Zenisky and Maria Elena Oliveri), 5 years, \$3,800,000.
2021	Caveon, <i>Item Innovations for 21st-Century Assessments</i> , 2 years, \$108,000, year 3 (Co-PI Craig Wells) \$67,000

Selected Funded Research (continued)

- 2020 Chan Zuckerberg Initiative, *Improving Educational Measurement Through Diversity* (Co-PI with Jennifer Randall), \$225,000
- 2019 WestEd, *Evaluating Interventions for Vulnerable Populations*, \$11,000
- 2019 Massachusetts Department of Education: *Developing and Validating Assessments for Adult Learners in Massachusetts* (4 years, approximately \$1,640,000)
- 2018 Pearson Education, *Technical Assistance to Support the National Board of Professional Teaching Standards Examinations* (1 year, \$80,000)
- 2017 Smarter Balanced Assessment Consortium (Co-PI, (Craig Wells, PI), \$91,000
- 2017 Pearson Education, *Technical Assistance to Support the National Board of Professional Teaching Standards Examinations* (2 years, \$577,000)
- 2016 College Board: *Linking AP Physics and Calculus and TIMSS Advanced Scales*. \$140,000
- 2015 Massachusetts Department of Education: *Developing and Validating Assessments for Adult Learners in Massachusetts* (4 years, approximately \$1,400,000)
- 2015 Educational Testing Service: *Improving Educational Assessment through Psychometric Research* (6 years, Co-PI R. Hambleton, approximately \$1,000,000)
- 2012 College Board: *Developing A Measure of Pedagogical Content Knowledge for AP Biology Teachers* (3 years, approximately \$185,000)
- 2011 Massachusetts Department of Education: *Developing and Validating Assessments for Adult Learners in Massachusetts* (4 years, approximately \$900,000)
- 2010 Measured Progress: *Score Equating and other technical work for the Massachusetts Comprehensive Assessment System* (4 years, Co-PI (R. Hambleton,PI) approximately \$600,00)
- 2010 Educational Testing Service: *Improving Educational Assessment through Psychometric Research* (5 years, Co-PI with R. Hambleton, approximately \$1,200,000)
- 2010 World Bank, *Developing a World Class Master's Degree Program in Educational and Psychological Measurement at the Higher School of Economics (Moscow)* (Co-PI R. Hambleton, approx \$90,000)
- 2009 Pearson Educational Measurement: *Enhancing the Validity of Educational Achievement Tests* (5 years, Co-PI J. Randall, approximately \$400,000)
- 2008 American Institute of Certified Public Accountants: *Standard Setting Research for the Uniform CPA Exam* (Co-PI R. Hambleton, approximately \$45,000)
- 2007 Massachusetts Department of Education: *Developing and Validating Assessments for Adult Learners in Massachusetts* (4 years, approximately \$1,400,000)
- 2007 College Board: *Calibrating IRT Item Statistics & Equating AP Tests* (2 years, Co-PI R. Hambleton, approximately \$250,000)
- 2007 Pearson Educational Measurement: *Enhancing the Validity of Educational Achievement Tests* (Co-PI R. Hambleton, approximately \$52,000)
- 2007 National Science Foundation: *Electronic Delivery and Criterion-referencing of Assessment Materials for Chemistry* (2 years, Co-PIs D. Hart, S. Battisti, approximately \$58,000 for CEA)
- 2006 College Board: *Identifying Key Characteristics of Public Postsecondary Institutions Fostering Success for Under-Represented Students* (Co PI K. O'Meara, approximately \$180,000)
- 2005 U.S Department of Education: *Comprehensive Evaluation of NAEP* (3-year subcontract through Buros/University of Nebraska, approximately \$600,000)
- 2003 Massachusetts Department of Education: *Designing Quality Program Monitoring and Evaluation Systems for Massachusetts Adult and Community Learning Services* (5 years, approximately \$1,000,000)
- 2003 All Kinds of Minds of Minds Institute: *Evaluating Student Achievement* (Co-PI Lisa Keller, approximately \$320,000)
- 2001 Educational Testing Service: *Applying/Evaluating Emerging Measurement Models* (Co-PIs R. Hambleton & H. Swaminathan, approximately \$100,000)

Selected Funded Research (continued)

- 2001 Evaluation of STEMTEC Program (co-evaluator for NSF-funded grant)
- 2002 Evaluation of STEMTEC-II Program (co-evaluator for NSF-funded grant)
- 1999 American Institute of Certified Public Accountants: *Psychometric research for the Uniform Certified Public Accountants Examination* (5 years, \$240,000)
- 1999 Microsoft Corporation: *Develop and evaluate computerized-adaptive test algorithms and item cloning techniques* (2 years, Co-PI R. Hambleton, approximately \$280,000)
- 1999 The College Board: *Research alternative designs for setting standards on AP examinations.*
- 1998 Massachusetts Department of Education: *Psychometric properties of MCAS exams* (Co-PIs R. Hambleton, H. Swaminathan, approximately \$150,000)
- 1997: Microsoft Corporation: *Study of Computerized-adaptive Test Algorithms and Translation Equivalence of Microsoft exams* (Co-PI R. Hambleton, approximately \$180,000)
- 1996: Novell, Inc.: *Investigate comparability of computer examinations across multiple languages* (approximately, \$18,000)

Publications

Books

- International Test Commission and Association of Test Publishers (2022). *Guidelines for technology-based assessments* (J. Weiner & S. Sireci, Eds.). Available at <https://www.intestcom.org/page/28> and <https://www.testpublishers.org/white-papers>.
- Gordon, E. W., Sireci, S. G., Armour-Thomas, E., Baker, E. L., Everson, H. T., Tucker, E. (Eds.) (2025). *Handbook for Assessment in the Service of Learning Book Series*. Amherst, MA: University of Massachusetts Libraries Press. [Editorial team for 3-volume book series.]
- Sireci, S. G., Tucker, E., & Gordon, E. (Eds.) (2025). *Handbook on assessment in the service of learning: Volume II*. Amherst, MA: University of Massachusetts Library Press. Available at <https://openpublishing.library.umass.edu/plugins/books/4/>.

Peer-Reviewed Journal Articles and Book Chapters

- Allalouf, A., Hambleton, R. K., & Sireci, S. G. (1999). Identifying the sources of differential item functioning in translated verbal items. *Journal of Educational Measurement*, 36, 185-198.
- Allalouf, A., & Sireci, S. G. (2012). Guest editorial: Dissemination of measurement concepts and knowledge to the public. *Educational Measurement: Issues and Practice*, 31(20), 1.
- Araneda, S., Lee, D., Lewis, J., Sireci, S. G., Moon, J. A., Lehman, B., Arslan, B., & Keehner, M. (2022). Exploring relationships among test takers' behaviors and performance using response process data. *Education Sciences*, 12, 104. <https://doi.org/10.3390/educsci12020104>.
- Arias, A., & Sireci, S. G. (2021). Validez y validación para pruebas educativas y psicológicas. *Revista Iberoamericana de Psicología* 14 (1), 11-22. Available at <https://reviberopsicologia.iberro.edu.co/article/view/rip.14102/1674>.
- Arias, A., & Sireci, S. G. (in press). Standardizing test administrations. Chapter to appear in A.N. Herrera, E.M. Árias, A.I. Berrio, S.L. Camargo and A.C. Santana (Eds.) *Manual de Psicometría (Volume 2)*.

Publications (continued)

- Badham, L., Oliveri, M. E., & Sireci, S. G. (2025). Navigating multi-language assessments: Best practices for test development, linking, and evaluation. *Psicothema*, 37(3), 1-11. <https://doi.org/10.70478/psicothema.2025.37.19>
- Baron, P., Sireci, S. G., & Slater, S. (2020). Evaluating panelists' understanding of standard setting data. *Educational Measurement: Issues and Practice* 40(2) 16-25.
- Benitez, I. B., Padilla, J.-L., Sireci, S. G., & Hildalgo, D. (2016). Using mixed methods to interpret differential item functioning. *Applied Measurement in Education*, 29, 1-16, DOI: 10.1080/08957347.2015.1102915.
- Brown-Chidsey, R., Boscardin, M. L., & Sireci, S. G. (2001). Computer attitudes and opinions of students with and without learning disabilities. *Journal of Educational Computing Research*, 24, 183-204.
- Casas Hernandez, M., & Sireci, S. G. (2024). Measurement invariance across immigrant and non-immigrant populations on PISA non-cognitive scales. *International Journal of Testing*.
- Chakwera, E., Khembo, D., & Sireci, S. G. (2004). High-stakes testing in the warm heart of Africa: The challenges and successes of the Malawi National Examinations Board. *Education Policy Analysis Archives*, 12(29) (see <https://epaa.asu.edu/ojs/article/view/184>).
- Chulu, B. W., & Sireci, S. G. (2011). Importance of equating high-stakes educational measurements. *International Journal of Testing*, 11, 38-52.
- Crespo Cruz, E. J., Immanuel, A., Keller, L. A., Ketan, McIntee, K., Mena Serrano, F. J., Sireci, S. G., Smith, N., Suárez-Álvarez, J., Wells, C. S., Woodland, R., & Zenisky, A. L. (2024). Commentary: What is truly foundational? *Educational Measurement: Issues and Practice*.
- Crotts-Roohr, K., & Sireci, S. G., (2017). Evaluating computer-based test accommodations for English learners. *Educational Assessment*, 22, 35-52. DOI: 10.1080/10627197.2016.1271704.
- Crotts, K., Sireci, S. G., & Zenisky, A. L. (2012). Evaluating the content quality of a multistage-adaptive test. *Journal of Applied Testing Technology* 13(1), 1-26.
- Crotts, K., Sireci, S. G., & Zenisky, A. L., & Lee, X. (2013). Estimating measurement precision in reduced-length multistage adaptive testing. *Journal of Computerized Adaptive Testing*, 1, 67-87.
- Davison, M.L., & Sireci, S. G. (2000). Multidimensional scaling. In H.E.A. Tinsley & S. Brown (Eds.), *Handbook of multivariate statistics and mathematical modeling* (pp. 325-349). Washington, DC: American Psychological Association.
- Diao, H., & Sireci, S. G. (2018). Item response theory-based methods for estimating classification accuracy and consistency. *Journal of Applied Testing Technology*, 19, 20-25.
- Doğan, N., & Sireci, S. G. (2017). Interview with Stephen G. Sireci on validity. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi* [Journal of Measurement and Evaluation in Education and Psychology], 8,158-168

Publications (continued)

- Faulkner-Bond, M. F., Wolf, W. K., Wells, C. S., & Sireci, S. G. (2018). Exploring the factor structure of a K-12 English language proficiency assessment. *Language Assessment Quarterly*. DOI: 10.1080/15434303.2017.1419247.
- Faulkner-Bond, M. F., & Sireci, S. G. (2015). Validity issues in assessing linguistic minorities. *International Journal of Testing*, 15, 114-135.
- Feng, W., Ghosh, A., Sireci, S., & Lan. A. (2023). Balancing test accuracy and security in computerized adaptive testing. *arXiv preprint arXiv:2305.18312*. Available at <https://arxiv.org/abs/2305.18312>
- Feng, W., Tran, P., Sireci, S., & Lan. A. (2025). Reasoning and sampling-augmented MCQ difficulty prediction via LLMs. *arXiv preprint arXiv:2503.08551*. Available at <https://arxiv.org/pdf/2503.08551>
- Ghosh, A., Feng, W., Sireci, S., & Lan. A. (2024). Fully data-driven computerized-adaptive testing. In H. Jiao & R. Lissitz (Eds.), *Machine learning, natural language processing, and psychometrics*. Charlotte, NC: Information Age Publishing.
- Gökçe, S., Berberoglu, G., Sireci, S.G., & Wells, C. S. (2021). Linguistic distance and translation DIF on TIMSS mathematics assessment items. *Journal of Psychoeducational Assessment*. <https://doi.org/10.1177/07342829211010537>.
- Gomez, J., Sireci, S., Padilla, J. L., Hidalgo, M. D., and Benítez, I. (2018). Differential Item Functioning: Beyond validity evidence based on internal structure. *Psicothema*, 30, 104-109.
- Gordon, E. W., Sireci, S. G., Armour-Thomas, E., Baker, E. L., Everson, H. T., & Tucker, E. M. (2025). Handbook for Assessment in the Service of Learning series preface. In E. M. Tucker, E. Armour-Thomas, & E. W. Gordon (Eds.), *Handbook for assessment in the service of learning*, Volume I-III (pp. 13-17). Amherst, MA: University of Massachusetts Amherst Libraries.
- Green, P., & Sireci, S. G. (1999). Legal and psychometric issues in testing students with disabilities. *Journal of Special Education Leadership*, 12(2), 21-29.
- Greiff, S., Iliescu, D., Ketan, & Sireci, S. G. (2024). Challenges for psychological assessment. *European Journal of Psychological Assessment*, 40(4), 269-271.
- Hambleton, R. K., & Sireci, S. G. (1997). Future directions for norm-referenced and criterion-referenced educational assessments. *International Journal of Educational Research*, 27 (5), 379-393.
- Hambleton, R. K., Sireci, S. G., & Robin, F. (1999). Adapting credentialing exams for use in multiple languages. *CLEAR Exam Review*, 10 (2), 24-28.
- Hambleton, R. K., Sireci, S. G., & Smith, Z. (2009). Evaluating NAEP achievement levels in the context of international assessments. *Applied Measurement in Education*, 22, 376-393.

Publications (continued)

- Han, K., Wells, C. S., & Sireci, S. G. (2012). The impact of multidirectional item parameter drift on IRT scaling coefficients and proficiency estimates. *Applied Measurement in Education, 25*, 97-117.
- Hattie, J., Baker, E., & Sireci, S. (2025). Mind Frames for Improving Educational Assessment. In E. Tucker, E. Armour-Thomas, & E.W. Gordon (Eds.). *Handbook for Assessment in the Service of Learning: Volume I* (pp. 255-289). Amherst, MA: University of Massachusetts Library Press. https://www.assessment-for-learning.org/_files/ugd/e901ef_d6327b296c6a4304aa20a2c78461be26.pdf.
- Hauger, J. B, & Sireci, S. G. (2008). Detecting differential item functioning across examinees tested in their dominant language and examinees tested in a second language. *International Journal of Testing, 8*, 237-250.
- Huff, K. L., & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice, 20* (3), 16-25.
- Huff, K. L., Koenig, J. A., Treptau, M. S., & Sireci, S. G. (1999). Validity of MCAT scores for predicting clerkship performance of medical students grouped by sex and ethnicity. *Academic Medicine, 74* (10, supplement), S41-S44.
- Iliescu, D., Bartram, D., Zeinoun, P., Ziegler, M., Elosua, P., Sireci, S. G., Geisinger, K., Odendaal, A., Twing, J., & Camara, W. (2024). Test adaptation reporting standards (TARES): Reporting test adaptations. *International Journal of Testing*. DOI: 10.1080/15305058.2023.2294266.
- Kaira, L. T., & Sireci, S. G. (2010). Evaluating content validity in multistage adaptive testing. *CLEAR Exam Review, 21*(2), 15-23.
- Kaira, L. T., & Sireci, S. G. (2017). Using item mapping to evaluate alignment between curriculum and assessment. *Revista Examen 1*, 49-71.
- Karantonis, A., & Sireci, S. G. (2006). The bookmark standard setting method: A literature review. *Educational Measurement: Issues and Practice, 25* (1), 4-12.
- Keller, L. A. & Sireci, S.G. (2005). Equating 21st century licensure and certification tests. *CLEAR Exam Review, 16*(2), 16-23.
- Keller, L. A., Swaminathan, H., & Sireci, S. G. (2003). Evaluating scoring procedures for context-dependent item sets. *Applied Measurement in Education, 16*, 207-222.
- Koljatic, M., Silva, M., & Sireci, S. G. (2021). College admission tests and social responsibility. *Educational Measurement: Issues and Practice*. <https://doi.org/10.1111/emip.12425>.
- Koljatic, M., Silva, M., & Sireci, S. G. (2021). College admission tests and social responsibility: A response to comments. *Educational Measurement: Issues and Practice*.
- Koenig, J. A., Sireci, S. G., & Wiley, A. (1998). Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Academic Medicine, 73*, 65-76.

Publications (continued)

- Lee, D., Buzick, H., Sireci, S. G., Lee, M., & Laitusis, C. (2021). Embedded Accommodation and Accessibility Support Usage on a Computer-Based Statewide Achievement Test. *Practical Assessment, Research & Evaluation, 26*(25). Available online: <http://pareonline.net/getvn.asp?v=26&n=25>
- Lee, J. J., Lee, M., Wells, C. S., & Sireci, S. G. (2016). A unified factor-analytic approach to the detection of item and test bias: Illustration with the effect of providing calculators to students with dyscalculia. *The Quantitative Methods for Psychology, 12*, 9-29.
- Lewis, J. L., Lim, H., Padellaro, F., Sireci, S. G., & Zenisky, A.L. (2022). *Setting and validating multiple standards on a multistage-adaptive test. Educational Measurement: Issues & Practice, 41*(1), 80–92.
- Lewis, J. L., & Sireci, S. G. (2022). Digital module 30: Validity and educational testing: Purposes and uses of educational tests. *Educational Measurement: Issues and Practice, 41*(4), 81-82.
- Li, X., & Sireci, S. G. (2013). A new method for analyzing content validity data using multidimensional scaling. *Educational & Psychological Measurement, 73*, 365-385. <https://doi.org/10.1177/001316441247>.
- Lim, H., & Sireci, S. G. (2017). Linking TIMSS and NAEP assessments to evaluate international trends in achievement. *Educational Policy Analysis Archives, 25*(11), <http://dx.doi.org/10.14507/epaa.25.2682>
- Luecht, R. L., & Sireci (2011). A review of models for computer-based testing. *Research report 2011-2012*. New York: The College Board.
- Marland, J., Harrick, M., & Sireci, S. G. (2019). Student assessment opt out and the impact on value-added measures of teacher quality. *Educational and Psychological Measurement, 80*, 365-388. DOI: 10.1177/0013164419860574.
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessments, and instruction, *Review of Educational Research 4*, 1332-1361.
- Militello, M., Schweid, J., & Sireci, S. G. (2010). Formative assessment systems: evaluating the fit between school districts' needs and assessment systems' characteristics, *Educational Assessment, Evaluation, and Accountability, 29*-52.
- Meara, K. P., Robin, F., & Sireci, S. G. (2000). Using multidimensional scaling to assess the dimensionality of dichotomous item data. *Multivariate Behavioral Research, 35* (2), 229-259.
- Meara, K. P., Hambleton, R. K., & Sireci, S. G. (2001). Setting and validating standards on professional licensure and certification exams: A survey of current practices. *CLEAR Exam Review, 12* (2), 17-23.
- Matthews, W. J., Conti, J. M., & Sireci, S. G. (2001). The effects of intercessory prayer, positive visualization, and expectancy on the well-being of kidney dialysis patients. *Alternative Medicine, 7* (5), 42-54.

Publications (continued)

- Noble, T., Sireci, S. G., Wells, C. S., Kachchaf, R. R., Rosebery, A. S., & Wang, Y. C. (2020). Targeted linguistic simplification of science test items for English learners. *American Educational Research Journal, 57*, 2175-2209.
- Nguyen, T.-A. T., Crespo Cruz, E., & Sireci, S. G. (in press). Validity and related measurement concepts. In F. J. Wertz, L. L. Harlow, and R. T. Fouladi (Eds.) *Handbook of Methodology for Psychology in the 21st Century: Progress, Critique, and Future*. Guilford Press.
- O'Donnell, F., & Sireci, S. G. (2018). Score reporting issues for licensure, certification, and admissions programs. In D. Zapata-Rivera (Ed.). *Score reporting: Research and applications* (pp. 77-90). New York: Routledge.
- O'Donnell, F., & Sireci, S. G. (2021). Language matters: Teacher and parent perceptions of achievement labels from educational tests. *Educational Assessment*. DOI: 10.1080/10627197.2021.2016388
- Ong, S. L., & Sireci, S. G. (2008). Using bilingual students to link and evaluate different language versions of an exam. *US-China Education Review, 5*, 37-46.
- O'Neil, T., Sireci, S. G., & Huff, K. F. (2004). Evaluating the consistency of test content across two successive administrations of a state-mandated science assessment. *Educational Assessment, 9*, 129-151.
- Padilla, J., Benitez, I., Sireci, S. G., & Flores-Galaz, M. (2012). Evaluating structural equivalence in psychological questionnaires using multidimensional scaling. *Cross-Cultural Research, 46*, 348-365.
- Pitoniak, M. J., Sireci, S. G., & Luecht, R. M. (2002). A multitrait-multimethod validity investigation of scores from a professional licensure exam. *Educational and Psychological Measurement, 62*, 498-516.
- Randall, J., Sireci, S. G., Li, X., & Kaira, L. (2013). Evaluating the comparability of paper- and computer-based science tests across sex and SES subgroups. *Educational Measurement: Issues and Practice, 31(4)*, 2-12.
- Rios, Joseph A., & Sireci, S. G., (2014). Guidelines versus practices in cross-lingual assessment: A disconcerting disconnect. *International Journal of Testing, 14*, 289-312.
- Robin, F., Sireci, S. G., & Hambleton, R. K. (2003). Evaluating the equivalence of different language versions of a credentialing exam. *International Journal of Testing, 3*, 1-20.
- Sireci, S. G. (1997). Problems and issues in linking tests across languages. *Educational Measurement: Issues and Practice, 16(1)*, 12-19.
- Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment, 5*, 299-321.

Publications (continued)

- Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, 45, 83-117.
- Sireci, S. G. (2000). Recruiting the next generation of measurement professionals. *Educational Measurement: Issues and Practice*, 19(4), 5-9.
- Sireci, S. G. (2001). Standard setting using cluster analysis. In C.J. Cizek (Ed.), *Standard setting: Concepts, methods, and perspectives* (pp. 339-354). Mahwah, NJ: Lawrence Erlbaum.
- Sireci, S. G. (2003). Content validity. *Encyclopedia of psychological assessment* (pp. 1075-1077). London: Sage.
- Sireci, S. G. (2003). Validity. *Encyclopedia of psychological assessment* (pp. 1067-1069). London: Sage.
- Sireci, S. G. (2004). Computerized-adaptive testing: An introduction. In J. Wall and G. Walz (Eds.), *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 685-694), Greensboro, NC: CAPS Press.
- Sireci, S. G. (2005). Unlabeling the disabled: A perspective on flagging scores from accommodated test administrations. *Educational Researcher*, 34(1), 3-12.
- Sireci, S. G. (2005). Using bilinguals to evaluate the comparability of different language versions of a test. In R.K. Hambleton, P. Merenda, & C. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 117-138). Hillsdale, NJ: Lawrence Erlbaum.
- Sireci, S. G. (2005). The most frequently unasked questions about testing. In R. Phelps (Ed.), *Defending standardized testing* (pp. 111-121). Mahwah, NJ: Lawrence Erlbaum.
- Sireci, S. G. (2005). Validity theory and applications. *Encyclopedia of statistics in the behavioral sciences* (Volume 4, pp. 2103-2107). West Sussex, UK: John Wiley & Sons.
- Sireci, S. G. (2006). Content validity. In N. J. Salkind (Ed.) *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: Sage.
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, 36(8), 477-481.
- Sireci, S. G. (2007). Are educational tests inherently evil? *Nonpartisan Education Review / Essays*, 3(4). <https://nonpartisaneducation.org/Review/Essays/v3n4.htm>
- Sireci, S. G. (2008). Standardized testing is useful. In D. A. Henningfeld (Ed.). *At issue: Standardized testing* (pp. 10-16). Detroit: Thompson Gale.
- Sireci, S. G. (2008). Validity issues in accommodating reading tests. *Educators and Education (Pendidik dan Pendidikan)*, 23, 81-110.

Publications (continued)

- Sireci, S. G. (2009). No more excuses: New research on assessing students with disabilities. *Journal of Applied Testing Technology*, 10 (2). Available at <http://www.testpublishers.org/Documents/Special%20Issue%20article%201%20.pdf>.
- Sireci, S. G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. In R. Lissitz (Ed.), *The Concept of Validity: Revisions, New Directions and Applications* (pp. 19-37). Charlotte, NC: Information Age Publishing Inc.
- Sireci, S. G. (2010). National Council on Measurement in Education. In N. Salkind (Ed.) *Encyclopedia of research design*. Thousand Oaks, CA: Sage.
- Sireci, S. G. (2010). Validity issues and empirical research on translating educational achievement tests. In P. Winter (Ed.), *Evaluating the comparability of scores from achievement test variations* (pp. 153-183). Washington, DC: Council of Chief State School Officers.
- Sireci, S. G. (2011). Evaluating test and survey items for bias across languages and cultures. In D. Matsumoto and F. van de Vijver (Eds.) *Cross-cultural research methods in psychology* (pp. 216-240). Oxford, UK: Oxford University Press.
- Sireci, S. G. (2013). Agreeing on validity arguments. *Journal of Educational Measurement*, 50, 99-104.
- Sireci, S. G. (2013). Standard Setting in an international context: Introduction to the special issue. *International Journal of Testing*, 13, 2-3.
- Sireci, S. G. (2013). Trafność symulacyjnych gier jako narzędzi oceny. *Personel Plus*, 08(69), 8-11. [Validating simulation games as assessment tools. Published in Polish.]
- Sireci, S. G. (2015). Beyond ranking of nations: Innovative research on PISA. *Teachers College Record*, 117(1), 1-8.
- Sireci, S. G. (2015). A theory of action for validation. In H. Jiao & R. Lissitz (Eds.). *The next generation of testing: Common core standards, Smarter-Balanced, PARCC, and the nationwide testing movement* (pp. 251-269). Charlotte: Information Age Publishing Inc.
- Sireci, S. G. (2016). Comments on valid (and invalid?) commentaries. *Assessment in Education: Principles, Policy & Practice*, 23, 319-321.
- Sireci, S. G. (2016). On the validity of useless tests. *Assessment in Education: Principles, Policy & Practice*, 23, 226-235.
- Sireci, S. G. (2016). Prologue: A conference retrospective. In C. Wells & M. F. Bond (Eds., pp. v-vii). *Educational measurement: From foundations to future*. Guilford Press.
- Sireci, S. G. (2016). Using technology to enhance assessment: Comments on four chapters. In F. Drasgow (Ed.) *Technology and Testing* (pp. 104-108). New York: Routledge.
- Sireci, S. G. (2016, September). Want to understand your child's test score report? Here's what to ignore. *The Conversation*. <https://doi.org/10.64628/AAI.f37uje4mr>.

Publications (continued)

- Sireci, S. G. (2020). Educating the measurement community: Introduction to the special issue “Using Educational Assessments to Educate,” *Educational Measurement: Issues and Practice*, 39, 3, 70–71.
- Sireci, S. G. (2020). Standardization and UNDERSTANDARDIZATION in educational assessment, *Educational Measurement: Issues and Practice*, 39(3) 100-105. DOI: 10.1111/emip.12377.
- Sireci, S. G. (2020) "De-“Constructing” Test Validation," *Chinese/English Journal of Educational Measurement and Evaluation*, 1. 教育测量与评估双语季刊: Available at: <https://www.ce-jeme.org/journal/vol1/iss1/3> [DOI <https://doi.org/10.59863/CKHH8837>].
- [Also available in Chinese: Sireci, S. G. (2020) "去“构念”的测验效度验证," *Chinese/English Journal of Educational Measurement and Evaluation*, 1. 教育测量与评估双语季刊. Available at: <https://www.ce-jeme.org/journal/vol1/iss1/4> [DOI <https://doi.org/10.59863/EOEL9922>].
- Sireci, S. G. (2021). Valuing educational measurement. *Educational Measurement: Issues and Practice*, 40(1), 7-16. DOI: 10.1111/emip.12415.
- Sireci, S. G. (2021, September). 6 big changes in standardized tests – including less focus on grading students and more on learning. *The Conversation*. DOI: [10.64628/AAI.cggm9jvn5](https://doi.org/10.64628/AAI.cggm9jvn5).
- Sireci, S. G. (2022). In memoriam: Ronald K. Hambleton. *Psychological Test and Assessment Modeling, Volume 64*,(2), 210-214.
- Sireci, S. G. (2022). Ronald K. Hambleton (1943–2022): Setting the standard for measurement excellence. *Educational Measurement: Issues and Practice*, 41(4), 6-9.
- Sireci, S. G. (2022). National Council on Measurement in Education. In B. Frey (Ed.) *Encyclopedia of research design* (2nd Edition), 1010-1011. Thousand Oaks, CA: Sage. DOI: <https://dx.doi.org/10.4135/9781071812082.n367>.
- Sireci, S. G. (2022). Validating technology-based assessments. In (J. Weiner & S. Sireci, editors). *Guidelines for technology-based assessments* (pp. 86-91). Available at <https://www.intestcom.org/page/28>.
- Sireci, S. G., Araneda, S., & McIntee, K. (2025). Social justice in educational assessment: A blueprint for the future. In S. G. Sireci, E. M. Tucker, & E.W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume II: Reconceptualizing assessment to improve learning* (pp. 187-210). Amherst, MA: University of Massachusetts Amherst Libraries.
- Sireci, S. G., & Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language Testing*, 20, 148-166.
- Sireci, S. G., & Banda, E., & Wells, C. S. (2018). Promoting valid assessment of students with disabilities and English learners. In Elliott, S. N., Kettler, R. J, Beddow, P. A., & Kurz, A., (Eds.), *Handbook of Accessible Instruction and Testing Practices: Issues, Innovations, and Application* (pp.231-246). Sage.

Publications (continued)

- Sireci, S. G., & Benítez Baena, I. (2023). Evidence for test validation: A guide for practitioners. *Psicothema*, 35(3) 217-226. Available at <https://www.psicothema.com/pi?pii=4805>.
- Sireci, S. G. & Berberoglu, G. (2000). Using bilingual respondents to evaluate translated-adapted items. *Applied Measurement in Education*, 35 (2), 229-259.
- Sireci, S. G. & Biskin, B. H. (1992). A survey of national professional licensure examination programs. *CLEAR Exam Review*, 3, 21-25.
- Sireci, S. G., & Botha, S. (2020). Timing considerations in test development and administration. In M. Margolis & J. Feinberg (Eds). *Integrating timing considerations to improve testing practices* (pp. 32-46). Routledge Press.
- Sireci, S. G., & Clauser, B. E. (2001). Issues to be considered in setting standards on computerized-adaptive tests. In C.J. Cizek (Ed.), *Standard setting: Concepts, methods, and perspectives* (pp. 355-369). Mahwah, NJ: Lawrence Erlbaum.
- Sireci, S. G. & Crabtree, D. (2025). Validity for assessments intended to serve learners. In S. G. Sireci, E. M. Tucker, & E. W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume II: Reconceptualizing assessment to improve learning* (pp. 167-186). Amherst, MA: University of Massachusetts Amherst Libraries.
- Sireci, S. G., Crespo Cruz, E. J, (in press). Content validity (validity evidence based on test content). In S. Sinharay (Ed.), *Encyclopedia of Measurement in Social Sciences, Second Edition*. Amsterdam: Elsevier.
- Sireci, S. G., Crespo Cruz, E., Suárez-Álvarez, J., & Rodriguez, G. (2025). Understanding UNDERSTANDARDization research. In *Socioculturally Responsive Assessment: Implications for Theory, Measurement, and Systems-Level Policy* (R. Bennett, L. Darling-Hammond, & A. Badrinarayan, Eds., pp. 415-433). New York: Routledge.
- Sireci, S. G., DeLeon, B., & Washington, E. (2002, Spring). Improving teachers of minority students' attitudes towards and knowledge of standardized tests. *Academic Exchange Quarterly*, 162-167.
- Sireci, S. G., & Faulkner-Bond (2014). Validity evidence based on test content. *Psicothema*, 26, 100-107. doi: 10.7334/psicothema2013.256.
- Sireci, S. G. & Faulkner-Bond, M. F. (2015). Promoting validity in the assessment of English learners and other linguistic minorities. *Review of Research in Education*, 39, 215-252.
- Sireci, S. G., & Faulkner-Bond, M. F. (2016). The times they are a'changing, but the song remains the same: Future issues and practices in test validation. In C. Wells & M. F. Bond (Eds., pp. 435-448). *Educational measurement: From foundations to future*. Guilford Press.
- Sireci, S. G., & Forte, E., (2012). Informing in the information age: How to communicate measurement concepts to education policy makers. *Educational Measurement: Issues and Practice*, 31(2), 27-32.

Publications (continued)

- Sireci, S. G. & Gándara, M. F. (2016). Testing in educational and developmental settings. In F. Leong et al. (Eds.). *International Test Commission handbook of testing and assessment* (pp. 187-202). Oxford University Press.
- Sireci, S. G., & Garcia, A. (2018). The Angoff method for setting standards on educational tests. In B. Frey (Ed.). *The Sage Encyclopedia of Educational Measurement, Research, and Evaluation*. <http://dx.doi.org/10.4135/9781506326139.n44>
- Sireci, S. G. & Geisinger, K. F. (1992). Analyzing test content using cluster analysis and multidimensional scaling. *Applied Psychological Measurement, 16*, 17-31.
- Sireci, S. G., & Geisinger K. F. (1995). Using subject matter experts to assess content representation: An MDS analysis. *Applied Psychological Measurement, 19*, 241-255.
- Sireci, S. G., & Geisinger, K. F. (1998). Equity issues in employment testing. In J.H. Sandoval, C. Frisby, K.F. Geisinger, J. Scheuneman, & J. Ramos-Grenier (Eds.), *Test interpretation and diversity* (pp. 105-140). American Psychological Association: Washington, D.C.
- Sireci, S.G., & Green, P.C. (2000). Legal and psychometric criteria for evaluating teacher certification tests. *Educational Measurement: Issues and Practice, 19(1)*, 22-31, 34.
- Sireci, S. G., & Greiff, S. (2019). On the importance of educational tests. *European Journal of Psychological Assessment, 35*, 297-300. <https://doi.org/10.1027/1015-5759/a000549>.
- Sireci, S.G., & Hambleton, R.K. (2009). Mission: Protect the public: Licensure and certification testing in the 21st century. In R. Phelps (Ed.). *Correcting fallacies about educational and psychological testing*. Washington, DC: American Psychological Association.
- Sireci, S. G., Hambleton, R. K., & Pitoniak, M. J. (2004). Setting passing scores on licensure exams using direct consensus. *CLEAR Exam Review 15(1)*, 21-25.
- Sireci, S. G., Han, K. T., & Wells, C. S. (2008). Methods for evaluating the validity of test scores for English language learners. *Educational Assessment, 13*, 108-131.
- Sireci, S. G., Harter, J., Yang, Y., & Bhola, D. (2003). Evaluating the equivalence of an employee attitude survey across languages, cultures, and administration formats. *International Journal of Testing, 3*, 129-150.
- Sireci, S. G., Hauger, J. B, Wells, C. S., Shea, C., & Zenisky, A. L. (2009). Evaluation of the standard setting on the 2005 grade 12 National Assessment of Educational Progress mathematics test. *Applied Measurement in Education, 22*, 339-358.
- Sireci, S. G., & Khaliq, S. N. (2002). NCME members' suggestions for recruiting new measurement professionals. *Educational Measurement: Issues and Practice, 21(3)*, 19-24.
- Sireci, S. G., & Kingston, N. (2025). Removing the “psycho” from education metrics. In S. G. Sireci, E. M. Tucker, & E. W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume II: Reconceptualizing assessment to improve learning* (pp. 299-326). Amherst, MA: University of Massachusetts Amherst Libraries.

Publications (continued)

- Sireci, S. G., Longe, B., Suárez-Álvarez, J., & Oliveri, M. E. (2024). Analyzing the dimensionality of O*NET cognitive ability Ratings to inform assessment design. *Education Sciences*, *14*, 1202. <https://doi.org/10.3390/educsci14111202>.
- Sireci, S. G., & Marland, J. (2017). Measurement. In D. Wyse, N. Selwyn, E. Smith, & L. Sutter (Eds.) *The BERA/SAGE handbook of educational research* (pp. 659-676). Los Angeles: Sage.
- Sireci, S. G., & Meijer, R. (2009). Editor's introduction. *International Journal of Testing*, *9*, 1-2.
- Sireci, S. G., & Mullane, L. A. (1994). Evaluating test fairness in licensure testing: The sensitivity review process. *CLEAR Exam Review*, *5* (2) 22-28.
- Sireci, S. G., & O'Riordan, M. (2020). Comparability issues in assessing individuals with disabilities. In A.I. Berman, E.H. Haertel, & J.W. Pellegrino (Eds.), *Comparability Issues in Large-Scale Assessment: Issues and recommendations* (p. 117-204). Washington, DC: National Academy of Education Press.
- Sireci, S. G., & Padilla, J.-L. (2014). Validating assessments: Introduction to the special issue. *Psicothema*, *26*, 97-99. DOI 10.7334/psicothema2013.255.
- Sireci, S. G., & Parker, P. (2006). Validity on trial: Psychometric and legal conceptualizations of validity. *Educational Measurement: Issues and Practice*, *25*(3), 27-34.
- Sireci, S. G., Patsula, L., & Hambleton, R. K. (2005). Statistical methods for identifying flawed items in the test adaptations process. In R.K. Hambleton, P. Merenda, & C. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 93-115). Hillsdale, NJ: Lawrence Erlbaum.
- Sireci, S. G., & Pitoniak, M. J. (2007). Assessment accommodations: What have we learned from research? In C. C. Laitusis & L. Cook (Eds.) *Large scale assessment and accommodations: What works?* (pp. 53-65). Arlington: Council for Exceptional Children.
- Sireci, S. G., & Randall, J. (2021). Evolving notions of fairness in testing in the United States. In M. Bunch & B. Clauser (Eds.), *The history of educational measurement: Key advancements in theory, policy, and practice* (pp. 111-135). New York: Routledge
- Sireci, S. G., Randall, J., & Zenisky, A. (2012). Setting valid performance standards on educational tests. *CLEAR Exam Review*, *23*(2), 18-27.
- Sireci, S. G., & Rios, J. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation*, *19*, 170-187.
- Sireci, S. G., Rios, J. A., & Powers, S. (2016). Comparing test scores from tests administered in different languages. In N. Dorans & L. Cook (Eds.) *Fairness in educational assessment and measurement* (pp. 181-202). New York: Routledge.
- Sireci, S. G., Robin, F., & Patelis, T. (1999). Using cluster analysis to facilitate standard setting. *Applied Measurement in Education*, *12*, 301-325.

Publications (continued)

- Sireci, S. G., Robin, F., Meara, K., Rogers, H. J., & Swaminathan, H. (2000). An external evaluation of the 1996 Grade 8 NAEP Science Framework. In N. Raju, J.W. Pellegrino, M.W. Bertenthal, K.J. Mitchell & L.R. Jones (Eds.), *Grading the nation's report card: Research from the evaluation of NAEP* (pp. 74-100). Washington, D.C.: National Academy Press.
- Sireci, S. G., & Rodriguez, G. (2022). Validity in educational testing. In S. Brookhart (Ed.). *Routledge Encyclopedia of Education* New York: Routledge.
<https://doi.org/10.4324/9781138609877-REE180-1>. Available at
<https://www.academia.edu/?oa=2091560105>.
- Sireci, S. G., Rogers, H.J., Swaminathan, H., Meara, K., & Robin, F. (2000). Appraising the dimensionality of the 1996 Grade 8 NAEP Science Assessment Data. In N. Raju, J.W. Pellegrino, M.W. Bertenthal, K.J. Mitchell & L.R. Jones (Eds.), *Grading the nation's report card: Research from the evaluation of NAEP* (pp. 101-122). Washington, D.C.: National Academy Press.
- Sireci, S. G., Scarpati, S., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75, 457-490.
- Sireci, S. G., & Soto, A. (2016). Validity and accountability: Test validation for 21st-century educational assessments. In H. Braun (Ed.). *Meeting the challenges to measurement in an era of accountability* (pp. 149-167). New York: Routledge.
- Sireci, S. G., & Suárez-Alvárez, J. (2022). Deriving decisions from disrupted data. *Educational Measurement: Issues and Practice*, 41(1), 23-27.
<https://onlinelibrary.wiley.com/doi/abs/10.1111/emip.12499>.
- Sireci, S. G., Suárez-Alvárez, J., Zenisky, A. L., & Oliveri, M. E. (2024). Evolving educational testing to meet students' needs: Design-in-real-time assessment. *Educational Measurement: Issues and Practice*, 43(4), 112-118.
- Sireci, S. G., & Sukin, T. (2013). Test validity. In K. F. Geisinger (Editor-in-chief). *APA handbook of testing and assessment in psychology* (Vol. 1, pp.61-84). Washington, DC: American Psychological Association.
- Sireci, S. G., & Talento-Miller, E. (2006). Evaluating the predictive validity of Graduate Management Admissions Test Scores. *Educational and Psychological Measurement*, 66, 305-317.
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237-247.
- Sireci, S. G., & Tucker, E. M. (2025). Reflections on reconceptualizing assessment to improve learning. In S. G. Sireci, E. M. Tucker, & E. W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume II: Reconceptualizing assessment to improve learning* (pp.)389-392). Amherst, MA: University of Massachusetts Amherst Libraries.
- Sireci, S. G., Wainer, H., & Braun, H. (1998). Psychometrics, overview. In *Encyclopedia of biostatistics*. New York: John Wiley & Sons.

Publications (continued)

- Sireci, S. G., & Wells, C. S. (2010). Evaluating the comparability of English and Spanish video accommodations for English language learners. In P. Winter (Ed.), *Evaluating the comparability of scores from achievement test variations* (pp. 33-68). Washington, DC: Council of Chief State School Officers.
- Sireci, S. G., Wiley, A., & Keller, L. A. (2002). An empirical evaluation of selected multiple-choice item writing guidelines. *CLEAR Exam Review*, 13(2), 20-26.
- Sireci, S. G., Yang, Y., Harter, J., & Ehrlic, E. (2006). Evaluating guidelines for test adaptations: A methodological analysis of translation quality. *Journal of Cross-Cultural Psychology*, 37, 557-567.
- Sireci, S. G., Zanetti, M. L., & Berger, J. B. (2003). Recent and anticipated changes in postsecondary admissions: A survey of New England colleges and universities. *Review of Higher Education*, 26, 323-342.
- Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In S.M. Downing and T.M. Haladyna (Eds.), *Handbook of Testing* (pp. 329-347). Mahwah, NJ: Lawrence Erlbaum.
- Sireci, S. G., & Zenisky, A. L. (2016). Computerized innovative item formats: Achievement and Credentialing. In S. Lane, T. Haladyna, & M. Raymond (Eds.). *Handbook of test development* (pp. 313-334). Washington, DC: National Council on Measurement in Education.
- Sukin, T., Sireci, S. G., & Ong, S. L. (2015). Using bilingual examinees to evaluate the comparability of test structure across different language versions of a mathematics exam. *Actualidades en Psicología*, 29 (19)131-139. DOI: <http://dx.doi.org/10.15517/ap.v29i119.19244>.
- Suárez-Ávarez, J., Oliveri, M.E., Zenisky, A., and Sireci, S.G (2024). Five key actions for redesigning adult skills assessments from learners, employees, and educators. *Journal for Research on Adult Education*, <https://doi.org/10.1007/s40955-024-00288-8>.
- Swaminathan, H., Hambleton, R. K., Sireci, S. G., Xing, D., & Rizavi, S. M. (2003). Small sample estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates. *Applied Psychological Measurement*, 27, 27-51.
- Tran, P., Feng, W., Sireci, S. G., McNichols, H., & Lan, A. (2025). Using item features to calibrate educational test items: Comparing artificial intelligence and classical approaches. *American Journal of Education and Learning*, 10(2), 178–189. <https://doi.org/10.55284/ajel.v10i2.1543>
- Tucker, E. M., & Sireci, S. G. (2025). Reconceptualizing assessment to improve learning. In S. G. Sireci, E. M. Tucker, & E. W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume II: Reconceptualizing assessment to improve learning* (pp. 19-34). Amherst, MA: University of Massachusetts Amherst Libraries.
- Wainer, H., & Sireci, S. G. (2005). Item and test bias. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement volume 2*, 365-371. San Diego: Elsevier.

Publications (continued)

- Wainer, H., Sireci, S. G., & Thissen, D. (1991). Differential testlet functioning: Definitions and detection. *Journal of Educational Measurement*, 28, 197-219.
- Weiner, J., & Sireci, S. G. (2023). New guidelines for technology based assessments. *The Industrial Psychologist*, 60 (4), 604. <https://www.siop.org/Research-Publications/Items-of-Interest/ArtMID/19366/ArticleID/7465/New-Guidelines-for-Technology-Based-Assessment>
- Wells, C. S., Baldwin, S., Hambleton, R. K., Sireci, S. G., Karantonis, A. & Jirka, S. (2009). Evaluating score equity assessment for state NAEP. *Applied Measurement in Education*, 22, 394-408.
- Wells, C. S., & Sireci, S. G. (2020). Evaluating random and systematic error in student growth percentiles. *Applied Measurement in Education*. DOI:10.1080/08957347.2020.1789139.
- Wells, C. S., Sireci, S. G., & Han, K. T. (2008). Identifying item parameter drift in multistage adaptive tests. *CLEAR Exam Review*, 19(1), 14-21.
- Woo, S.-E., Willie, B., & Sireci, S. G. (2022). Introduction to the International Journal of Testing special issue on equity and fairness in testing and assessment in school admissions. *International Journal of Testing*, 22, 1-4, DOI:10.1080/15305058.2021.2019753
- Ying, L., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, 26(4), 29-37.
- Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2002). Identification and evaluation of local item dependencies in the Medical College Admissions Test. *Journal of Educational Measurement*, 39, 291-309.
- Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2009). Getting the message out: An evaluation of NAEP score reporting practices with implications for disseminating test results. *Applied Measurement in Education*, 22, 359-375.
- Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15, 337-362.
- Zenisky, A. L., & Sireci, S. G. (2025). Purposeful design for useful tests: Considering choices in multistage-adaptive testing. In D. Yan, A. von Davier, & D. Weiss (Eds.) *Research for practical issues and solutions in multistage-adaptive testing* (pp. 17-35). Oxfordshire: Taylor and Francis.

Book Reviews

- Copella, J., & Sireci, S. G. (2013). Review of Cutscores: A manual for setting standards of performance on educational and occupational tests. *Applied Measurement in Education*, 26, 73-76.
- Sireci, S. G. (1997). Review of the Twelfth Mental Measurements Yearbook. *Journal of Educational Measurement*, 34, 184-187.

- Sireci, S. G. (2000). Review of Modern Methods for Business Research. *Structural Equation Modeling*, 7(3), 484-488.
- Sireci, S. G. (2000). Review of The New Rules of Measurement: What Every Psychologist and Educator Should Know. *Applied Psychological Measurement*, 24, 284-286.
- Sireci, S. G. (2003). Review of Modern *Multidimensional Scaling: Theory and Applications*. *Journal of Educational Measurement*, 40, 277-280.
- Sireci, S. G., & Rios, J. A. (2012). Review of Uneducated Guesses: Using evidence to uncover misguided educational policies. *Journal of Educational Measurement*, 49, 330-334.

Magazines/Newsletters

- Copella, J. M., & Sireci, S. G. (2010). The consequences of educational assessment: Who should evaluate what and why. *NCME Newsletter*, 18 (1) 5-7. Available at http://www.ncme.org/pubs/pdf/vol_18_num_1_v2.pdf
- Sireci, S. G. (1999). Guidelines for adapting certification tests for use across multiple languages. *PES News: A publication of the Professional Examination Service*, 19(2), 8-9.
- Sireci, S. G. (2002, June). No psychometrician left behind. *NCME Newsletter*, 10 (2), 6.
- Sireci, S. G. (2004, Spring). ACLS, SABES, and UMASS: Perfect Together! *Adventures in Assessment*, 16, 39-42.
- Sireci, S. G. (2007, June). NCLB: The good, the bad, and the ugly unanswered research questions. *NERA Researcher*, 45(2). <https://www.nera-education.org/docs/NERAResearcherJune07.pdf>
- Sireci, S. G. (2007, September). NCLB: Passionate research. *NERA Researcher*, 45(3). <https://www.nera-education.org/docs/NERAResearcherSeptember07.pdf>
- Sireci, S. (2019, June). You, me, and NCME! *NCME Newsletter*. National Council on Measurement in Education. <https://www.ncme.org/blogs/megan-welsh1/2019/06/30/you-me-and-ncme>.
- Sireci, S. (2019, September). 2020 vision. *NCME Newsletter*. National Council on Measurement in Education. <https://www.ncme.org/blogs/megan-welsh1/2019/09/30/from-the-president-2020-vision>.
- Sireci, S. (2019, December). Communities and communication. *NCME Newsletter*. National Council on Measurement in Education. <https://www.ncme.org/blogs/megan-welsh1/2019/12/20/from-the-president-communities-and-communication>.
- Sireci, S. (2020, April). Perspectives on our past, present, and future. *NCME Newsletter*. National Council on Measurement in Education. <https://www.ncme.org/blogs/megan-welsh1/2020/04/09/from-the-president-perspectives-on-our-past-presen>.
- Sireci, S. G. (2022, December). The International Testing Community. *Testing International*, 48, 2-4.
- Sireci, S. G. (2023, June). The hibernation is over: ITC on the move. *Testing International*, 49, 2-7.

- Sireci, S. G. (2023, December). Presidential Promulgations and Ponderings. *Testing International*, 50, 2-5.
- Sireci, S. G. (2024, June). Five Loaves, Two Fish, and the International Test Commission. *Testing International*, 51, 2-3.
- Sireci, S. G., Kingston, N., & ChatGPT (2024, June). The Contributions of Jerry Garcia to Criterion-Referenced Testing: A Cautionary Tale of Chat GPT. *Testing International*, 51, 11-12.

Selected Presentations

Selected Keynote Addresses

- Sireci, S. G. (1999, August). *Adapting and Evaluating Tests for Use Across Multiple Languages and Cultures*. Keynote address delivered at the annual meeting of the Association of Test Publishers, Boston, MA.
- Sireci, S. G. (2006, July). *Ensuring validity in cross-lingual assessments: Issues, methods, and future research directions*. Keynote presentation delivered at the biennial conference of the International Test Commission, Brussels, Belgium.
- Sireci, S. G. (2007, February). *Social Considerations and Equity Issues in Testing*. Keynote delivered at the X Congreso de Metodología de las Ciencias Sociales y de la Salud, University of Barcelona, Spain.
- Sireci, S. G. (2011, May). *Assessing adult learners in the 21st century: Challenges, innovations, and future directions*. National Training and Technical Assistance Assessment Institute, Washington, DC.
- Sireci, S. G. (2011, May). *Computer-based testing for teacher licensure/certification: Innovations, challenges, and a look to the future*. 2011 Praxis Client Conference, Princeton, NJ.
- Sireci, S. G., (2012, May). *How We Should Measure Teaching "Effectiveness" - Or Should We?* Keynote presentation delivered at the 44th Annual Meeting of the New England Educational Research Association, Portsmouth, NH.
- Sireci, S. G., (2012, July). *What Have We Learned from 100 Years of Validity Theory and Test Validation?* Keynote address delivered at the 8th annual conference of the International Test Commission, Amsterdam.
- Sireci, S. G. (2013, September). *Using 21st-century technology to improve educational assessments*. Keynote Presentation delivered at the XIII Congreso de Metodología de las Ciencias Sociales y de la Salud Tenerife, Spain.
- Sireci, S. G. (2017, July). *How would Messick validate 20th-Century language assessments?* Samuel J. Messick Memorial Award presentation delivered at the Language Testing Research Colloquium, Bogotá, Colombia.
- Sireci, S. G. (2018, June). *Do educational tests do more good than harm? Criticisms, benefits, and evidence*. Keynote Presentation Delivered at ICAP 2018: The 29th International Congress of Applied Psychology, Montreal.

- Sireci, S. G. (2018, July). *21st-century validation procedures for 21st-century tests*. Keynote Presentation Delivered at the biennial conference of the International Test Commission, Montreal.
- Sireci, S. G. (2020, June). *Using computer-based process data to improve assessment validity: Challenges and opportunities*. Closing Keynote for Beyond Results: Paving the Way for the Use of Process Data, Frankfurt (virtual). Available at https://www.youtube.com/watch?v=4tqx4YsIMlw&feature=emb_logo.
- Sireci, S. G. (2020, September). *Psychometricians in the hands of an angry mob*. Presidential address delivered (virtually) at the annual meeting of the National Council on Measurement in Education. Available for viewing at <https://www.youtube.com/watch?v=5X2uYYax9yw>.
- Sireci, S. G. (2020, October). *Understanding UNDERSTANDARDization: Promoting Inclusivity in Educational Assessment*. Presented at Beyond Multiple Choice [virtual]. Available at <https://youtu.be/4tqx4YsIMlw>.
- Sireci, S. G. (2021, October). *Reconceptualizing Educational Measurement*. Presentation delivered at the Educational Assessment and Measurement for Sustainable Development conference, Vietnam National Institute of Educational Sciences, (virtual).
- Sireci, S. G. (2023, November). *Aligning Curriculum, Instruction, & Assessment to Promote Learning: Technological Advances*. Keynote presentation delivered at the Association of Language Testing Europe's 59th Meeting and Conference, Rome.
- Sireci, S. G. (2023, November). *What is "Fairness" in Educational Testing?* Keynote Presentation delivered at the annual conference of the Florida Educational Research Association, Hollywood Beach, FL.
- Sireci, S. G. (2024, July). *The Failure and Future of Psychometrics*. Presidential Address delivered the 2024 Conference of the International Test Commission, Granada, Spain.

American Educational Research Association

- Allalouf, A., & Sireci, S. G. (1998, April). *Detecting the causes of differential item functioning in translated verbal items*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Boyer, M., & Sireci, S. G. (2017, April). Evaluating the dimensionality of tests comprising traditional and technology-enhanced items. Paper presented at the annual meeting of the American Educational Research Association, San Antonio, TX.
- Crehan, K. D., Sireci, S. G., Haladyna, T. M., & Henderson, P. A., (1993, April). *A comparison of testlet reliability for polytomous scoring methods*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.
- Foster, D., Olsen, J. B., Ford, J., & Sireci, S. G. (1997, March). *Administering computerized certification exams in multiple languages: Lessons learned from the international marketplace*. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

- Keller, L. A., Sireci, S. G., & Swaminathan, H. (2001, April). *Alternatives for scoring simulated performance tasks*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Ketan, & Sireci, S. G. (2025, April). *Is rapid and revolutionary also valid? Exploring Validity evidence for large-scale Annual Status of Education Report assessment in India*. Poster presented at the annual meeting of the American Educational Research Association, Denver.
- Li, S., Scarpati, S., & Sireci, S. G. (2004, April). *Test accommodations and students with disabilities: An analysis of the interaction hypothesis*.
- Ma, X., & Sireci, S. G. (2004, April). *An investigation of polytomous scoring of multiple response items on a certification exam*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Martone, A., & Sireci, S. G. (2007, April). *Exploring the impact of teachers' participation in an assessment-standards alignment study*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Nguyen, T.-A. T., & Sireci, S. G. (2025, April). *Examining social consequences of standardized admission tests: A literature review and suggestions for practice*. Poster presented at the annual meeting of the American Educational Research Association, Denver.
- Noble, T., & Sireci, S. G., Wells, C. S., Kachchaf, R. R., Rosebery, A., & Wang, Y. (2016, April). *Targeted linguistic modifications of science test items for English learners*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.
- Pitoniak, M. J., Hambleton, R. K., & Sireci, (2002, April). *Advances in standard setting for professional licensure examinations*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Qi, S., & Sireci, S. G., (1996, April). *Why did they drop out? And who came back? Comparing high school graduates, dropouts, and returnees using NELS:88*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
- Sireci, S. G. (1998, April). *Evaluating content validity using multidimensional scaling*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Sireci, S. G. (2013, April). *Incorporating a Theory of Action into a Validity Argument*. Presentation delivered at the Test Validity Research and Evaluation Special Interest Group of the American Educational Research Association, San Francisco, CA.
- Sireci, S. G. (2021, April). *The role of understandardization in international assessment*. Paper presented at the annual meeting of the American Educational Research Association [virtual].
- Sireci, S. G., & Berberoglu, G. (1997, March). *Evaluating translation DIF using bilinguals*. Paper presented at the annual meeting of the American Educational Research Association (Division D), Chicago, IL.

- Sireci, S. G., Fitzgerald, C., & Xing, D. (1998, April). *Adapting credentialing examinations for international uses*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Sireci, S. G., & Han, K. T. (2007, April). *Methods for evaluating the validity of test scores for English language learners*. Paper presented at the annual meeting of the American Educational Research Association, as part of the symposium Design and Evaluation of Accessible Assessment Items for English Learners (R. Duran, Chair), Chicago.
- Sireci, S. G., Powers, S., Rios, J. A. (2013, May). *Contemporary Methods for Evaluating the Comparability of Translated Tests*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Sireci, S. G., & Rizavi, S. M. (1997, March). *Defining social studies content domains using multidimensional scaling*. Poster presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Sireci, S. G., & Schweid, J. A. (2011, April). *Beyond alignment: Important questions to ask (and answer) to evaluate content validity*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Sireci, S. G., Zanetti, M., & Berger, J. (2001, April). *Recent and anticipated changes in the postsecondary admissions process*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Wang, X., Sireci, S. G. (2013, April). *Investigating the Relationship Between Item Response Time and Cognitive Level*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Zenisky, A. L., & Sireci, S. G. (2005, April). *No adult left behind either: Creating large-scale computer-based tests for adult basic education students*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- American Psychological Association***
- Meara, K., & Sireci, S. G. (1999, August). *Appraising the dimensionality of the Medical College Admissions Test across diverse applicant groups*. Paper presented at the annual meeting of the American Psychological Association, Boston, MA.
- Robin, F., Sireci, S. G., & Hambleton, R. K. (1999, August). *Evaluating credentialing exams administered in multiple languages*. Poster presented at the annual meeting of the American Psychological Association, Boston, MA.
- Scarpatti, S., & Sireci, S. G. (1998, August). *Including students with disabilities in state and district tests: Perceptions of accommodations and score validity*. Paper presented at the annual meeting of the American Psychological Association (Division 5), San Francisco, CA.
- Shelley-Sireci, L. M., & Sireci, S. G. (1998, August). *Controlling for uncontrolled variables in cross-cultural research*. Paper presented at the annual meeting of the American Psychological Association (Division 5), San Francisco, CA.

- Sireci, S. G., (1992, August). *The utility of IRT in small-sample testing applications*. Poster presented at the centennial annual conference of the American Psychological Association, Washington, D.C.
- Sireci, S. G., (1995, August). *Using cluster analysis to solve the problem of standard setting*. Paper presented at the annual conference of the American Psychological Association (Division 5), New York, NY, August.
- Sireci, S. G., (1996, August). *Psychos and psychometrics: Careers in quantitative psychology*. Invited paper presented at the annual meeting of the American Psychological Association, Toronto, Canada.
- Sireci, S. G., (1996, August). *Using bilinguals to evaluate the comparability of a test administered in different languages*. Invited paper presented at the annual meeting of the American Psychological Association, (Division 5), Toronto, Canada.
- Sireci, S. G., Bastari, B., & Allalouf, A. (1998, August). *Evaluating construct equivalence across adapted tests*. Invited paper presented at the annual meeting of the American Psychological Association (Division 5), San Francisco, CA.
- Sireci, S. G., & Shelley-Sireci, L. M. (2011, August). *Identifying and resolving ethical issues in international assessment*. Invited paper presented at the annual meeting of the American Psychological Association, Washington, DC.
- Sireci, S. G. (2022, August). *Overcoming the “racism reflex” to improve college admissions testing*. Invited presentation delivered at the annual meeting of the American Psychological Association, Minneapolis, MN.

Association of Test Publishers

- Davis-Becker, S., & Sireci, S. G. (2019, November). *But my psychometrician told me it was fine! Understanding psychometric standards and practice*. Presentation delivered at the annual conference of the Institute for Credentialing Excellence, San Diego.
- Foster, D., & Sireci, S. G. (2023, March). *The value of unpredictability for enhancing security and fairness in high-stakes exams*. Presentation delivered at the annual conference of the Institute for Credentialing Excellence, Dallas.
- Hambleton, R., K., & Sireci, S. G. (2002, February). *Translating credentialing exams: Practical guidelines and steps to ensure validity*. Invited address delivered at the annual meeting of the Association of Test Publishers, Carlsbad, CA.
- Sireci, S. G. (2002, February). *Balancing efficiency and validity when setting standards on computer-based tests*. Presentation delivered at the annual meeting of the Association of Test Publishers, Carlsbad, CA.
- Sireci, S. G. (2003, February). *Test statistics*. Invited workshop presented at the annual meeting of the Association of Test Publishers, Amelia Island, FL.
- Sireci, S. G. (2003, February). *Test translations: Localization issues*. Invited presentation delivered at the annual meeting of the Association of Test Publishers, Amelia Island, FL.

- Sireci, S. G. (2006, February). *Testing across languages and cultures: Methods for ensuring and evaluating measurement equivalence*. Presentation delivered at the annual meeting of the Association of Test Publishers, Orlando.
- Sireci, S. G. (2008, March). *Preventive maintenance: Building validity through sound test construction and evaluation processes*. Presentation delivered at the 2008 Annual Conference of the Association of Test Publishers as Part of the Symposium “Maintaining Quality in Large-Scale Assessments,” (K. Huff, Chair), Dallas, TX
- Sireci, S. G. (2013, February). *Using the Standards for Educational and Psychological Testing as a validation framework*. Workshop presentation delivered at the annual meeting of the Association of Test Publishers, Hollywood, FL.
- Sireci, S. G. (2016, March). *Validity theory and test validation: An introduction*. Presentation delivered at the annual meeting of the Association of Test Publishers, Orlando.
- Sireci, S. G. (2018, February). *Building validity into the standard setting process*. Presentation delivered at the annual meeting of the Association of Test Publishers, San Antonio, TX.
- Sireci, S. G., & Botha, S. (2018, February). *Setting reasonable policies for exam time limits*. Presentation delivered at the annual meeting of the Association of Test Publishers, San Antonio, TX.
- Sireci, S. G. (2022, March). *Empowering the marginalized through algorithmic item design and UNDERSTANDARDization* Presentation delivered at the annual meeting of the Association of Test Publishers, Orlando.
- Weiner, J., & Sireci, S. G. (2019, March). *Technology-based assessment guidelines: Present and future*. Presentation delivered at the annual meeting of the Association of Test Publishers, Orlando.
- Weiner, J., & Sireci, S. G. (2019, September). *Guidelines for technology-based assessment guidelines: Exploring the European perspective*. Presentation delivered at the European annual meeting of the Association of Test Publishers, Madrid.
- Weiner, J., & Sireci, S. G., von Davier, A., & Tong, A. (2020, September). *Update on the ITC/ATP guidelines for technology-based assessment*. Presentation delivered at the European annual meeting of the Association of Test Publishers [online].
- Weiner, J., Sireci, S. G., Oliveri, M. E., Mills, C., & Schoenig, R. (2022, March). *New guidelines for technology-based assessments*. Presentation delivered at the annual meeting of the Association of Test Publishers, Orlando.
- Coalition on Adult Basic Education***
- Green, A., Vanek, J., Oliveri, M.E. & Sireci, S. G. (2023, April). *New measurable skills gain models are coming: Help us build them!* Symposium delivered at the annual conference of the Coalition on Adult Basic Education, Atlanta, GA.

- Sireci, S. G. (2025, March). Research and practical solutions: *Digital tools to support teaching and learning from the CREATE Network: ASAP assessments*. Presentation delivered at the annual meeting of the Coalition on Adult Basic Education (COABE) National Conference. Dallas.
- Sireci, S. G., Suárez-Álvarez, J., Oliveri, M.E., & Sabatini, J. (2022, April). *Assessing adult skills through technology: The MAPT and ASAP* [Symposium]. Coalition on Adult Basic Education (COABE) National Conference. Seattle.
- Suárez-Álvarez, J., and Sireci, S. (2022, April). *Adult skills assessment program: Actionable assessments for adult learners*. Coalition on Adult Basic Education (COABE) National Conference. Seattle, WA.
- Suárez-Álvarez, J., & Sireci, S. G. (2023, April). *Adult Skills Assessment Program (ASAP): Progress Update Year 2*. Coalition on Adult Basic Education (COABE) National Conference. Atlanta, GA, United States.
- Suárez-Álvarez, J., Sireci, S. G., Schuler-Jones, H., & Varn, D. (2023, April). *Designing digital tasks for adult numeracy learners to solve real-world problems* [Symposium]. Coalition on Adult Basic Education (COABE) National Conference. Atlanta, GA, United States.

National Council on Measurement in Education

- Araneda, S., Lee, D., Lewis, J., Sireci, S. G., Arslan, B., Keehner, M., Moon, J. A. (2021, September). *Exploring relationships among examinees' behaviors and performance using response process data*. Paper presented at the annual meeting of the National Council on Measurement in Education. [Virtual]
- Araneda, S., & Sireci, S. G. (2023, April). Sketching a validity argument from user experiences shared on twitter. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.
- Baron, P., Slater, S. C., & Sireci, S. G. (2019, April). *Evaluating panelists' understanding of standard-setting data*. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto.
- Berberoglu, G., Sireci, S. G., & Hambleton, R. K. (1997, March). *Comparing translated items using bilingual and monolingual examinees*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Casas, M., & Sireci, S. G. (2023, April). Measurement invariance across immigrant and non-immigrant populations on PISA non-cognitive scales. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Copella, J., & Sireci, S. G. (2009, April). *Interpreting non-uniform DIF*. Poster presented at the annual conference of the National Council on Measurement in Education. San Diego.
- Crespo Cruz, E., & Sireci, S. G. (2025, April). *Evaluating alignment criteria for educational assessments*. Poster presented at the annual conference of the National Council on Measurement in Education. Denver.

- Crotts, K., Sireci, S. G., & Zenisky, A. L. (April, 2011). *Evaluating content validity in a multistage adaptive test*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Crotts, K., Zenisky, & Sireci, S. G. (2012, April). *Estimating measurement precision in reduced-length multistage-adaptive testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver.
- Crotts Roohr, K., & Sireci, S. G. (2014, April). *Evaluating computer-based test accommodations for English learners*. Paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia.
- Egan, K. L., Sireci, S. G., & Swaminathan, H. (1998, April). *Effect of item bundling on the assessment of test dimensionality*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Foster, C., Wells, C., Sireci, S. G., Randall, J. (2013, April). *Taking the Next Step in Erasure Analysis: An Evaluation of the Development and Accuracy of Modern Methods*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Gándara, F., & Sireci, S. G. (2021, June). The economic consequences of ignoring testing consequences. Paper presented at the annual meeting of the National Council on Measurement in Education [virtual].
- Hambleton, R. K., Sireci, S. G., & Li, S. (2003, April). *Identifying common problems in item translations: A meta-analysis*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Hambleton, R. K., Swaminathan, H., Sireci, S. G., Xing, D., & Rizavi, S. (1998, April). *Estimating item statistics with judgmental data and Bayesian statistical procedures*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Han, K., Sireci, S. G., Wells, C., & Zenisky-Laguilles, A. (2006, April). *Methods for evaluating gain at the program level*. Presentation delivered at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Kaira, L. T., & Sireci, S. G. (2011, April). *Using item mapping to evaluate alignment between curriculum and assessment*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Khaliq, S. N., & Sireci, S. G. (2004, April). *Evaluating essay scoring programs: Beyond percent agreement and Pearson correlations*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Lee, D., Buzick, H. M., Sireci, S. G., Lee, M., Laitusis, C. (2020, September). Understanding and illustrating the use of test accessibility supports. Paper presented at the annual meeting of the National Council on Measurement in Education (virtual).

- Lee, M., Wells, C., & Sireci, S. G. (2011, April). *Assessing measurement invariance in the context of disparate sample sizes and proficiency distributions*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Lewis, J. L., Lim, H., Padellaro, F., & Sireci, S. G. (2019, April). *Setting multiple standards on a vertically scaled multistage-adaptive test*. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto.
- Li, S., Wang, S., Sireci, S. G., & Keller, L. (2004, April). *Accounting for testlet structure in vertical scaling*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Li, X., & Sireci, S. G. (2012, April). *Analyzing alignment data using multidimensional scaling*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver.
- Lukhele, R. & Sireci, S. G., (1995, April). *Using IRT to combine multiple-choice and free-response sections of a test onto a common scale using a priori weights*. Paper presented at the annual conference of the National Council on Measurement in Education, San Francisco, CA.
- Marland, J., Wells, C. S. & Sireci, S. G. (2015, April). *Investigating the amount of systematic and random error in classroom-level SGPs*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Marland, J., Sireci, S. G., Zenisky, A. L., & Pham, D. (2015, April). *What can item response times tell us about differential item functioning for English learners?* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Martineau, J., Sireci, S. G., McCall, M., & Gallagher, C. (2013, April). *The current state of the Smarter Balanced Assessment Consortium research agenda*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Maxfield, M., Lewis, J., Crespo Cruz, E., & Sireci, S. G. (2025, April). *An English proficiency test with Spanish directions: Helpful or irrelevant?* E-board presented at the annual conference of the National Council on Measurement in Education. Denver.
- O'Donnell, F., Sireci, S. G., & Zenisky, A. L. (2019, April). *Understanding teachers' interpretation of achievement labels from score reports*. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto.
- O'Donnell, F., & Sireci, S. G. (2000, September). *Stakeholder perceptions of achievement labels from tests*. Paper presented at the annual meeting of the National Council on Measurement in Education (virtual).
- O'Neil, T., & Sireci, S. G. (2002, April). *Evaluating the content validity of a state-mandated science assessment across two successive administrations*. Paper presented at the annual conference of the National Council on Measurement in Education, New Orleans, LA.

- Padilla, J., Benitez, I., Hidalgo, M. D., & Sireci, S. G. (2012, April). *Can cognitive interviewing help in interpreting DIF?* Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver.
- Rick, F., Sireci, S. G., Dickey, J., Slater, S. C., & Kannan, P. (2017, April). *Parents' perspectives on summative test score reports.* Paper presented at the annual conference of the National Council on Measurement in Education, San Antonio, TX.
- Sireci, S. G., (1995, April). *The central role of content representation in test validity.* Paper presented at the annual conference of the National Council on Measurement in Education, San Francisco, CA.
- Sireci, S. G., (1996, April). *Technical issues in linking assessments across languages.* Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.
- Sireci, S. G. (1998, April). *"I can't get the chalk off my butt!" and other cries from the ivory tower.* Invited address delivered at the annual meeting of the National Council on Measurement in Education as part of the symposium "Career directions in educational measurement" (Cyndy Schmeiser, Chair), San Diego, CA.
- Sireci, S. G. (1999, April). *Training the next generation of measurement professionals.* Invited paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Quebec, Canada.
- Sireci, S. G. (2004, April). *The role of sensitivity review and differential item functioning analyses in reducing the achievement gap.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Sireci, S. G. (2005, April). *Measurement problems revisited.* Presentation delivered at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.
- Sireci, S. G. (2005, April). *No modification necessary: Some reflections on Dr. Angoff.* Presentation delivered at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.
- Sireci, S. G. (2005, April). *The most frequently UNasked questions about standardized testing.* Presentation delivered at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.
- Sireci, S. G., (2012, April). *De-"Constructing" Test Validation.* Paper presented at the NCME symposium "Beyond Consensus: The Changing Face of Validity," (P. Newton, Chair), Vancouver.
- Sireci, S. G. (2016, April). *Are we ready for college and career readiness?* Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, DC.

- Sireci, S. G. (2016, April). *Using technology to enhance educational assessments: Comments on four chapters*. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, DC.
- Sireci, S. G. (2017, April). *International standards for educational and psychological testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Antonio, TX.
- Sireci, S. G. (2020, June). Making measurement matter: A practical guide for assessment developers and test users. Workshop presented (virtually) at the annual meeting of the National Council on Measurement in Education. Available at: https://www.youtube.com/watch?v=1rgryQ1V_OE.
- Sireci, S. G. (2022, February). *Overcoming the racism reflex*. Presentation delivered as part of the NCME: Diversity, Equity, and Inclusion in Assessment Webinar Series [online]. Available for viewing at <https://www.youtube.com/watch?v=oyaVWY6lWcM>.
- Sireci, S. G., Baldwin, P., Martone, D., & Han, K. T. (2007, April). Determining cut points on a multi-stage test for federally established proficiency levels. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.
- Sireci, S. G., & Benitez, I. (2023). Evidence for test validation: A guide for practitioners. *Psicothema*, 35(3) 217-226.
- Sireci, S. G., Lewis, C., & Martone, A. (2006, April). *Why can't we all just get along? How psychometricians can work with school districts to improve student learning*. Presentation delivered at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Sireci, S. G., Foster, D., Olsen, J. B., & Robin, F. (1997, March). *Comparing dual-language versions of international computerized certification exams*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Sireci, S. G., & Geisinger, K. F., (1993, April). *Using subject matter experts to assess content representation: a MDS analysis*. Paper presented at the annual conference of the National Council on Measurement in Education, Atlanta, GA.
- Sireci, S. G., & Gonzalez, E. J. (2003, April). *Evaluating the structural equivalence of tests used in international comparisons of educational achievement*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Sireci, S. G., Harter, J., Yang, Y., & Bhola, D. (2000, April). *Evaluating the construct equivalence of an international employee survey*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Sireci, S. G., & Khaliq, S. N. (2002, April). *An analysis of the psychometric properties of dual language test forms*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

- Sireci, S. G., & Lim, H., Rodriguez, G., Banda, E., & Zenisky, A. (2018, April). *Evaluating criteria for validity evidence based on test content*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.
- Sireci, S. G. & O’Riordan, M. (2020, September). Presentation delivered at the annual meeting of the National Council on Measurement in Education, virtual. Available at https://www.youtube.com/watch?v=b0HCMouOc0Y&list=PLPxHgAQNdRRqj2S2cri9AmiP1xqRM5_0E&index=15.
- Sireci, S. G., & Park, Y. (2019, April). *Evaluating Initiatives to Recruit a Larger and More Diverse Educational Measurement Workforce*. Presentation delivered at the annual meeting of the National Council on Measurement in Education, Toronto.
- Sireci, S. G., & Parker, P. (2006, April). *Enforcing the Standards: Exploring the use of the Standards by the courts*. Presentation delivered at the annual meeting of the National Council on Measurement in Education, San Francisco, CA
- Sireci, S. G., Patelis, T., Rizavi, S., Dillingham, A., & Rodriguez, G. (2000, April). *Setting standards on a computerized-adaptive placement examination*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Sireci, S. G., & Randall, J. (2022, April). *Evolving notions of fairness in testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego.
- Sireci, S. G., Rios, J., & Powers, S. (2017, April). *Comparing scores from tests administered in different languages*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Antonio, TX.
- Sireci, S. G., Robin, F., & Patelis, T. (1997, March). *Empirically-based standard setting using cluster analysis*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Sireci, S. G., Wells, C., Bahry, L. (2013, April). *Student Growth Percentiles: More Noise Than Signal?* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Sireci, S. G., Wells, C., Hu, H. (2014, April). *Using internal structure validity evidence to evaluate test accommodations*. Paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia.
- Sireci, S. G., Yang, Y., Harter, J., & Ehrlich, E. J. (2004, April). *Evaluating guidelines for test adaptations: An empirical analysis of translation quality*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Sireci, S. G., Xing, D., & Fitzgerald, C. (1999, April). *Evaluating translation DIF across multiple groups: Lessons learned from the information technology industry*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

- Suárez-Álvarez, J., Oliveri, M.E. & Sireci, S.G. (2023, April). *DIRTY CATs and Other DIRTY Assessments: The Adult Skills Assessment Program*. Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago, IL.
- Tran, P., Ketan, & Sireci, S. G. (2025, April). *Evaluating the content quality of items written with and without artificial intelligence*. Poster presented at the annual meeting of the National Council on Measurement in Education, Denver.
- Wells, C. S., Sireci, S. G., Bahry, L. (2014, April). *The effect of conditioning years on the reliability of SGPs*. Paper presented at the 2014 annual meeting of the National Council of Measurement in Education, Philadelphia, PA.
- Zenisky, A., Sireci, S. G., & Casas, M. (2019, April). *Picture this: using visualizations in standard setting for educator certification*. Paper presented at the annual meeting of the National Council on Measurement in Education Toronto, ON.
- Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2000, April). *Effects of local item dependence on the validity of IRT item, test, and ability statistics*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Zenisky, A. L., & Sireci, S. G. (2009, April). *Performing At or Above Proficient: The Reporting of NAEP Results in the Internet Age*. Paper presented at the annual conference of the National Council on Measurement in Education. San Diego.
- Zenisky, A., & Sireci, S. G. (2013, April). *Innovative Items to Measure High-Order Thinking: Development and Validity Considerations*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Zenisky, A., & Sireci, S. G. (2020, September). Purposeful design for useful tests: Considerations of choice in multistage-adaptive testing. Paper presented at the annual meeting of the National Council on Measurement in Education (virtual). Available at https://www.youtube.com/watch?v=3ieCJWwrk1c&list=PLPxHgAQNdRRqj2S2cri9AmiP1xqRM5_0E&index=25).
- Zenisky, A., Sireci, S. G., & Casas, M. (2019, April). *Picture This: Using Visualizations in Standard-Setting for Educator Certification*. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto.
- Zumbo, B. D., Sireci, S. G., & Hambleton, R. K. (2003, April). *Revisiting exploratory methods for construct comparability: Is there something to be gained from the ways of old?* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Northeastern Educational Research Association***
- Allalouf, A., Bastari, B., Hambleton, R. K., & Sireci, S. G. (1997, October). *Comparing the dimensionality of a test administered in two languages*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.

- Araneda, S. Lee, D., Lewis, J., O’Riordan, M., Sireci, S. G., & Zenisky, A. L. (2021, October). *Incorporating SmartItem technology on a multistage-adaptive test*. Presentation delivered at the annual meeting of the Northeastern Educational Research Association [virtual].
- Banda, E., & Sireci, S. G. (2016, October). Current practices In test accommodations in the USA. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Banda, E., & Sireci, S. G. (2018, October). Adaptation of the ATMI scale into Chichewa: Language validity, and preliminary psychometric investigation. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Chulu, B. W., & Sireci, S. G. (2002, October). *Evaluating the content validity of the MSCE Physical Science Exam*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.
- Chulu, B. W., Sireci, S. G., Wells, C. S., & Abedi, J. (2005, October). *Revisiting simplified English as a test accommodation for English language learners*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.
- Crespo Cruz, E., Duan, L., Pham, D., & Sireci, S. G. (2024, October). *Exploring the quality of AI-generated numeracy items: The human-in-the-loop approach*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Crotts, K., Sireci, S. G., & Wells, C. S. (2011, October). *Examining the structural invariance of English and Spanish video accommodations for English learners using multidimensional scaling*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Fan, F., Sireci, S. G., & Wells, C. S. (2014, October). *Evaluating internal structure validity evidence for the 2011 TIMSS mathematics test*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Feng, W., Tran, P., Sireci, S. G., & Lan, A. (2024, October). *Using artificial intelligence to scale multiple-choice math items*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Faulkner-Bond, M., & Sireci, S. (October 2012). *Investigating large score declines on a low-stakes, multistage-adaptive proficiency test*. Paper presented at the Northeastern Educational Research Association Annual Meeting, Rocky Hill, CT.
- Faulkner-Bond, M. , Wolf, M., Wells, C. S., & Sireci, S. (2014, October). *Exploring the Factor Structure of a K-12 English Language Proficiency Assessment*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Foster, C., & Sireci, S. G. (2011, October). *Relative judgmental scaling process for estimating item difficulty*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.

- Gándara, F., & Sireci, S. G. (2014, October). *Evaluating the misalignment between information and content specifications by means of statistically equivalent tests*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Green, P., & Sireci, S. G. (1998, October). *Legal issues in teacher certification testing*. Paper presented at the meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Gubin, A., Pearlman, L. A., & Sireci, S. G. (1999, October). *An evaluation of the dimensionality of the Traumatic Stress Institute Scale*. Paper presented at the meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Han, N., & Sireci, S. G. (2003, October). *Evaluating the equivalence of multiple language versions of TIMSS using a generalized Mantel-Haenszel procedure*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.
- Hauger, J. B., & Sireci, S. G. (2003, October). *Detecting differential item functioning across examinees tested in their dominant language and examinees tested in a second language*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.
- Huff, K. L., & Sireci, S. G. (2000, October). *Validity issues in computer-based testing*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Huff, K. L., & Sireci, S. G. (2001, October). *Appraising the dimensionality of a large-scale science assessment across demographic groups*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.
- Keller, L., Rodriguez, G., Zenisky, A., & Sireci, S. G. (1999, October). *Assessing the dimensionality of the grade 4 MCAS science test: A multi-method analysis*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Ketan, Mena, F., Crespo Cruz, E., Duan, L., Pham, D., & Sireci, S. G. (2024, October). *Large language models for item generation: Comparing ChatGPT-4, Gemini, Latimer, and Copilot*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Khaliq, S. N. & Sireci, S. G. (2001, October). *Methods for evaluating construct equivalence*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.
- Lee, M., Wells, C., & Sireci, S. G. (2010, October). *A comparison of linear and nonlinear factor analysis in examining the effect of a calculator accommodation on math performance*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Li, S., & Sireci, S. G. (2003, October). *Applying logistic regression DIF detection procedures to the 1999 TIMSS science multiple-choice items*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.

- Li, X., & Sireci, S. G. (2011, October). *Analyzing content validity ratings using multidimensional scaling*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Marland, J., & Sireci, S. G. (2014, October). *Age matters: Grouping approaches for detecting differential item functioning due to age*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Marland, J. J., Wells, C. S., Sireci, S. G., & Castellano, K. (2015, October). *Investigating the Amount of Systematic and Random Error in Aggregate SGPs*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Mena, F., Ketan, Crespo Cruz, E., Pham, D. & Sireci, S. G. (2024, October). *Generating Numeracy items with ChatGPT-4 for the Adult Skills Assessment Program: Opportunities and limitations*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- O'Neil, T., & Sireci, S. G. (2001, October). *The consistency of dimensionality across administrations of a large-scale science assessment*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.
- Padilla, J. L., Benitez, I., Hildalgo, M. D., & Sireci, S. G. (2011, October). *Cognitive interviewing evidence of DIF in polytomous items on the PISA 2006 student questionnaire*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Pitoniak, M. J., & Sireci, S. G. (2000, October). *A multitrait-multimethod validity investigation of scores from a professional licensure examination*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Pitoniak, M. J., & Sireci, S. G. (2001, October). *The relationship between class size and students' course evaluations: An analysis of the SRTI*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.
- Robin, F., Patelis, T., & Sireci, S. G. (1996, October). *Empirical methods for setting standards on tests*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Sireci, S. G. (2008, October). *Fairness issues in cross-lingual assessment*. Presentation delivered at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Sireci, S. G., (1991, October). *"Sample-independent item parameters?" An investigation of the stability of IRT item parameters estimated from small sample sizes*. Paper presented at the annual conference of the Northeastern Educational Research Association, Ellenville, NY.
- Sireci, S. G., (1993, October). *An assessment of ethics and the ethics of assessment: reactions to the NCME code of ethical assessment practices in education*. Invited paper presented at the annual conference of the Northeastern Educational Research Association, Ellenville, NY.

- Sireci, S. G., (1995, October). *Problems and issues in linking assessments across languages*. Paper presented at the annual conference of the Northeastern Educational Research Association, Ellenville, NY.
- Sireci, S. G. (1997, October). *Dimensionality assessment: Implications for psychometric theory and practice*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Sireci, S. G. (2013, October). *Suggestions for navigating the publication process*. Invited presentation delivered at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Sireci, S. G. (2016, October). A guide to writing effective and successful grants. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Sireci, S. G., (2021, July). Publishing educational research: The peer review process. Presentation delivered as part of the Northeastern Educational Research Association webinar Demystifying the peer review process: A how-to conversations for proposals and manuscripts. [Virtual]
- Sireci, S. G., & Crespo Cruz, E. (2023, October). *The new potential of K12 educational assessments*. Presentation delivered at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Sireci, S. G., & Crotts, K. (2010, October). *The importance of content validation in educational testing (then and now)*. Presentation delivered at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Sireci, S. G., & Faulkner-Bond, M. (2011, October). *If I were king of the forest: Designing an effective and valid statewide testing program*. Presentation delivered at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Sireci, S. G., Geisinger, K.F., & Lee, S. (1990, November). *Applying empirical analyses to the evaluation of test content*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Sireci, S. G. & Rodriguez, G. (2014, October). *Research-based guidelines on test accommodations for English learners*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Sireci, S. G., & Swaminathan, H. (1996, October). *Evaluating translation equivalence: So what's the big DIF?* Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Sireci, S. G., & Valdiviezo, L. (2019, October). Beyond qualitative and quantitative: Working together to research social justice issues. Presentation delivered at the annual meeting of the Northeastern Educational Research Association, Trumbull, CT.
- Sireci, S. G., Wiley, A., & Keller, L. A. (1998, October). *An empirical evaluation of multiple-choice item writing guidelines*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.

- Sireci, S. G., & Zenisky, A. L., & Randall, J. (2008, October). *New methods for building validity into the standard setting process*. Paper presented at the annual conference of the Northeastern Educational Research Association, Rocky Hill, CT.
- Sireci, V. P., & Sireci, S. G. (2007, October). *Has the MCAS helped students learn?* Poster presented at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Soto, A., Sireci, S. G., Keller, L. A., & O'Malley, K. (2011, October). *Evaluating teachers using value-added models: Current practices and validity evidence*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Suárez-Álvarez, J., Zenisky, A., Oliveri, M.E., and Sireci, S. (2022, October). *The adult skills assessment program: A competency framework for 21st-century learners*. Northeastern Educational Research Association (NERA). Trumbull, CT.
- Washington, E. D., De León, B., Smith, T. J., & Sireci, S. G. (1997, October). *Leveling the playing field: Improving teachers of minority students' attitudes towards and knowledge of standardized tests*. Paper presented at the annual conference of the Northeastern Educational Research Association, Ellenville, NY.
- Wiley, A., & Sireci, S. G. (1994, October). *Determining the reliability of a test with free-response and multiple-choice items*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Ying, L., & Sireci, S. G. (2003, October). *Validity issues in test speededness*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Kerhonkston, NY.
- Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (1999, October). *Effects of item dependencies on the validity of IRT item, test, and ability statistics*. Paper presented at the meeting of the Northeastern Educational Research Association, Ellenville, NY.
- Zhao, Y., & Sireci, S. G. (2005, October). *Validity issues in automated essay scoring*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.

Presentations: Other

- Banda, E., Park, Y., Rick, F., & Sireci, S. G. (2016, April) *A Comparison of Validity Arguments Across Three High School Equivalency Tests University of Massachusetts Amherst*. Paper presented at the annual meeting of the New England Educational Research Organization, Portsmouth, NH.
- Benitez, I. B., Padila, J. L., Hildalgo, M. D., & Sireci, S. G. (2010, October). *Detecting Sources of Differential Item Functioning in Polytomous Items by Cognitive Interviewing*. Detecting Sources of Differential Item Functioning in Polytomous Items by Cognitive Interviewing
- Berberoglu, G., Sireci, S. G., & Hambleton, R. K. (1997, July). *A comparison of the graded response model and the Mantel-Haenszel method for detecting DIF across different language groups*. Paper presented at Fifth European Congress of Psychology, Dublin, Ireland.

- Chakwera, E., Khembo, D., & Sireci, S. G. (2002, April). *High-stakes testing in the warm heart of Africa: Challenges and success of the Malawi National Examinations Board*. Paper presented at the annual meeting of the New England Educational Research Organization, Northampton, MA.
- Crespo Cruz, E., Horton, O., Immanuel, A., & Sireci, S. G. (2025, April). *Is the Adult Skills Assessment Program truly culturally responsive? A proactive self-audit*. Presentation delivered at the 8th International Conference of the Center for Culturally Responsive Evaluation and Assessment, April 10, 2025, Chicago.
- Colvin, K.F., Sireci, S. G., & Keller, L. A. (2011, April). *Investigation of item parameter drift in a computerized multistage adaptive test*. Paper presented at the annual meeting of the New England Educational Research Organization, New Bedford, MA.
- Diao, H., & Sireci, S. G. (2016, April). *Using differential item functioning to evaluate performance differences on technology enhanced items*. Paper presented at the annual meeting of the New England Educational Research Organization, Portsmouth, NH.
- Gökçe, S., Berberoglu, G., Sireci, S. G., & Wells, C. S. (2018, July). Can language and cultural differences play an important role in TIMSS mathematics assessments? Paper presented at the biannual meeting of the International Test Commission, Montreal.
- Hambleton, R. K., & Sireci, S. G. (1999, September). *Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving credentialing testing practices*. Invited address delivered at the annual meeting of the Council on Licensure, Enforcement, and Regulation, Portland, OR.
- Hambleton, R. K., & Sireci, S. G. (2009, June). *Setting performance standards: Methods, validity issues, & research for improving practice*. Seminar for ETS Interns, Princeton, NJ.
- Horton, O., & Sireci, S. G. (2025, July). *Building and evaluating culturally responsive assessments*. Presentation delivered at XI European Congress of Methodology as part of the symposium “Innovative Approaches to Address Validity” (I. Benitez. Chair). Tenerife, Spain
- Huff, K. L., Koenig, J. A., Treptau, M. S., & Sireci, S. G. (1999, November). *Validity of the MCAT for predicting clerkship performance of medical students grouped by sex and ethnicity*. Paper presented at the Research in Medical Education conference, Washington, DC.
- Khaliq, S. N., & Sireci, S. G. (2001, May). *Methods for evaluating construct equivalence*. Paper presented at the annual conference of the Canadian Society for the Study of Education, Quebec.
- Martone, A., & Sireci, S.G. (2008, September). *The Massachusetts Adult Proficiency Tests: Plans for Fiscal 2009 and Beyond*. Presentation delivered at the Massachusetts Coalition for Adult Education Network Conference, Marlborough, MA.
- Pham, D., Sireci, S. G., Ketan, Crespo Cruz, E., Mena Serrano, F., Duan, L. (2024, July). Building Culturally Responsive Assessments to Support Adult Learners: From Co-design Studies to Assessment Development and Profile Reporting. Presentation delivered at the biennial meeting of the International Test Commission, Granada, Spain.

- Rick, F, Park, Y, Banda, E, and Sireci, S. G., (2016, April) *Validation Frameworks for Evaluating 21st-Century Assessments*. Paper presented at the annual meeting of the New England Educational Research Organization, Portsmouth, NH.
- Rios, J. A., & Sireci, S. G., (2012, July). *Guidelines versus Practices in Cross-Lingual Assessment: A Disconcerting Disconnect*. Presentation delivered at the 8th annual conference of the International Test Commission, Amsterdam.
- Robin, F., Sireci, S. G., & Hambleton, R. K. (1999, May). *Evaluating credentialing exams administered in multiple languages*. Poster presented at the International Conference on Adapting Tests for Use in Multiple Languages and Cultures, Washington, DC.
- Sánchez-Martín, M., Niessen, S., Luesia, J. F., & Sireci, S. G. (2025, July). *Evaluating the concurrent validity of traditional and alternative measures used in university admission processes in Spain*. Presentation delivered at XI European Congress of Methodology. Tenerife, Spain
- Sireci, S. G. (1996, November). *Evaluating the predictive validity of the MCAT across diverse applicant groups*. Invited paper presented at the annual meeting of the Association of American Medical Colleges, San Francisco, CA.
- Sireci, S. G. (1999, May). *Statistical methods for determining problematic items in test adaptations*. Workshop presented at the International Conference on Adapting Tests for Use in Multiple Languages and Cultures, Washington, DC.
- Sireci, S. G. (2003, December). *Test accommodations for English language learners: A review of the literature*. Invited presentation delivered at the U.S. Department of Education's Office of English Language Acquisition's "Celebrate Our Rising Stars Summit," Washington, DC.
- Sireci, S. G. (2008, October). *Packing and Unpacking Sources of Validity Evidence: History Repeats Itself Again*. Presentation delivered at the 9th Annual Maryland Conference: The Concept of Validity, College Park, MD.
- Sireci, S. G. (2009, April). *Educational Evaluation in the United States*. Presentacion a la Inspectores de Educacion de Asturias (Spain).
- Sireci, S. G. (2009, June). *Fostering dialogue among psychos and policy wonks*. Presentation delivered at the National Conference on Student Assessment. Los Angeles, CA.
- Sireci, S. G. (2011, February). *Conquering (two) problems in 21st-century educational and psychological testing*. Presentation for the feast of Juan Huarte de San Juan, Universidad de Oviedo (Spain).
- Sireci, S. G. (2011, July). *An historical perspective on validity theory and test validation*. Paper presented at the 12th European Congress of Psychology, Istanbul, Turkey.
- Sireci, S. G., (2012, July). *Conducting Research Worth Publishing: Illustrations from the International Journal of Testing*. Presentation delivered at the 8th annual conference of the international Test Commission, Amsterdam.

- Sireci, S. G., (2012, July). *Standards for Educational and Psychological Testing: A Validation Framework*. Paper delivered at the V European Congress of Methodology, as part of the symposium “Standards and Practices in Validating Tests and Questionnaires: An International Perspective,” Santiago de Compostela, Spain.
- Sireci, S. G. (2012, November). *Standards for Educational and Psychological Testing: A Validation Framework*. Visiting scholar presentation to ACT Staff, Iowa City.
- Sireci, S. G. (2013, June). *Smarter balanced validation: Incorporating systemic objectives into a validity argument*. Paper presented at the National Conference on Student Assessment, National Harbor, MD.
- Sireci, S. G. (2013, November). *A theory of action for test validation*. Invited presentation delivered at the 13th annual Maryland Assessment Conference, College Park.
- Sireci, S. G. (2014, February). *Computer-based testing: Promises, challenges, and future directions*. Invited presentation delivered at the University of Barcelona.
- Sireci, S. G. (2014, August). *Evaluating the comparability of test variations*. Invited presentation to ACT Staff, Iowa City.
- Sireci, S. G. (2015, June). *Are we ready for college and career readiness?* Presentation at the National Conference on Student Assessment as part of the symposium, Validating Claims of College and Career Readiness (T. Patelis, Chair), June 23, 2015, San Diego, CA.
- Sireci, S. G. (2016, May). *Gathering validity evidence using multidimensional scaling*. Workshop presented at the 2016 ATINER conference, Athens, Greece.
- Sireci, S. G. (2017, October). *Issues and practices in test speededness: An historical perspective and empirical analyses*. Presentation delivered at the Timing Impact on Measurement in Education conference, Philadelphia, PA.
- Sireci, S. G. (2017, December). *Comments on Comunicación y uso válido de resultados*. Remarks delivered at the Seminario en Transparencia y Validez en la Selección en Educación Superior, Universidad de Chile, Santiago.
- Sireci, S. G. (2017, December). *How should we validate university admissions tests?* Presentation a la Seminario DEMRE Transparencia y Validez en la Selección en Educación Superior Universidad de Chile, Santiago.
- Sireci, S. G. (2018, June). *Statistical and qualitative approaches for facilitating comparability in cross-lingual assessments*. Presentation delivered at the OECD Seminar on Translating and Adapting Instruments in Large-Scale Assessments, Paris.
- Sireci, S. G. (2021, June). *Deriving decisions from disrupted data: Opportunities from the COVID-19 pandemic*. Presentation delivered at the National Conference on Student Assessment. [Virtual].
- Sireci, S. G. (2021, May). *Process data literacy*. Presentation delivered at the OECD Webinar 21st-century readers: Using process data to understand literacy skills in a digital world [online].

- Sireci, S. G. (2022, November). *How can assessments be integrated with teaching and learning?* Presentation delivered at the virtual conference “Extending His Reach: Promoting the Educational Legacy of Edmund W. Gordon.”
- Sireci, S. G. (2023, May). Test scores must always be comparable: Myth or reality? Presentation delivered to the State and Local Assessment Leaders Special Interest Group of the National Council on Measurement in Education [virtual].
- Sireci, S. G. (2023, May). What is standardization? Presentation delivered at the convening on Performance Assessment by the National Education Association and Fairtest. Washington, DC.
- Sireci, S. G. (2023, June). *The emperor has no clothes: The state of college admissions testing in the USA.* Presentation delivered to the Technical Advisory Committee of the Swedish Scholastic Assessment Test [virtual].
- Sireci, S. G. (2023, June). *What does fairness mean in educational testing?* Presentation delivered at National Conference on Student Assessment, New Orleans.
- Sireci, S. G., (2023, October) *Evaluating Perceptions of Test Content Using Multidimensional Scaling.* Presentation delivered to the Turkish Association of Measurement and Evaluation in Education and Psychology (EPODDER).
- Sireci, S. G. (2023, October). Validity Evidence Based on Test Content: Definitions and Methods. Presentation delivered to Stellenbosch University, Department of Industrial-Organizational Psychology (online).
- Sireci, S. G. (2023, November). *Understanding the ITC/ATP Guidelines for Technology-Based Assessments: Implications for Language Testing.* Presentation delivered at the Association of Language Testing-Europe’s 59th Meeting and Conference, Rome.
- Sireci, S.G., (2024, March). *Guidelines & Strategies for Valid Cross-Lingual Assessment.* Presentation delivered at the webinar, “ITC Guidelines for Cross-Cultural Testing, sponsored by the International Test Commission & Squadra Internazionale di Scienze Moderne NPO.
- Sireci, S. G., (2024, June). *Be Intelligent Rather than Artificial: The Proper Role of AI in Test Development.* Presentation delivered at the National Conference on Student Assessment as part of the symposium “Generative AI...Advice for State Assessment” Seattle, WA.
- Sireci, S. G. (2024, June). *Possible Futures for the Educational Testing Industry.* Presentation delivered to the Educational Assessment Group, University College London [online].
- Sireci, S. G. (2025, July). *Difficult Conversations in Revising the Standards for Educational and Psychological Testing.* Presentation delivered at XI European Congress of Methodology. Tenerife, Spain
- Sireci, S. G., Longe, B., Suárez, J., & Oliveri, M. E. (2022, July). *Assessing the dimensionality of O*NET cognitive ability ratings across job zones.* Poster presented at the international meeting of the Psychometric Society, Bologna, Italy.

- Sireci, S. G., & Faulkner-Bond, M. (2012, July). *Validity and fairness issues in assessing linguistic minorities*. Paper presented at the biannual meeting of the International Test Commission, Amsterdam.
- Sireci, S. G., & Faulkner-Bond, M. (2016, July). *Guidelines on developing fair assessments for linguistic minorities*. Paper presented at the biannual meeting of the International Test Commission, Vancouver.
- Sireci, S. G., & Green, P. (1999, May). *Legal and psychometric issues in teacher testing*. Invited seminar presented at the Assessment Literacy Conference, Shutesbury, MA.
- Sireci, S. G., & Nguyen, T. A. T. (2023, November). *Professional Testing Guidelines as Tools for Improving Educational Assessment: The Role of the International Test Commission*. Paper presented at the Annual Meeting of the Association for Educational Assessment – Europe, Malta.
- Sireci, S. G., Oliveri, M. E., & Badham, L. (2024, July). *Developing, Facilitating, & Evaluating Validity and Comparability in Multi-Language Assessment: A Literature Review*. Paper presented at the biennial meeting of the International Test Commission, Granada, Spain.
- Sireci, S. G. & Pitoniak, M. J. (2006, March). *Assessment accommodations: What have we learned from the research?* Invited presentation for the national conference “Accommodating Students With Disabilities on State Assessments: What Works. Savannah, GA.
- Sireci, S. G., & Pham, D. (2024, April). *Constraining for Fairness*. Presentation delivered at the annual meeting of the National Council on Measurement in Education as part of the symposium “Addressing Fairness Concerns in Digital-First, AI-Powered Testing Environments,” Philadelphia.
- Sireci, S. G., & Robin, F. (1996, June). *Setting passing scores on tests using cluster analysis*. Paper presented at the annual conference of the Classification Society of North America, Amherst, MA.
- Sireci, S. G., & Scarpati, S. (2003, October). *Effects of test accommodations on test performance*. Invited presentation delivered at the Education Policy Reform Research Institute conference “The effect of accommodations on accountability.” Washington, DC.
- Sireci, S. G., & Suárez, J. (2022, April). *Assessing adult skills through technology: The MAPT and ASAP*. Presentation delivered at the annual meeting of the Coalition for Adult Basic Education, Seattle.
- Sireci, S. G., Tong, Y., Ho, A., Norton, S., & Schwinn, P. (2019, June). *How can NCME better serve state assessment professionals?* Presentation delivered at the National Conference on Student Assessment, Orlando.
- Sireci, S. G., & Wells, C. S. (2008, June). *Evaluating the comparability of video accommodations for English language learners*. Paper presented at the National Conference on Student Assessment, Orlando, FL.

- Sireci, S. G., & Wells, C. S. (2009, June). *Methods for evaluating the comparability of video accommodations for English language learners*. Presentation delivered at the National Conference on Student Assessment. Los Angeles, CA.
- Sireci, S. G., Wells, C. S., Han, K. T., & Baldwin, P. (2007, April). *Evaluating item parameter drift in computerized-adaptive testing*. Paper presented at the annual meeting of the Society for Industrial-Organizational Psychology, New York, NY.
- Sireci, S. G., & Wells, C. S., & Keller, L. (2016, June). *Why we should abandon student growth percentiles*. Presentation delivered at the National Conference on Student Assessment as part of the symposium “How Should We Measure Growth, or Should We?” Philadelphia.
- Sireci, S. G., & Zenisky, A. L. (2006, June). *Testing linguistic minorities*. Invited presentation delivered at the Large-Scale Assessment Conference, San Francisco, CA.
- Sireci, S. G., & Zenisky, A. L. (2009, June). *Evaluating standard setting on the 2005 Grade 12 NAEP mathematics exam*. Presentation delivered at the National Conference on Student Assessment. Los Angeles, CA.
- Sireci, S. G., & Zenisky, A. L. (2016, May). *The next generation MAPT: Measuring college and career readiness*. Presentation delivered at the MCAE assessment conference, Marlboro, MA.
- Sireci, S. G., & Zenisky, A. L. (2021, July). *Validity argument for the Massachusetts Adult Proficiency Test: An example of 21st-century test validation*. Paper presented at the 9th European Conference on Methodology as part of the symposium “A Dialogue on Validity Theory, Validation Methods, and Implications for Researchers,” (J..L. Padilla, Chair), Universitat de València, Spain.
- Skorupski, W., & Sireci, S. G. (2002, April). *Current trends in computer-based testing*. Paper presented at the annual meeting of the New England Educational Research Organization, Northampton, MA.
- Suárez-Álvarez, J., & Sireci, S. G. (2024, July). *The Role of Fairness, Validity, Comparability, and Reliability in the New PISA Quality Standards*. Presentation delivered at the biennial meeting of the International Test Commission, Granada, Spain.
- Weiner, J., Sireci, S. G., Camara, W. & Stark, S. (2022, April). *New guidelines for technology-based assessments: A joint collaboration*. Presentation delivered at the annual meeting of the Society for Industrial-Organizational Psychology, Seattle.
- Wells, C. S., Hambleton, R. K., Baldwin, S., Karantonis, A., Jirka, S., Keller, R., & Sireci, S. G. (2009, June). *Evaluating population invariance (score equity) of NAEP results across states*. Presentation delivered at the National Conference on Student Assessment. Los Angeles, CA.
- Yang, Y., Sireci, S. G., & Hayes, T. L. (2013, April). *Assessments (Truly) Enhanced by Technology: Rationale, Validity, and Value*. Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, Houston, TX.

Zenisky, A., Lewis, J., O’Riordan, M., Suárez-Álvarez, J., & Sireci, S. G. (2022, April). *Tests at the service of learners and educators*. Presentation delivered at the annual conference of the Massachusetts Coalition for Adult Education [online].

Zenisky, A., Sireci, S. G., & Noonan, M. (2013, March). *Making the most of the MAPT: Integrating Assessment, Curriculum, and Instruction*. Presentation delivered at the Massachusetts Adult Basic Education Director’s Conference, Marlborough, MA.

Research Brief

Sireci, S. G., Wells, C. S., & Keller, L. A. (2016, June). Why we should abandon student growth percentiles. *Center for Educational Assessment Research Brief 16-1*. Amherst, MA: Center for Educational Assessment, University of Massachusetts. Available at http://www.umass.edu/rempe/news_SGPsResearchBrief.html.

Selected Commissioned Papers and Reports:

Auchter, J. C., Sireci, S. G., & Skaggs, G. (1993). *Technical Manual for the Tests of General Educational Development*. Washington, D.C.: American Council on Education.

Brown, J. D., & Sireci, S. G. (1995). Investigating the Stability of Item Statistics and Parameters Across GED Candidate and High School Senior Populations. *Technical Report No. 95-1*, GED Testing Service, American Council on Education, Washington, D.C.

Buckendahl, C.W., Davis, S. L., Plake, B. S., Sireci, S. G., Hambleton, R. K., Zenisky, A., & Wells, C. S. (2009). *Evaluation of the National Assessment of Educational Progress: Study reports*. Washington, DC: U.S. Department of Education.

Carey, J. D., Sireci, S. G., & Blanchard, J. (1999). *Evaluation of the Lawrence Public School Teaching and Learning Technology Project: September 1998-August 1999*. School of Education, Amherst, MA, University of Massachusetts.

Keller, L. A., & Sireci, S. G. (1998). *Annotated bibliography on teacher licensure assessment*. Princeton, NJ: Educational Testing Service. Commissioned by Educational Testing Service.

Lazzaro, C.C., Loveless, T., Sireci, S. G., & Webb, D.C. (2021, June). *Advanced Placement® calculus and physics and TIMSS Advanced 2015: Performance report*. New York: The College Board

Lukhele, R. & Sireci, S. G. (1994). Combining the multiple-choice and essay portions of the GED Writing Skills Test onto an IRT scale using a priori weights. *Technical Report No. 94-3*, GED Testing Service, American Council on Education, Washington, D.C.

Meara, K., & Sireci, S. G. (2000, August). Appraising the dimensionality of the Medical College Admissions Test. *MCAT monograph 2*. Washington, DC: Association of American Medical Colleges.

Popham, W. J., Baker, E. L., Berliner, D. C., Yeakey, C. C., Pelligrino, J.W., Quenemoen, R. F., Roderiquez-Brown, F. V., Sandifer, P. D., Sireci, S. G., & Thurlow, M. L. (2001, October). *Building tests to support instruction and accountability: A guide for policymakers*. Commission on Instructionally Supportive Assessment.

- Rizavi, S., & Sireci, S. G. (1999). Comparing computerized and human scoring of WritePlacer Essays. *Laboratory of Psychometric and Evaluative Research Report No. 354*, School of Education, University of Massachusetts, Amherst, MA.
- Sireci, S. G. (1989). *District-wide testing results*. Newark Board of Education: Newark, NJ.
- Sireci, S. G. (1991). Equating and scaling of the 1990 Accredited Personal Financial Specialist Examination. *Technical Report No. 91-1*, Examinations Division, American Institute of Certified Public Accountants, New York, NY.
- Sireci, S. G. (1993). *GED Testing Service Sensitivity Review Guidelines*. American Council on Education, Washington, D.C.
- Sireci, S. G. (1993). *Wisconsin 1993 GED Norming Study*. American Council on Education, Washington, D.C.
- Sireci, S. G., (1994). Linking the English- and Spanish-language versions of the Tests of General Educational Development: Recommendations of the GED-STEP psychometric feasibility panel. *Technical Report No. 94-1*, GED Testing Service, American Council on Education, Washington, DC.
- Sireci, S. G., (1997). *Dimensionality issues related to the National Assessment of Educational Progress*. Commissioned paper by the National Academy of Sciences/National Research Council's Committee on the Evaluation of National and State Assessments of Educational Progress, [Document Number 619]. Washington, DC: National Research Council.
- Sireci, S. G. (2004, February). Validity issues in accommodating NAEP reading tests. *Center for Educational Assessment research report no. 515*. Amherst, MA: Center for Educational Assessment, University of Massachusetts. Commissioned by Educational Testing Service.
- Sireci, S. G. (2012, December). *Smarter Balanced Assessment Consortium: Comprehensive validity agenda*. Available at <https://portal.smarterbalanced.org/library/en/comprehensive-research-agenda.pdf>
- Sireci, S. G., Baldwin, P., Martone, A., Zenisky, A., Hambleton, R. K., & Han, K. T. (2006). *Massachusetts Adult Proficiency Tests technical manual* (Center for Educational Assessment Research Report No. 600). Amherst, MA: University of Massachusetts, Center for Educational Assessment.
- Sireci, S. G., Baldwin, P., Martone, A., Zenisky, A., Hambleton, R. K., Han, K. T., Lam, W., Karia, L., & Deng, N. (2008). *Massachusetts Adult Proficiency Tests technical manual version 2* (Center for Educational Assessment Research Report No. 677). Amherst, MA: University of Massachusetts, Center for Educational Assessment.
- Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of tests accommodations on test performance: A review of the literature*. Commissioned paper by the National Academy of Sciences/National Research Council's Board on Testing and Assessment. Washington, DC: National Research Council.

- Sireci, S. G., & Rodriguez, G. (2016, December). *Evaluating the alignment of the Medición y Evaluación para la Transformación Académica to the Puerto Rico Core Standards*. Northampton, MA: Sireci Psychometric Services.
- Sireci, S. G., Rogers, H. J., Swaminathan, H., Meara, K., & Robin, F. (1997). *Evaluating the content representation and dimensionality of the 1996 Grade 8 NAEP Science Assessment*. Commissioned paper by the National Academy of Sciences/National Research Council's Committee on the Evaluation of National and State Assessments of Educational Progress, Washington, DC: National Research Council.
- Sireci, S. G., Zanetti, M. L., Slater, S., & Berger, J. B. (2001, September). *STEMTEC evaluation report for year 4. Center for Educational Assessment Report No. 426*. School of Education, University of Massachusetts, Amherst, MA.
- Vanchu, M. M. & Sireci, S. G. (1995). Analysis of differential item and differential test functioning on the GED tests: an application of SIBTEST. *Technical Report No. 95-2*, GED Testing Service, American Council on Education, Washington, D.C.
- Walker, E., Kopacci, R. M., Sireci, S.G., Humell, P., & Azumi, J. A. (1989). *Basic Skills Evaluation: 1988-1989*. Office of Planning, Evaluation, & Testing, Newark Board of Education, Newark, NJ.
- Wiley, A., & Sireci, S. G. (1994). Determining the reliability of GED Writing Skills Test scores using classical test theory. *Technical Report No. 94-2*, GED Testing Service, American Council on Education, Washington, D.C.
- Zenisky, A. L., Sireci, S. G., Lewis, J., Lim, H., O'Donnell, F., Wells, C. S., Padellaro, F., Jung, H., Banda, E., Pham, D. Hong, S., Park, Y., Botha, S., Lee, M., & Garcia, A. (2018, September). Massachusetts Adult Proficiency Tests for College and Career Readiness: Technical manual. *Center for Educational Assessment research report No. 974*. Amherst, MA: Center for Educational Assessment.
- Zenisky, A. L., Sireci, S. G., Lewis, J., Lim, H., O'Donnell, F., Wells, C. S., Padellaro, F., Jung, H., Banda, E., Pham, D. Hong, S., Park, Y., Botha, S., Lee, M., & Garcia, A. (2024, March). Massachusetts Adult Proficiency Tests for College and Career Readiness: Technical manual, Volume 2. *Center for Educational Assessment research report No. 1002*. Amherst, MA: Center for Educational Assessment.
- Zenisky, A.L., Sireci, S.G., Lewis, J., Lim, H., O'Donnell, F., Wells, C. S., Padellaro, F., Jung, H. J., Banda, E., Pham, D., Hong, S. E., Park, Y., Botha, S., Lee, M., Crespo Cruz, E., Mena Serrano, F., Gundula, A., Ketan, O'Rourke, K., Nguyen, T.-A. T., Gitiria, L., Tran, P., & Garcia, A. (2025). Massachusetts Adult Proficiency Tests - College and Career Readiness (MAPT-CCR) Technical Manual, 5th Edition. *Center for Educational Assessment Report No. 1017*. Amherst, MA: Center for Educational Assessment.
- Zenisky, A. L., Sireci, S. G., Martone, A., Baldwin, P., & Lam, W. (2009). Massachusetts Adult Proficiency Tests Technical manual supplement: 2008-2009. *Center for Educational Assessment Research Report No. 715*. Amherst, MA: University of Massachusetts, Center for Educational Assessment.

International Experience

International Advisory Committee, SweSAT, (Umea, Sweden) 2023-present
Fulbright Specialist Roster, Approved Candidate 2/24/17-2/24/20
International Scientific Advisory Board, Centre for Educ. Measurement, University of Oslo (2018-2021)
Member, Technical Working Group, Ministry of Education, Chile (2021-2022)
Visiting Professor, Classes, Workshops, and Lectures Delivered at the Following Universities:
 University of Barcelona (Spain): 2014
 University of Brasilia (Brazil): 2015
 Universidad de Chile: 2017
 Universidad de Granada (Spain): 2009, 2013, 2016, 2018
 Universidad Iberoamericana, México, 2019
 Uniwersytet Jagielloński (Jagiellonian University, Poland): 2012, 2013, 2013
 Universidad Nacional de Colombia: 2017
 University of Oviedo (Spain): 2009 (sabbatical), 2011, 2013, 2016
 Vietnam National University: 2017
Technical Assistance, Student Recruitment: Chancellor's College, University of Malawi: 2001
External Examiner, University of Granada
External Examiner, Umeå University, Sweden
External Examiner, University of Melbourne
External Examiner, University of Oslo

Selected Professional Service

President, *International Test Commission (2022-2024, President-Elect, 2020-2022, Past-President 2024-2026)*
Member, AERA, APA, & NCME Joint Committee to revise the *Standards for Educational & Psychological Testing (2024-2026)*
Past-President, *National Council on Measurement in Education (4/20-4/21)*
President, *National Council on Measurement in Education (4/19-4/20)*
Vice-President (President-Elect), *National Council on Measurement in Education (4/18-4/19)*
Council Member, *International Test Commission, 2014-2020*
Member, Executive Committee, Conference on Test Security, 2025
Co-Chair, Conference Committee, *International Test Commission 2025-present*
Member, Culturally Responsive Assessment Collaborative, 2023-2025
Member, Blue Ribbon Commission on Graduation Measures, New York State, 2022-2024
Member, Committee on Psychological Tests & Assessments, *American Psychological Association 2018-21*
Member, Board of Directors, *National Council on Measurement in Education, April 2006-April 2009*
President, *Northeastern Educational Research Association, 2006-2007 (Past-President 2007-2008).*
Co-editor, *International Journal of Testing, September 2008-August 2013*
Co-editor, *Journal of Applied Testing Technology, December 2000—May 2008*
Editorial Board, *Applied Measurement in Education, since January 1996*
Editorial Board, *Psicothema, since November 2000*
Editorial Board, *International Journal of Testing, January 2002—September 2008, 2014-present*
Editorial Board, *Educational and Psychological Measurement, since December 2004*
Editorial Board, *Educational Assessment, since 2014*
Editorial Board, *Talincrea: Talento, Inteligencia, y Creatividad, 2016-present*
Editorial Board, *Practical Assessment, Research, and Evaluation, 2019-present*
Editorial Board, *Educational Measurement: Issues and Practice, 5/2000—12/2003, 2009-2012*
Editorial Board, *European Journal of Psychological Assessment, 2005-2012*
Associate Editor, *Journal of Educational Measurement, 2019 (6 months)*
Editorial Board, *Journal of Educational Measurement, 2019-present*
Manuscript Reviewer: Several other journals including *AERA Open, Applied Psychological Measurement, Contemporary Educational Psychology, Educational Assessment, Educational Evaluation and Policy Analysis, Educational Researcher, Equity and Excellence in Education, European Journal of*

Selected Professional Service (continued)

Psychological Assessment, Exceptional Children, Journal of Applied Social Psychology, Journal of Behavioral Education, Journal of Educational Measurement, Journal of Special Education Leadership, Journal of Teacher Education, Language Testing, Psychometrika, Multivariate Behavioral Research, Psychological Reports, and Psychological Methods

Board of Directors, Northeastern Educational Research Association, 1996-1999

International Committee, Division D, American Educational Research Assoc. 2018-21 (Chair, '20)

Nominations Committee, Division D, American Educational Research Association, 2022

Chair, Membership Committee, Northeastern Educational Research Association, 2002-2003

Chair, Public Affairs Committee, Division 5, American Psychological Association, 2002-2004

Chair/Member, NCME Recruitment Committee, 1999-2003

Conference program chair, 1998 annual meeting, Division 5, American Psychological Association

Conference program co-chair, 1996 annual meeting, Northeastern Educational Research Association

Advisory Board, *NCME Newsletter*, 1999-2003

Grant Report Reviewer, U.S. Department of Education, OERI, 1996

Grant Reviewer, U.S. Department of Education, IES, 2021

Proposal Reviewer, National Assessment Governing Board, 1997

Proposal Reviewer, Social Sciences and Humanities Research Council (Canada), 2002

Proposal Reviewer for annual conferences of AERA, APA, NCME, and NERA since 1992

Member, AERA Teller's Committee, 1995

Member, AERA Division D Graduate Student Seminar Committee, 1999

Member, AERA Division D Nominations Committee 2021

Member, Conference Committee, International Test Commission, 2019-present

Member, Publications Committee of NCME, 1996-1999

Member, External Relations Committee of NCME, 1994-1996

Member, Elections Committee of NCME, 1993-1994

Member, Graduate Student Issues Committee of NCME, 1991-1993

Professional Affiliations

American Educational Research Association (Fellow; 2009)
 American Psychological Association (Fellow, Division 5; 2007)
 Association of Test Publishers
 Coalition on Adult Basic Education
 European Association of Methodology
 International Test Commission (Fellow, 2024)
 National Council on Measurement in Education
 Northeastern Educational Research Association
 Psychometric Society

Consulting

Consult with a variety of national testing organizations, professional licensure organizations, boards of education, federal government agencies, and other educational research or service organizations since 1987 (Sireci Psychometric Services, Inc.). Current and former clients include ACT, American Institute of Certified Public Accountants, Association of American Medical Colleges, Chartered Alternative Investment Analyst Association, College Board, Curriculum Associates, Educational Records Bureau, Duolingo, Educational Testing Service, Federation of State Medical Boards, the Gallup Organization, the Graduate Management Admissions Council, Microsoft, National Academy of Sciences, Newark (NJ) Board of Education, Novell, and Westfield Public Schools.

Selected Volunteer Activities

United Way volunteer, American Council on Education, 1992-1995

Guitarist and vocalist, Voices of Higher Education, American Council on Education, 1993-1995

Big Brother, Big Brothers and Big Sisters of Hampshire County, 1999-present

CCD Instructor, Sacred Heart Church, Northampton, Massachusetts, 2000-2002

Assistant (Youth) Soccer Coach, Northampton Recreation Department, 2002, 2003, 2010

T-Ball Coach, Northampton Recreation Department, 2004

Parish Council, Sacred Heart Church, 2004-2010, St. Elizabeth Ann Seton Parish, 2010-2011

Cot Shelter Team, Northampton, MA, 2004-2012, 2022-2023

Coach, Assistant Coach, Northampton Youth Soccer, Pioneer Valley Junior Soccer League, 2001-2012

Member, Circle of Care (refugee resettlement assistance), Northampton, MA, 2017-present

Doctoral Dissertations Chaired

Name	Dissertation Title	Year
Frederic Robin	Development and evaluation of test assembly procedures for computerized adaptive testing	2001
Kevin Meara	Validity issues in standard setting	2001
Kristen Huff	An item modeling approach to descriptive score reports	2003
Xiaoying Ma	An investigation of alternative approaches to scoring multiple response items on a certification exam	2004
Elias Chakwera	Content validity of independently-constructed curriculum-based examinations	2004
Shameem Khaliq	An evaluation of automated scoring programs designed to score essays	2004
Shuhong Li	Evaluating the consistency and accuracy of proficiency classifications using item response theory	2006
Bob Chulu	Equating high stakes educational measurements : A study of design and consequences.	2006
Andrea Martone	Exploring the impact of teachers' participation in an assessment-standards alignment study.	2007
Jeffrey Hauger	Empirical comparison of the Bookmark and modified Angoff standard setting methods and the impact on student classification	2007
Leah Kaira	Using Item Mapping to Evaluate Alignment between Curriculum and Assessment	2010
Timothy O'Neil	Maintenance of Vertical Scales Under Conditions of Item Parameter Drift and Rasch Model-data Misfit	2010
Matthew DiBartolomeo	Impact Of Judges' Consensus On Accuracy Of Anchor-Based Judgmental Estimates Of Multiple-Choice Test Item Difficulty: The Case Of The NATABOC Examination	2010
Katrina Crotts Rohr	Evaluating the Validity of Accommodations for English Learners through Evidence Based on Response Processes	2013
Joseph Rios	Identifying Examinees Who Possess Distinct and Reliable Subscores When Added Value is Lacking for the Total Sample	2016
Molly Faulkner-Bond	Who is like Whom? Reclassification and Performance Patterns for Different Groupings of English Learners	2016
Joshua Marland	Investigating the Impact of Student Opt Out on Value-Added Measures of Teacher Quality	2017
Yooyoung Park	Constructing multistage adaptive tests using response times	2020
Francis O'Donnell	What's in a Label? Unpacking the Meaning of Achievement Labels from Tests	2020
Maritza Casas	Measurement Invariance Across Immigrant and Non-Immigrant Populations on PISA Cognitive and Non-Cognitive Scales	2021
Sandra Botha Sweeney	Avatars & Choice: Fostering Student Agency, Engagement, and Equity in Assessment	2022
Sergio Araneda	An experiential approach to test design and validation	2022
Maura O'Riordan	Impact of special education status and text-to-speech on test-taker engagement	2023
Michael Buoniconti	Impact of Reasonable Accommodations on Disabled Student-Veteran Higher Educational Achievement	2024

Updated: December, 2025