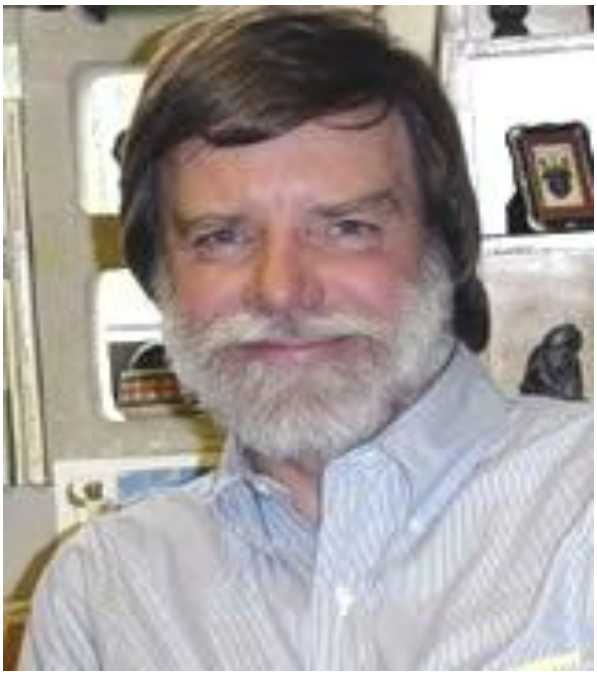
PROFESSOR LARRY LUDLOW
Larry H. Ludlow, Ph.D.
Professor and Chair
Department of Educational Research,
Measurement, and Evaluation
Boston College
Lynch School of Education



CURRICULUM VITAE

Larry Houston Ludlow
23 Duggan Drive
Framingham, MA 01702

January 27, 2010

Communications:
Phone (617) 552-4221
FAX (617) 552-1981
email ludlow@bc.edu

Education:

1979-1983 University of Chicago: The Department of Education
Degree: Doctor of Philosophy, December 9, 1983
Special Field: Measurement, Evaluation, and Statistical Analysis (MESA)
Dissertation: The Analysis of Rasch Model Residuals. Chairman: Benjamin D. Wright, Ph.D.

1973-1976 California State University, Sacramento: The Department of Psychology
Degree: Master of Arts, May 26, 1976

Special Field: Statistical analysis
Thesis: A Multidimensional Comparison of Short and Long-Term Shape Imagery
Chairman: Bruce Behrman, Ph.D.

1968-1972 California State University, Sacramento: The Department of Psychology
Degree: Bachelor of Arts, June 8, 1973
Major: Clinical psychology

1967-1968 San Juan High School, Citrus Heights, CA

1964-1967 General H.H. Arnold High School, Weisbaden, Germany

Employment History:

a) *Teaching*

1983- Boston College, Chestnut Hill, MA
Lynch School of Education
Chair, Department of Educational Research, Measurement, and Evaluation (2001-present)
Rank: Professor 2003
Associate Professor 1989-2003
Assistant Professor 1984-1989
Instructor 1983

Courses:

1. Child Development I and II - undergraduate
2. Introduction to Computer Science - undergraduate
3. Quantitative Data Collection Procedures: Theory and Practice - graduate
4. Achievement Test Construction - graduate
5. Interpretation and Evaluation of Research - graduate
6. Statistics for Behavioral Research - graduate
7. Statistics I - graduate
8. Intermediate Statistics - graduate
9. Design of Experiments - graduate
10. Psychometric Theory - graduate
11. Introduction to Multivariate Analysis – graduate
12. General Linear Models - graduate
13. Topics in Multivariate Analysis – graduate
14. Multivariate Analysis - graduate
15. Seminar in Educational Research and Measurement - graduate
16. Research Methods – undergraduate
17. Doctoral dissertation seminar – graduate
18. Visiting professor: National Institute for Education-Singapore: Rasch Models
19. Visiting professor: St. Patrick's College-Dublin, Ireland: Quantitative Methods.

b) *Other*

1981-1983 University of Chicago: The Department of Education
Title: Research Assistant
Project: 1. Taxonomic skills in understanding fiction

2. Meta-Analysis of composition literature
Supervisor: *George Hillocks, Jr*., Ph.D.

1980-1983 United States Veterans Administration Hospital (Hines, IL)
Rehabilitation Research & Development Laboratory
Title: Psychology Technician
Project: Measurement of attitudes toward blindness. Supervisor: *Ross Lambert,* M.D.

1979-1983 University of Chicago: The Department of Education
Title: Research Assistant. Supervisor: *Benjamin D. Wright,* Ph.D.

1980 University of Chicago: The Department of Education
Title: Teaching Assistant. Supervisor: *David Rogosa*, Ph.D.

1980 University of Chicago: The Department of Education
Title: Teaching Assistant. Supervisor: *Benjamin D. Wright,* Ph.D.

1977-1979 Stat-Rats: Sacramento, CA
Proprietor of research design company

1976-1979 California State University, Sacramento: Instructional Support Group (Computer Center)
Title: Statistical Consultant. Supervisor: *Linda Downing,* M.A.

1976 Research Consulting Service: Sacramento, CA
Title: Research Assistant. Supervisor: *Neal Grossen,* Ph.D.

1975-1976 California State University, Sacramento: The Department of Psychology
Title: Graduate Assistant. Supervisor: *John Doolittle*, Ph.D.

1973-1976 Stanford Lathrop Memorial Home: Sacramento, CA
Residential child-care counselor. Supervisor: *Vernon Rowe*, M.A.

1972 California Youth Authority (Eldorado Ward): Stockton, CA
Counselor in Transactional Analysis/Gestalt Therapy. Supervisor: *Lynn Fitzpatrick,* M.A.

Publications:

a) Articles

Eris O, Chachra D, Chen H, Sheppard S, Ludlow L, Rosca C, Bailey T & Toye G (under review). Outcomes of a longitudinal administration of the Persistence in Engineering Survey.
Gable RK, Ludlow LH, McCoach DB & Kite SL (under review). Validation of the Survey of Knowledge of Internet Risk and Internet Behavior.
Chapman L & Ludlow LH (under review). Can downsizing college class sizes augment student outcomes: An investigation of the effects of class size on student learning.
Haley SM, Coster WJ, Kao YC, Dumas HM, Fragala-Pinkham MA, Kramer JM, Ludlow LH & Moed (**in press**). Lessons from the Pediatric Evaluation of Disability Inventory (PEDI): Where do we go from here? *Pediatric Physical Therapy*.

Cochran-Smith M, Mitescu E, Shakman K and the Boston College TNE Evidence Team. (**in press**). Just measures:
Social justice as a teacher education outcome. *Teacher Education and Practice*.
Cochran-Smith M and the Boston College TNE Evidence Team. (2009). "Re-culturing" teacher education: Evidence, Inquiry and Action. *Journal of Teacher Education*, 5, 458-468.
Morgan M, Ludlow LH, Kitching K, O'Leary M & Clarke A. (2009). What makes teachers tick? Sustaining events in new teachers lives. *British Educational Research Journal*, iFirst Article, 1-18. http://dx.doi.org/10.1080/01411920902780972
Haley SM, Ni PS, Jette AM, Tao W, Moed R, Meyers D & Ludlow LH (2009). Replenishing computer adaptive testing (CAT) of daily activities: Analyzing the impact of pre-test items. *Quality of Life Research*. 18: 461-471. http://dx.doi.org/10.1007/s11136-009-9463-5
Vasilyeva M, Ludlow LH, Casey B & St. Onge C (2009). Examination of the psychometric properties of the Measurement Skills Assessment (MeSA). *Educational and Psychological Measurement*, 69, 106-130.
Dumas HM, Haley SM & Ludlow LH (2008). Achieving a minimally important difference in physical function
during pediatric inpatient rehabilitation. *International Journal of Rehabilitation Research*, 31, 257-260.
Enterline S. Cochran-Smith M, Ludlow LH & Mitescu E. (2008). Learning to teach for social justice:
Measuring change in the beliefs of teacher candidates. *The New Educator*, 4, 267-290.
Ludlow LH, Pedulla J, Enterline SE, Cochran-Smith M, Loftus F, Salomon-Fernandez Y & Mitescu E. (2008). From students to teachers: Using surveys to build a culture of evidence and inquiry. *European Journal of Teacher Education*, 31(4), 319-337.
Kharasch VS, Dumas HM, Haley SM, Wright EA, Ludlow LH, Jones DT & O'Brien JE. (2008)
Bronchoscopy findings in children and young adults with tracheostomy due to congenital anomalies and neurological impairment. *Journal of Pediatric Rehabilitation Medicine: An Interdisciplinary Approach.* 1, 137-143.
Ludlow, LH, Enterline, S & Cochran-Smith, M (2008) Learning to Teach for Social Justice—Beliefs scale: An application of Rasch measurement principles. *Measurement and Evaluation in Counseling and Development*, 20, 194-214.
Haley SM, Pengsheng N, Ludlow LH & Fragala-Pinkham M. (2006) Measurement precision and efficiency of multidimensional computer adaptive testing of physical functioning using the Pediatric Evaluation of Disability Inventory. *Archives of Physical Medicine and Rehabilitation*, 87, 1223-1229.
Burns S. & Ludlow LH. (2006). Understanding student evaluations of teaching quality: The unique contributions of class attendance. *Journal of Personnel Evaluation in Education*. http://dx.doi.org/10.1007/s11092-006-9002-7
Maloy RW, Pine GJ, Seidman I & Ludlow LH. (2006). Arriving on a fast Track: Perceptions of teachers from
alternative, campus-based and PDS teacher preparation programs about their first four years in the classroom. *The Teacher Educator,42*, 106-121.
Ludlow LH, Haley SM & Andres P. (2005). Measuring change: Functional activity item calibrations at
admission vs. discharge. In Bezruczko, N. (ed.) *Rasch Measurement in Health Sciences*. Maple Grove.
MN: JAM Press.

Mahalik JR, Morray EB, Coonerty-Femiano A, Ludlow LH, Slattery S & Smiler A. (2005). Development of
the Conformity to Feminine Norms Inventory. *Sex Roles, 52(7/8),* 417-435.
Ludlow LH (2005). A longitudinal approach to understanding course evaluations. *Practical Assessment Research and Evaluation,* http://pareonline.net/pdf/v10n1.pdf.
Coster WJ, Haley SM, Ludlow LH, Andres PL & Ni PS. (2004) Development of an applied cognition scale to measure rehabilitation outcomes. *Archives of Physical Medicine and Rehabilitation, 85, 2030-2035.*
Dumas HM, Haley SM, Ludlow LH & Carey TM. (2004). Recovery of ambulation during inpatient rehabilitation:
Physical therapist prognosis for children and adolescents with traumatic brain injury. *Physical Therapy 84(3),* 232-242.
Haley SM, Coster WJ, Andres PL, Ludlow LH, Ni P, Bond TLY, Sinclair SJ & Jette AM. (2004).
Activity Outcome Measurement for Post-acute Care. *Medical Care,* 42, 49-61.
Coster WJ, Haley SM, Andres PL, Ludlow LH, Bond TLY & Ni P. (2004). Refining the conceptual basis
for rehabilitation outcome measurement: Personal Care and Instrumental items. *Medical Care,* 42. 62-72.
Dumas HM, Haley SM, Carey TM, Ludlow LH, & Rabin JP. (2003). Lower extremity spasticity as an early marker
of ambulatory recovery following traumatic brain injury. *Child' s Nervous System,* 19, 114-118.
Kharasch VS, Haley SM, Dumas HM, Ludlow LH & O'Brien JE. (2003). Oxygen and ventilator weaning during inpatient pediatric pulmonary rehabilitation. *Pediatric Pulmonology,* 35, 280-287.
Mahalik JR, Locke B, Ludlow L, Diemer M, Scott PJ, Gottfried M, & Freitas G. (2003). Development of the Conformity to Masculine Norms Inventory. *Psychology of Men and Masculinity, 4,* 3-25.
Ludlow LH, Shirley D & Rosca C. (2002). The case that won't go away: Besieged institutions and the Massachusetts teacher tests. *Educational Policy Analysis Archives,* 10(50). http://epaa.asu.edu/epaa/v10n50.html.
Ludlow LH (2002). Rethinking practice: Using faculty evaluations to teach statistics *Journal of Statistics Education,* 10(3). www.amstat.org/publications/jse/v10n3/ludlow.html.
Dumas HM, Haley SM, Ludlow LH, Rabin JP. (2002). Functional recovery in pediatric traumatic brain injury during inpatient rehabilitation. *American Journal of Physical Medicine and Rehabilitation,* 81(9), 661-669.
Ludlow LH. (2002). Residuals: Trash or treasure? *Popular Measurement,*4, 1-7. http://www.rasch.org/publish.htm.
Haley SM, Ludlow LH, & Kooyoomjian JT. (2002). Extending the range of functional assessment in older adults: Development of the Late Life Function and Disability Instruments. *Journal of Aging and Physical Activity,* 10, 453-465.
Haley SM, Ludlow LH, Coster WJ, Langmuir L. (2002). Self- reporting of capable versus typical functional activity performance in community-dwelling older adults: Is there a difference? *Journal of Geriatric Physical Therapy, 25*; 1:02, 3-10.
Ludlow LH & Mahalik JR. (2001). Congruence between a Theoretical Continuum of Masculinity and the Rasch Model: Examining the Conformity to Masculine Norms Inventory. *Journal of Applied Measurement, 2,* 205-221.

Haley SM, Dumas HM & Ludlow LH. (2001). Variation by diagnostic and practice pattern groups in the mobility outcomes of inpatient rehabilitation program for children and adolescents. *Physical Therapy, 81,* 1425-1436.
Ludlow LH & Alvarez-Salvat R. (2001). Spillover in the academy: Marriage stability and faculty evaluations. *Journal of Personnel Evaluation in Education, 15:2,* 111-119.
Ludlow LH (2001). Teacher test accountability: From Alabama to Massachusetts. *Education Policy Analysis Archives, 9* (6). http://epaa.asu.edu/epaa/v9n6.html.
Ludlow LH & Alvarez-Salvat R. (2000). Fechner: The man in the mask. *Popular Measurement, 3,* 5-6.
Ludlow LH & Haley SM. (2000). New directions in pediatric rehabilitation measurement: The growing challenge. *Journal of Outcome Measurement, 4,* 482-490.
Ludlow LH. (1999). The structure of the Job Responsibilities Scale: A multi-method analysis. *Educational and Psychological Measurement, 59,* 962-975.
Coster WJ, Mancini MC & Ludlow LH. **(**1999**).** Factor structure of the School Function Assessment. *Educational and Psychological Measurement, 59,* 665-677.
Ludlow LH & O'Leary M. (1999). Omitted and not reached items: Practical data analysis implications. *Educational and Psychological Measurement, 59,* 615-630.
Ludlow LH & Haley KC. (1999). Newton: The pinball wizard?. *Popular Measurement, 2,* 5-7.
Coster W, Ludlow LH & Mancini M. (1999). Using IRT variable maps to enrich understanding of rehabilitation data. *Journal of Outcome Measurement, 3,* 123-133.
Ludlow LH. (1999). Putting the psych in psychometrics. *Popular Measurement, 2,* 50-51.
Ludlow LH & Lunz M. (1998). The Job Responsibilities Scale: Invariance in a longitudinal prospective study. *Journal of Outcome Measurement, 2,* 326-337.
Ludlow LH. (1998). Galton: The first psychometrician?. *Popular Measurement, 1,* 13-14.
Ludlow LH. (1998). Scale invariance from a three-dimensional graphical perspective: Visualizing an eigenvector. *Educational and Psychological Measurement, 58,* 166-178.
Gonzalez EJ, Adams RJ, Wu M & Ludlow LH.(1997). Item response theory in the Third International Mathematics and Science Study. In: Wilson M, Engelhard Jr. G & Draney K. (eds.). *Objective Measurement: Theory into Practice*. Volume 4.Norwood, NJ:, Ablex Publishing.
Ludlow,L.H. & Bell,K.N. (1996). Psychometric characteristics of the Attitudes Towards Mathematics and its Teaching (ATMAT) Scale. *Educational and Psychological Measurement, 56,* 864-880.
Ludlow,L.H. (1996). Instructor evaluation ratings: A longitudinal analysis. *Journal of Personnel Evaluation in Education,* 10, 83-92.
Ludlow,L.H. & Haley,S.M. (1996). Effect of context in rating of mobility activities in children with disabilities. *Educational and Psychological Measurement, 56,* 122-129.
Ludlow,L.H. & Haley,S.M. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement, 55,* 967-975.
Ludlow,L.H. & Haley,S.M. (1995). Displaying change in functional performance. In: Engelhard,G. & Wilson,M. (eds.). *Objective Measurement: Theory into Practice*. Volume 3. Norwood, NJ:, Ablex Publishing.
Fisher,A.G., Bryze,K.A., Granger,C.V. Haley,S.M., Hamilton,B.B., Heineman,A.W., Puderbaugh,J.K., Linacre,J.M., Ludlow,L.H., McCabe,M.A. & Wright,B.D. (1994). Applications of conjoint measurement to the development of functional assessment. *International Journal of Educational Research, 21,* 579-593.
Ludlow,L.H. (1994). An empirical cross-validation of alternative classification strategies applied to harness racing data for win bets. In: Hausch,D.B., LO,V.S.Y. & Ziemba,W.T. (eds). *Efficiency of Racetrack Betting Markets*. Academic Press.

Haley,S.M., Ludlow,L.H. & Coster,W.J. (1993). Pediatric Evaluation of Disability Inventory: Clinical interpretation of summary scores using Rasch rating scale methodology. In Gresham, G.E. & Granger, C.V. (eds). *New Developments in Functional Assessment* (*Physical Medicine and Pediatric Clinics of North America).* Philadelphia, W.B. Saunders, 529-540.
Haley,S.M. & Ludlow,L.H. (1992). Applicability of the hierarchical scales of the Tufts Assessment of Motor Performance for school-aged children and adults with disabilities (with Commentary and Author Response). *Physical Therapy, 72,* 191-202.
Ludlow,L.H., Haley,S.M. & Gans,B.M. (1992). A hierarchical model for functional performance in rehabilitation medicine: The Tufts Assessment of Motor Performance. *Evaluation and the Health Professions, 15,* 59-74.
Ludlow,L.H. & Haley,S.M. (1992). Polytomous Rasch models for behavioral assessment: The Tufts Assessment of Motor Performance. In Wilson,M. (ed). *Objective Measurement: Theory into Practice*. Volume 1. Norwood, NJ:, Ablex Publishing.
Ludlow,L.H. & Guida,F.V. (1991) The Test Anxiety Scale for Children as a generalized measure of academic anxiety. *Educational and Psychological Measurement, 51,* 1013-1022.
Haley,S.M., Costner,W.J. & Ludlow,L.H. (1991). Pediatric functional outcome measures. In Jaffe, K.M. (ed). *Pediatric Rehabilitation (Physical Medicine and Rehabilitation Clinics of North America).* Philadelphia, W.B. Saunders, 689-723.
Haley,S.M., Ludlow,L.H., Gans,B.M., Faas,R.M. & Inacio,C.A. (1991). Tufts Assessment of Motor Performance: An empirical approach towards identifying motor performance categories. *Archives of Physical Medicine and Rehabilitation, 72,* 359-366.
Theall,M., Franklin,J., & Ludlow, L.H. (1990). Attributions and retributions: Student ratings and the perceived causes of performance. *Instructional Evaluation, 11,* 12-17.
Ludlow,L.H. & Hwang,R. (1990). Evaluating district-level performance relative to the system. *Educational Research Quarterly, 14,* 29-37.
Gable,R.K., Ludlow,L.H. & Wolf,M.B. (1990). The use of Classical and Rasch latent trait models to enhance the validity of affective measures. *Educational and Psychological Measurement, 50,* 869-878.
Ludlow,L.H. & Howard,E.P. (1990). The family map: A graphical representation of family systems theory. *Educational and Psychological Measurement, 50,* 245-254.
Guida,F. & Ludlow,L. (1989). A cross-cultural study of test anxiety. *Journal of Cross-Cultural Psychology, 20,* 178-190.
Ludlow,L.H. (1987). The graphical representation of quantitative research synthesis residual variation. *Educational and Psychological Measurement, 47,* 941-952.
Ludlow,L.H. & Levy,S. (1987). Personal space as a function of infant illness: An application of multidimensional scaling. In Hair,J., Anderson,R. & Tatham,R. (eds). *Multivariate Data Analysis: With Readings.* Macmillan.
Ludlow,L.H. (1986). Graphical analysis of item response theory residuals. *Applied Psychological Measurement, 10,* 217-229.
Casey,M.B., Brabeck,M.M. & Ludlow,L.H. (1986). Familial handedness and its relation to spatial ability following strategy instruction. *Intelligence, 10,* 389-406.
Ludlow,L.H. (1985). A strategy for the graphical representation of Rasch model residuals. *Educational and Psychological Measurement, 45,* 851-859.
Ludlow,L.H. & Hillocks,Jr., G. (1985). Psychometric considerations in the analysis of reading skill hierarchies. *Journal of Experimental Education, 54,* 15-21.
Guida,F., Ludlow,L.H., & Wilson,M. (1985). The mediating effect of time-on-task on the academic anxiety/achievement interaction: A structural model. *Journal of Research and Development in Education,* 19, 21-26.
Hillocks,G. & Ludlow,L.H. (1984). A taxonomy of skills in reading and interpreting fiction. *American Educational Research Journal, 21,* 7-24.

Ludlow, L.H. & Levy,S. (1984). Personal space as a function of infant illness: An application of multidimensional scaling. *Journal of Pediatric Psychology, 9,* 331-347.
Courington,S.M., Lambert,R.W., Ludlow,L.H., Wright,B.D. & Becker,S.W. (1983). The measurement of attitudes toward blindness and its importance for rehabilitation. *International Journal of Rehabilitation Research, 6,* 67-72.
Hovsepian,B.A., Bond,N.A. & Ludlow,L.H. (1982). Multiattribute evaluation in baseball All-Star predictions. *Behavioral Science, 27,* 273-280.
Faria,I., Ludlow,L.H. & Frankel,M. (1981). An alternative scaling model for the prediction of body density in female athletes. *Journal of Sports Medicine and Physical Fitness, 21,* 336-341.
Bezruczko,N. & Ludlow,L.H. (1981). The alterability and measurement of learning rate. *Studies in Educational Evaluation, 7,* 97-104.
Bezruczko,N. & Ludlow,L.H. (1980). The alterability and measurement of learning rate. In B.S. Bloom (ed). *The state of research on selected alterable variables in education*. Chicago: The University of Chicago, Department of Education.

b) Books
Ludlow, L.H. (**in progress**). *Multivariate Statistics: Creative Design and Analysis.*
Gans,B.M., Nanna,M., Maiers,B., Trumbo,L., Dijkers,M., Haley,S.M., Ludlow,L.H. (1996). *Michigan Motor Performance Assessment*. Rehabilitation Institute of Michigan.
Haley,S.M., Coster,W.J., Ludlow,L.H., Haltiwanger,J.T. & Andrellos,P.J. (1992). *Pediatric Evaluation of Disability Inventory (PEDI): Development, Standardization and Administration Manual*. Boston, MA: New England Medical Center/PEDI Research Group.

c) Evaluation and Project Reports
Hudson Public Schools, Hudson, MA. *Teaching American History: Second Year Report*. PI: Todd Wallingford.
2009.
U.S. Department of Education, Institute of Education Sciences, Office of Indian Education and the National Center
for Education Statistic (2008). *National Indian Education Study Part I: Performance of American Indian and Alaska Native Students at Grades 4 and 8 on NAEP 2007 Reading and Mathematics Assessments.* NCES 2008-457.
U.S. Department of Education, Institute of Education Sciences, Office of Indian Education and the National Center
for Education Statistic (2006). *National Indian Education Study Part II: The Educational Experiences of American Indian and Alaska Native Students in Grades 4 and 8*. NCES 2008-458.
Chachra D, Eris O, Chen H, Donaldson K, Rosca C, Ludlow L, Sheppard S. 2008. Gender and academic self-confidence: results from the Persistence in Engineering survey. Center for the Advancement of Engineering Education, Academic Pathways Study. Technical Report CAEE-TR-08-07.
Chachra D, Eris O, Chen H, Donaldson K, Rosca C, Ludlow L, Sheppard S. 2008. Gender and academic
disengagement: results from the Persistence in Engineering survey. Center for the Advancement of Engineering Education, Academic Pathways Study. Technical Report CAEE-TR-08-06.
U.S. Department of Education, Institute of Education Sciences, Office of Indian Education and the National Center

for Education Statistics (2006). *Part I: The Performance of American Indian and Alaska Native Fourth- and Eight-Grade Students on NAEP 2005 Reading and Mathematics Assessment.* NCES 2006-463.
U.S. Department of Education, Institute of Education Sciences, Office of Indian Education and the National Center
for Education Statistics (2006). *Part II: The Educational Experiences of Fourth-and Eighth-Grade American Indian and Alaska Native Students.* NCES 2007-454
Bauer, S., Ludlow, L.H. & Stage, F. (2005). University of Massachusetts, Amherst; School of Education; Education
Policy, Research and Administration Department: *Academic Quality Assessment and Development External Review*. Chair: Joseph Berger. September 26-27, 2005.
Maloy, R.W., Pine, G., Seidman, I. & Ludlow, L.H. (2004). *Massachusetts Teacher Preparation and Induction Study Report on Third Year Findings*. Coordinator: Karen O'Connor, University of Massachusetts, Dartmouth.
Maloy, R.W., Pine, G., Seidman, I. & Ludlow, L.H. (2003). *Massachusetts Teacher Preparation and Induction Study Report on Second Year Findings*. Coordinator: Karen O'Connor, University of Massachusetts, Dartmouth.
Maloy, R.W., Pine, G., Seidman, I. & Ludlow, L.H. (2002). *Massachusetts Teacher Preparation and Induction Study Report on First Year Findings*. Coordinator: Karen O'Connor, University of Massachusetts, Dartmouth.
Ludlow,L.H. & Scanlon,D.J. (1999). Project NewBCTech. Newton Public Schools, MA.
Ludlow,L.H. (1994). Omitted and not reached items. *Summary of the technical advisory committee (Appendix F)*. Boston: TIMSS International Study Center.
Haney, W., Ludlow, L., Raczek, A., Stryker,S. and Jones, A. (1994). *Calibrating Scores on Two Tests of Adult Literacy: An Equating Study of the Test of Adult Literacy Skills (TALS) Document Test and the Comprehensive Adult Student Assessment System (CASAS) GAIN Appraisal Reading Test (Form 2)* . (Report prepared for the Manpower Demonstration Research Corporation) Chestnut Hill, MA: Boston College.
Ludlow,L.H., Ross-McBride,G. & November,A. (1986). *The Belmont Public Schools Computer Education*
Program, K-1. Final Evaluation Report: Belmont, MA.
Ludlow,L.H. (1979). *Reduction of fear and violence in the San Juan District*. ESEA, Title IV, Part C, Project No. 3940, California State Department of Education.

d) Abstracts
Ludlow,L.H. & Guida,F.V. (1990). The Test Anxiety Scale for Children: A measure of academic anxiety? Program Abstracts. New England Educational Research Organization.
Ludlow,L.H. & Haley,S.M. (1989). The Tufts Assessment of Motor Performance: An application of the Rasch rating scale model. Program Abstracts. New England Educational Research Organization.
Ludlow,L.H. & Howard,E. (1988). The family map: A graphical representation of family systems theory. Program Abstracts, New England Educational Research Organization.
Guida,F.V. & Ludlow,L.H. (1986). A cross-cultural study of test anxiety. *Behavior Today Newsletter*, 17, No. 35 & 36.

e) Book Reviews
Baker, F. The Basics of Item Response Theory. [Heineman Press, 1985.] *Educational Measurement*, Fall 1986.

f) Other

Ludlow, L.H., & Trong, K. (Eds.) (2004). A Longitudinal Analysis of One Professor's Course Evaluations: Vol. V. (Monograph). Chestnut Hill: MA. Boston College Press.
Ludlow, L.H., & Trong, K. (Eds.) (2003). A Longitudinal Analysis of One Professor's Course Evaluations: Vol. IV. (Monograph). Chestnut Hill: MA. Boston College Press.
Ludlow, L.H., & Rosca, C. (Eds.) (2002). A Longitudinal Analysis of One Professor's Course Evaluations: Vol. III. (Monograph). Chestnut Hill: MA. Boston College Press.
Ludlow, L.H., & Rosca, C. (Eds.) (2001). A Longitudinal Analysis of One Professor's Course Evaluations: Vol. II. (Monograph). Chestnut Hill: MA. Boston College Press.
Ludlow, L.H., Alvarez-Salvat, R. & Rosca, C. (Eds.) (2000). A Longitudinal Analysis of One Professor's Course Evaluations: Vol. I. (Monograph). Chestnut Hill: MA. Boston College Press.
Ludlow, L.H. (April 2000). What students see in statistics. *Educational Statistician's Newsletter*. http://comp.uark.edu/~seanm/.
Ludlow,L.H., Ferrucci,B., Gonzalez E.J., Wallace,D. & Carreras, I. (1996). Statistical Computing Laboratory Workbook: Student and Instructor Guides. Boston College: School of Education, Educational Research, Measurement and Evaluation program.
Ludlow,L.H. (1992). SCALE: A program for ordered response category data.(User's manual). Boston College: School of Education, Educational Research, Measurement and Evaluation program.
Ludlow,L.H. & Howard,E.P. (1988). The family map: A graphical representation of family systems theory. *The New England Educator*, Winter, 6-18.
Ludlow,L.H. (1986). The graphical representation of Rasch model residuals. *The New England Educator*, Spring, 4-11.
Wright,B.D., Mead,R.J. & Ludlow,L.H. (1980). KIDMAP: Person-by-item interaction mapping. Research Memorandum No. 29. MESA Psychometric Laboratory, Department of Education, University of Chicago.

Academic Activities/Presentations:
Organizer, Twelfth New England Objective Measurement Workshop. Topics in Item Response Theory and
Educational Measurement I and II. In conjunction with New England Educational Research Organization annual meeting. Hyannis, MA. May 8, 2008.
Gable R, Ludlow LH, Kite SL & McCoach B. Development and validation of the Survey of Internet Risk and
Behavior. Paper presented at the American Educational Research Association annual meeting. San Diego, CA. April 16, 2009.
Cannady M, Pedulla J, Ludlow LH & Chappe S. Does a stand-alone classroom assessment course make a difference?
Paper presented at the American Educational Research Association annual meeting. San Diego, CA. April 15, 2009.
Ludlow LH, Pedulla J, Mitescu E, Cannady M, Chappe S, Hu J & Enterline S. Measuring the "effect" of a culture of
evidence. Paper presented at the American Educational Research Association annual meeting. San Diego, CA. April 14, 2009.
O'Leary M & Ludlow LH. Learning to teach for social justice: Comparative data from Ireland, New Zealand, Puerto
Rico and US. Paper presented at the Development and InterCultural Education Conference. Froebel College
of Education. Sion Hill, Blackrock, Co. Dublin, Ireland. February 27-28, 2009.

Cochran-Smith, M & Ludlow, L.H. Creating a culture of inquiry, evidence and action in teacher education.
Presentations prepared for Pontificia Universidad Catolica de Valparaiso, Universidad Diego Portales, and
Pontificia Universidad Catolica de Chile. Valparaiso and Santiago, Chile. November 25-27, 2008.
Visiting Professor, "Research Methods (2): Building a powerful design". Course co-taught with Marilyn Cochran-
Smith, St Patrick's College, Dublin, Ireland. June 30-July 4, 2008.
Organizer, Eleventh New England Objective Measurement Workshop. Two sessions of Rasch model applications by
Boston College students. In conjunction with New England Educational Research Organization annual meeting. Hyannis, MA. April 11, 2008.
Enterline S, Ludlow LH, Mitescu E & Cochran-Smith M. Learning to Teach for Social Justice: Measuring changes
in beliefs. Paper presented at the American Educational Research Association annual meeting. New York, NY. March 26, 2008.
Cochran-Smith M & Ludlow LH. A series of workshops on quantitative and qualitative procedures for program
improvement and accountability. The University of Auckland, New Zealand. March 10-13, 2008.
Eris O, Chachra D, Chen H, Ludlow LH, Rosca C, Sheppard S & Donaldson K. A preliminary analysis of
correlates of engineering persistence. In: *Proceedings of the American Society for Engineering Education*
Annual Conference. Honolulu, HI. June 24-27, 2007.
Cochran-Smith, M. & Ludlow, L.H. (Symposium co-chairs). The Role of Research in Teaching and Teacher
Education in Ireland. Colleges of Education Research Consortium (CERC) Annual Meeting. Marino Institute of Education, Dublin, Ireland. June 20, 2007.
Cochran-Smith, M. & Ludlow, L.H. Transforming a teacher education program: Teachers for a New Era (TNE) at
Boston College. Professional Development Workshop. Church of Ireland College of Education and Froebel College, Dublin Ireland. June 18, 2007.
Cochran-Smith, M. & Ludlow, L.H. Using evidence to improve teacher education: A workshop. University of
Glasgow and University of Strathclyde, Glasgow, Scotland. June 15, 2007.
Cochran-Smith, M. & Ludlow, L.H. Boston College Teachers for a New Era (BC-TNE) and Scottish Teachers for a
New Era (S-TNE): A workshop. University of Aberdeen, Aberdeen, Scotland. June 13, 2007.
Organizer, Tenth New England Objective Measurement Workshop. One session of Rasch model applications by
Boston College students. In conjunction with New England Educational Research Organization annual meeting. Portsmouth, NH. April 27, 2007.
Ludlow LH, Enterline S, Pedulla J, Kafka A, Solomon-Fernandez Y& Mitescu E. Survey development and
utility of teacher candidates'/graduates' perceptions. Paper presented at the New England Educational Research Organization annual meeting. Portsmouth, NH. April 26, 2007.
Ludlow LH, Pedulla J, Enterline S, Loftus F, Kafka A, Salomon-Fernandez Y, Mitescu E & Gilligan, J. From

students to teachers: Teacher candidates'/graduates' perceptions and experiences. Paper presented at the American Educational Research Association annual meeting. Chicago, IL. April 12, 2007.
Morgan M, Ludlow LH, Kitching K, O'Leary M. What makes teachers tick? Understanding the motivation of classroom teachers. Paper presented at the American Educational Research Association annual meeting. Chicago, IL. April 13, 2007
Ludlow, L.H. Building a value-added model for teacher education: Observations about implementation. University
of Puerto Rico, San Juan, PR. March 8, 2007.
Cochran-Smith, M. & Ludlow, L.H. Development of a TNE-like teacher education research program. University of Puerto Rico, San Juan, PR. March 6, 2007.
Cochran-Smith, M & Ludlow, LH. Data access and collaboration with public school district offices. Second Annual
Meeting: Learning Network—Teachers for a New era. Philadelphia, PA. December 1, 2006.
Eris, O., Chachra, D., Ludlow, L.H. & Rosca, C. Longitudinal analysis of Cohort I Persistence in Engineering survey
data. 2006 Center for the Advancement of Engineering Education Academic Pathways Study meeting.
Stanford, CA. September 21, 2006.
Haley, S., Ni, P., Jette, A.M., Tao, W., Moed, R. & Ludlow, L.H. Replenishing CAT: Analyzing the impact of pre-
test items. Paper presentation at The Inaugural PROMIS Conference Patient-Reported Outcome Measures:
Gaithersburg, MD. September 12, 2006.
Organizer, Ninth New England Objective Measurement Workshop. Two sessions of Rasch model applications by
Boston College students and alumni. In conjunction with New England Educational Research
Organization annual meeting. Portsmouth, NH. April 28, 2006.
Ludlow, L.H., Enterline, S., Cochran-Smith, M. Learning to Teach for Social Justice: An Application of Rasch
Measurement Principles. Paper presentation at the American Educational Research Association. San Francisco, CA. April 10, 2006.
Ludlow,L.H. & Betebenner, D. Value-added models applied to teacher education. Paper presentation at the
American Educational Research Association. San Francisco, CA. April 8, 2006.
Ludlow, L.H. Faculty evaluations of teaching: A link between standardized ratings and student-centered drawings.
Paper presented at the 4th Annual Hawaii International Conference on Education. Waikiki, Hawaii. January 9, 2006.
Ludlow, L.H. & Enterline, S. Rasch analysis of BC TNE social justice questions. Paper presented at the New
England Educational Research Organization annual meeting. Northampton, MA. April 29, 2005.
Organizer, Eighth New England Objective Measurement Workshop. One session of Rasch model applications by
Boston College students and alumni. In conjunction with New England Educational Research
Organization annual meeting. Northampton, MA. April 29, 2005.

Ludlow,L.H. Value added modeling as designed for the Boston College Teachers for a New Era Project.
Presented at the BC TNE organized symposium titled "From teacher education to pupil learning: Evidence matters" at the New England Educational Research Organization annual meeting. Northhampton, MA. April 28, 2005.
Ludlow,L.H., Barrett,J., Salomon,Y., Jong,C. & Shakman,K. Teachers for a New Era evidence issues.
Roundtable presentation at the American Educational Research Association annual meeting. Montreal, Canada. April 14, 2005.
Ludlow,L.H., Enterline,S. & Famularo,L. Student drawings and student ratings of instruction: When numbers
bow to figures. Paper presented at the American Educational Research Association annual meeting. Montreal, Canada. April 14, 2005.
Ludlow,L.H., Famularo,L, & Enterline,S. Drawing conclusions III: Student drawings to inform teaching and
learning in a university setting. Roundtable presentation at the American Educational Research Association annual meeting. Montreal, Canada. April 13, 2005.
Ludlow,L.H., Diaconu,D., Erberber,E. & Muncaster,K. Does instructor availability matter: The size, time, outclass
model by degree. Paper presented at the American Association of Colleges for Teacher Education annual meeting. Washington, DC. February 21, 2005.
Ludlow,L.H. Consequences of federally mandated teacher testing: A longitudinal case study from Massachusetts.
Paper presented at the International Council on Education for Teaching World Assembly. Hong Kong
Special Administrative Region, China. July 15, 2004.
Chair, "Standards". Paper session at International Council on Education for Teaching World Assembly. Hong Kong
Special Administrative Region, China. July 14, 2004.
Organizer, Seventh New England Objective Measurement Workshop. Two sessions of Rasch model applications by
Boston College students. In conjunction with New England Educational Research Organization annual meeting. Portsmouth, NH. April 22, 2004.
Organizer and Chair, Symposium Session 7.3: "Creative applications of multivariate techniques". Paper presentations by Boston College, LSOE, ERME students. New England Educational Research Organization annual meeting. Portsmouth, NH. April 24, 2004.
Ludlow,L.H., Bebell,D. & Trong, K. What do students see in statistics classes. Poster presented at New England Educational Research Organization annual meeting. Portsmouth, NH. April 23, 2004.
Ludlow,L.H., Bebell,B. & Trong,K. Drawings conclusions II: Using student drawings to inform teaching and learning in a university setting. Paper presented at the American Educational Research Association annual meeting. San Diego, CA. April 14, 2004.
Leibowitz S. & Ludlow LH Measuring change in literacy instruction: The Bayside Readers Initiative Classroom Observations. Paper presented at the American Educational Research Association annual meeting. San Diego, CA. April 14, 2004.
Moderator, "MTEL and Short-term/Long-term Institutional Strategies". Framingham State College's 4th Annual
Conference on Massachusetts Teacher Test preparation. Framingham, MA. January 9, 2004.
Ludlow,L.H. & Bebell,D. Drawing conclusions: Using student drawings to inform teaching and learning in a

university setting. Paper presented at the American Educational Research Association annual meeting. Chicago, IL. April 23, 2003.
Ludlow,L.H., Mahalik,J. & Rosca,C. Examination of Rasch structure of feminine identity scales. Paper
presented at the American Educational Research Association annual meeting. Chicago, IL. April 24, 2003.
Ludlow,L.H. & Shirley,D. Street-level bureaucrats and the Massachusetts Teacher Tests. Paper
presented at the American Educational Research Association annual meeting. Chicago, IL. April 25, 2003.
Organizer, Sixth New England Objective Measurement Workshop. Two sessions of Rasch model applications by
Boston College students. In conjunction with NEERO annual meeting. Northampton, MA. April 11, 2003.
Ludlow LH, Shirley D & Rosca C. The case that won't go away: Besieged institutions and the Massachusetts
teacher tests. Paper presented at the American Association of College for Teacher Education annual meeting. New Orleans, LA. Jan. 27, 2003.
Ludlow LH (2002). Critiquing government-imposed teacher testing: A case study from the US. Paper
presented at the University Council for the Education of Teachers. England. Nov. 11, 2002.
Organizer, Fifth New England Objective Measurement Workshop. Two sessions of Rasch model applications by
Boston College students. In conjunction with NEERO annual meeting. Northampton, MA. April 25, 2002.
Rosca, C., Ludlow, L.H. & Mahalik, J. The psychometric structure of the Conformity to Feminine
Norms Inventory. Paper presented at the New England Educational Research Organization annual
Meeting. Northampton, MA. April 25, 2002.
Ludlow,L.H. A structural model for understanding faculty evaluations. Paper presented at the American Educational
Research Association annual meeting. New Orleans, LA. April 5, 2002.
Leibowitz S & Ludlow,L.H. Measuring longitudinal change in a multifaceted school reform initiative. Paper
presented at the American Educational Research Association annual meeting. New Orleans, LA. April 5, 2002.
Ludlow,L.H. & Leibowitz,S. Measuring readiness to change in the context of the family. Paper presented at the
American Educational Research Association annual meeting. New Orleans, LA. April 4, 2002.
Organizer, Fourth New England Objective Measurement Workshop. Two sessions of Rasch model applications by Boston College students. In conjunction with NEERO annual meeting. Portsmouth, NH. April 26, 2001.
Ludlow, L.H. & Mahalik, J.R. Congruence between a theoretical continuum of masculinity and the Rasch model.
Paper presented at the NEERO annual meeting. Portsmouth, NH. April 26, 2001.
Ludlow, L.H. & Alvarez-Salvat, R. Spillover in the academy: Marital stability and faculty evaluations. Paper

presented at the American Educational Research Association annual meeting. Seattle, WA. April 11, 2001.
Mahalik, J.R., Locke, B.D., Diemer, M., Ludlow, L.H., Scott, R.P.J., Gottfried, M., & Freitas, G. Development of the Conformity to Masculinity Norms Scale. Symposium - Masculine gender role conformity: Examining theory, research, and practice. Annual meeting of the American Psychological Association, Washington D.C., 2000.
Peck, A., Ludlow, L.H., & Scanlon, D. Student teachers and cooperating teachers as peer tutors in learning technology. Paper presented at the American Educational Research Association annual meeting. New Orleans, LA. April 29, 2000.
Ludlow, L.H. & Mahalik, J.R. Rasch structure of Conformity to Masculine Norms Inventory. Paper presented at the American Educational Research Association annual meeting. New Orleans, LA. April 27, 2000.
Ludlow, L.H. & Mahalik, J.R. Rasch structure of Masculine Identity Scales. Paper presented at the Tenth International Objective Measurement Workshops. New Orleans, LA. April 22, 2000.
Ludlow, L.H. & Mahalik, J.R. Rasch analysis of the Conformity to Masculine Norms Identity. Paper presented at the NEERO annual meeting. Portsmouth, NH. April 13, 2000.
Organizer, Third New England Objective Measurement Workshop: Two sessions of Rasch model applications by Boston College students. In conjunction with NEERO annual meeting. Portsmouth, NH. April 13, 2000.
Ludlow, L.H. & Mahalik, J.R. Initial construct validity of the Masculine Identity Continuum Inventory. Paper presented at the NEERO annual meeting. Portsmouth, NH. April 8, 1999.
Ludlow, L.H. Student drawings as course evaluations: What they see in statistics. Paper presented at the American
Educational Research Association annual meeting. Montreal, Canada. April 20, 1999. ERIC: ED 432 458.
Organizer, AERA Rasch Special Interest Group interactive symposium: Explaining Latent Trait Models to Non-Specialists. American Educational Research Association annual meeting. Montreal, Canada. April 21, 1999.
Organizer, Second New England Objective Measurement Workshop: Two sessions of Rasch model applications by Boston College students. In conjunction with NEERO annual meeting. Portsmouth, NH. April 8, 1999.
Ludlow,L.H. & Haley,S.M. New directions in pediatric rehabilitation measurement: The growing challenge. Paper presented at the Second International Outcomes Measurement Conference. Chicago, IL. May 16, 1998.
Organizer, (a) Symposium Session 3.1: “Classroom-level Issues in Assessment”, (b) Symposium Session 4.1: “Large-scale Issues in Assessment”. New England Educational Research Organization annual meeting. Portland, ME. April 30, 1998.
Organizer, First New England Objective Measurement Workshop: Two sessions of Rasch model applications by Boston College students. In conjunction with 30th NEERO annual meeting. Portland, ME. April 30, 1998.
Ludlow, L.H. The role of residuals in the evolution of scientific methods. Paper presented at the American Educational Research Association annual meeting. San Diego, CA, April 15, 1998.
Chair & Discussant (a) Symposium Session 4.1: "Creative Experimental Design and Statistical Analysis", (b) Symposium Session 6.4: "Rasch Model IRT Issues and Applications". New England Educational Research Organization annual meeting. Portsmouth, NH. May 1, 1997.

Ludlow, L.H. & O'Leary, M. Rasch analysis of omitted and not-reached test items: Some implications for large scale assessments. Paper presented at the New England Educational Research Association annual meeting. Portsmouth, NH. May 1, 1997.
Ludlow, L.H. & Lunz, M.E. The Job Responsibilities Scale: Invariance structure in a longitudinal study. Paper presented at the Fall meeting of the American Society of Clinical Pathologists. San Diego, CA, September 26, 1996.
Ludlow,L.H. & Bell,K.N. Psychometric characteristics of the Attitudes Towards Mathematics and its Teaching (ATMAT) Scale. New England Educational Research Association 1995 Distinguished Paper presented at
the American Educational Research Association annual meeting. NY, NY, April 9, 1996.
Ludlow,L.H. & Lunz,M.E. Comparative psychometric results on the Job Responsibilities Survey: Year one. Paper presented at the American Educational Research Association annual meeting. NY, NY, April 12, 1996.
Ludlow,L.H. & Bell,K.N. Psychometric characteristics of the Attitudes Towards Mathematics and its Teaching (ATMAT) Scale. Paper presented at the annual meeting of the New England Educational Research Organization. Portsmouth, NH, April 4, 1995.
Discussant, Paper session (6.5): "Assessments in Mathematics". New England Educational Research Organization annual meeting. Portsmouth, NH, May 4, 1995.
Ludlow,L.H. A comparison of classical and item response theory psychometric results on the ASCP Prospective Study 1st Year Survey. Paper presented at the Spring meeting of the American Society of Clinical Pathologists. Orlando, FL, April 21, 1995.
Ludlow,L.H. Instructor evaluation ratings: A longitudinal analysis. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA, April 19, 1995.
Ludlow,L.H. Omitted and not-reached item responses: How should they be scored? Paper presented at the eighth International Objective Measurement Workshop. Berkeley, CA, April 17, 1995.
Ludlow,L.H. A longitudinal analysis of student/instructor ratings. Paper presented at the Northeastern Educational Research Association annual meeting. Ellenville, NY, October 26, 1994.
Chair, Paper session (4.3): “Research Methodology on Education and Assessment”. New England Educational Research Organization annual meeting. Rockport, ME, April 21, 1994.
Chair, Paper session (3.3): “General Research on Assessment”. New England Educational Research Organization annual meeting. Rockport, ME, April 21, 1994.
Ludlow,L.H. & Haley,S.M. Measuring change in functional performance. Paper presented at the annual meeting of the New England Educational Research Organization. Rockport, ME, April 21, 1994.
Ludlow,L.H. & Haley,S.M. Effect of context in rating of functional performance. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA, April 4, 1994.
Gonzalez,E.J., Beaton,A.E. & Ludlow,L.H. Working with a national data base: The NAEP Primer. Paper presented at the annual meeting of the New England Educational Research Organization. Portsmouth, NH, April 30, 1993.
Ludlow,L.H. & Haley,S.M. Issues in the measurement of change in functional performance. Paper presented at the annual meeting of the New England Educational Research Organization. Portsmouth, NH, April 29, 1993.
Discussant, Paper session (25.44): “Partial Credit Models and Extensions”. American Educational Research Association annual meeting. Atlanta, GA, April 14, 1993.

Beaton,A.E., Gonzalez,E. & Ludlow,L.H. On developing a NAEP Primer. Paper presented at the annual meeting of the American Educational Research Association. Atlanta, GA, April 13, 1993.
Ludlow,L.H. & Haley,S.M. Displaying change in functional performance. Paper presented at the 7th International Objective Measurement Workshop. Emory University, Atlanta, GA, April 9, 1993.
Haley,S.M., Coster,W. & Ludlow,L.H. Discriminative and Evaluative Validity of the Pediatric Evaluation of Disability in Young Children with Brain Injury. Paper presented at the 11th Annual National Symposium of the National Head Injury Foundation. Boston, MA, December 11, 1992.
Ludlow,L.H. & Haley,S.M. Converting IRT logits into usable scaled scores. Paper presented at the annual meeting of the New England Educational Research Organization. Portsmouth, NH. May 7, 1992.
Discussant, Paper session: "Research Methodology". New England Educational Research Organization annual meeting. Portsmouth, NH. May 8,1992.
Ludlow,L.H. & Haley,S.M. The Pediatric Evaluation of Disability Inventory (PEDI). Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA. April 23, 1992.
Haley,S.M., Coster,W.J. & Ludlow,L.H. Rasch scaling of functional items in a normative sample: Development of the Pediatric Evaluation of Disability Inventory. Paper presented at the annual meeting of the American Academy of Physical Medicine and Rehabilitation. Washington, DC. October 27-31, 1991.
Discussant, Paper sessions: "Research Models and Analysis", and "Grouping Practices in Middle and Secondary Schools". New England Educational Research Organization annual meeting. Portsmouth, NH. April 24-26, 1991.
Ludlow,L.H. & Guida,F. Constructing a scale of academic anxiety. Paper presented at the American Educational Research Association annual meeting. Chicago, IL. April 1-6, 1991.
Franklin,J., Theall,M. & Ludlow,L. Grade inflation and student ratings: A closer look. Paper presented at the American Educational Research Association annual meeting. Chicago, IL, April 1-6, 1991.
Theall,M., Franklin,J. & Ludlow,L.H. Attributions or retributions: Student ratings and the perceived causes of performance. Paper presented at the American Educational Research Association annual meeting. Boston, MA. April 16, 1990.
Leibowitz,S. & Ludlow,L.H. Measuring ethical sensitivity in computer use. Paper presented at the American Educational Research Association annual meeting. Boston, MA, April 16, 1990.
Gable,R.K., Ludlow,L.H. & Wolf,M.B. Analyzing person characteristics and item structure to better understand Rasch model residuals. Paper presented at the American Educational Research Association annual meeting. Boston, MA. April 16, 1990.
Ludlow,L.H. & Guida,F.V. The Test Anxiety Scale for Children: A Measure of Academic Anxiety?. Paper presented at the New England Educational Research Organization annual meeting. Rockport, Maine, May 2-4, 1990.
Organizer, Symposium: “Measurement, Testing, and Statistical Analysis”. Paper presentations by Boston College, LSOE, ERME students. New England Educational Research Organization annual meeting. Rockport, ME. May 2-4, 1990.
Ludlow,L.H., Haley,S.M. & Gans,B.M. The Tufts Assessment of Motor Performance: An application of the Rasch rating scale model. Paper presented at the New England Educational Research Organization annual meeting. Portsmouth, New Hampshire. April 28, 1989.

Organizer, Symposium. “Item Response Theory Applications”. Paper presentations by Boston College, LSOE, ERME students. New England Educational Research Organization annual meeting. Portsmouth, NH. April 28, 1989.
Ludlow,L.H., Haley,S.M. & Gans,B.M. Scaling motor proficiency. Paper presented at the Fifth International Objective Measurement Workshop. Berkeley, CA. March 25, 1989.
Ludlow,L.H., Haley,S.M. & Gans,B.M. Methodological issues in the application of polytomous Rasch models to behavioral assessment: The Tufts Assessment of Motor Performance. Paper presented at the National Council on Measurement in Education annual meeting. San Francisco, CA. March 29, 1989.
Gable,R.K., Ludlow,L.H. & Wolf,M.B. The measurement of perceived school related stress using classical and Rasch latent trait models. Paper presented at the American Educational Research Association annual meeting. San Francisco, CA. March 28, 1989.
Haley,S.M., Ludlow,L.H. & Gans,B.M. Functional status assessment using the Rasch model: Development of the Tufts Assessment of Motor Performance. Paper presented at the American Physical Therapy Association annual meeting. February 2-5, 1989.
Chairperson and Discussant: "Psychometrics". Paper presentations by Boston College, LSOE, ERME students. New England Educational Research Organization annual meeting. Rockport, ME. April 27-29, 1988.
Organizer and Paper Presenter, Symposium: “Measurement and Testing: Innovations and Issues”. Paper presentations by Boston College, LSOE, ERME students. New England Educational Research Organization annual meeting. Rockport, ME. April 27-29, 1988.
Ludlow, L.H. A cross-validation of empirical classification strategies. Paper presented at the New England Educational Research Organization annual meeting. Stratton, VT. April 8-10, 1987.
Chairperson, Paper Session: "Statistics and Item Response". Paper presentations by Boston College, LSOE, ERME students. New England Educational Research Organization annual meeting. Stratton, VT. April 8-10, 1987.
Guida,F.V. & Ludlow,L.H. A cross-cultural study of test anxiety. Paper presented at the American Psychological Association annual meeting. Washington, D.C. August 1986.
Ludlow,L.H., Levy,S. & Howard,E. Mothers of failure to thrive infants: Identification of characteristics of perceived interpersonal familial relations. Paper presented at the New England Educational Research Organization annual meeting, Rockport, ME. May 1986.
Organizer, Symposium: “Innovative Applications of Multivariate Techniques”. Paper presentations by Boston College, LSOE, ERME students. New England Educational Research Organization annual meeting. Rockport, ME. May 1986.
Ludlow,L.H., Candee,D. & Murphy,C. On the construction of an ethical reasoning scale: An application of the Rasch model. Paper presented at the Eastern Educational Research Association annual meeting, Miami, FL. March 20-22, 1986.
Ludlow,L.H. A strategy for the graphical representation of Rasch model residuals. Paper presented at the New England Educational Research Organization annual meeting. Rockport, ME. April 23-25, 1985.
Chairperson. Paper Session: "Mathematics and Science". New England Educational Research Organization annual meeting. Rockport, ME. April 23-25, 1985.
Casey,M.B., Brabeck,M.M. & Ludlow,L.H. Familial handedness and its relation to spatial ability following strategy instruction. Paper presented at the American Educational Research Association annual meeting, Chicago, IL. March 29-April 4, 1985.
Ludlow,L.H. & Hwang,R. District vs. system level performance comparisons: Observation and proposals. Paper presented at the National Council on Measurement in Education annual meeting, Chicago, IL. April 1-3, 1985.

Ludlow,L.H. Diagnostic techniques in research synthesis. Paper presented at the American Educational Research Association annual meeting New Orleans, LA. April 1984, ERIC Document: ED 243-937.
Organizer, American Educational Research Association symposium. Title: Applications of Rasch models for ordered response category data. Montreal, Canada, April 1983.
Guida,F., Ludlow,L.H. & Wilson,M. Academic anxiety, time-on-task and achievement: A structural model. Paper presented at the American Educational Research Association annual meeting, Montreal, Canada. April 1983, ERIC document: ED 228-290.
Ludlow,L.H. Measuring change with the Rating Scale model. Paper presented at the American Educational Research Association annual meeting, Montreal, Canada. April 1983, ERIC document: ED 228-324.
Ludlow,L.H. The graphical representation of Rasch model residuals. Poster session presented at the American Educational Research Association annual meeting, Montreal, Canada. April 1983.
Hillocks, G. & Ludlow, L.H. A taxonomy of skills in reading and interpreting fiction. Paper presented at the American Educational Research Association annual meeting, New York, NY. March 22, 1982.
Ludlow, L.H. A residual is more than a chi-square: Patterns in attitudes toward blindness. Paper presented at the National Council on Measurement in Education annual meeting, New York, NY. March 20, 1982, ERIC document: ED 222-523.
Lecturer, workshop on "Measurement with Rasch Models". Sponsored by Instructional Objectives Exchange at the University of Chicago, Chicago, IL. June 15-19, 1981.
Ludlow, L.H. An exploratory investigation of Rasch model residuals. Paper presented at the American Educational Research Association annual meeting, Los Angeles, CA. April 1981, ERIC document: ED 206-652.
Lecturer, presession on "Rasch Measurement". Conducted by Benjamin D. Wright at the American Educational Research Association annual meeting, Boston, MA. April 4-6, 1980.
Ludlow,L.H. & Bond,N.A. Harness race handicapping: An empirical multivariate investigation. Paper presented at the Western Psychological Association annual meeting, San Francisco, CA. April 1978.
Ludlow,L.H. & Bond,N.A. A pari-mutuel betting system based on subjective win-loss utilities. Paper presented at the Conference on Decision Theory (hosted by Ward Edwards), Los Angeles, CA. February 16, 1978.
Behrman,B. & Ludlow,L.H. A multidimensional comparison of short and long-term shape imagery. Paper presented at the Western Psychological Association annual meeting, Northridge, CA, April 1976.

Invited addresses:
"Classical Test Theory, Item Response Theory, and Rasch Measurement: Transforming Theory Into Practice",
Boston College Graduate School of Social Work. Dr. Thanh Tran, April 27, 2009.
"Results from the TNE Surveys: Program effect analyses". Nanyang Technological Institute, National Institute of
Education. Boston College, April 9, 2008.
"Results from the TNE Surveys: Program effect analyses". The University of Auckland. Auckland, New Zealand.
March 11, 2008.
"Classical Test Theory, Item Response Theory, and Rasch Measurement: Transforming Theory Into Practice",
Boston College Connell School of Nursing. Dr. Anne Norris, October 30, 2007.

"Getting to pupil learning: Value-added assessment of teacher's impact". Visiting delegation from Nanyang
Technological University, National Institute of Education. Boston College, March 22, 2007.
"Making sense of course evaluations: A longitudinal approach". New England Institute of Technology. Faculty
Development Day. Warwick, RI. Dr. Richard L. Winstanley. January 27, 2007.
"Classical Test Theory, Item Response Theory, and Rasch Measurement", Boston College Connell School of
Nursing. Dr. Anne Norris, October 31, 2006.
"Value-added modeling applied to teacher preparation and pupil learning". Association of Jesuit Colleges and Universities Teacher Education Conference. Boston College. September 28, 2006.
"Advanced assessment and statistical modeling of Core data". Visiting Scholar. Nanyang Technological University, National Institute of Education, Singapore. July 3-July 14, 2006.
"Practical issues in the implementation of the Teachers for a New Era project". St. Mary Immaculate College, Limerick, Ireland; St. Patrick's College, Dublin, Ireland; and University of Ulster, Belfast, Northern Ireland, June 12-16, 2006.
"Survey development and measurement issues in longitudinal research". College of Education , The University of Auckland. Auckland, New Zealand. February 8, 2006.
"Online faculty evaluations versus paper administrations". College of Education , The University of Auckland. Dr. John Hattie. Auckland, New Zealand. February 7, 2006.
"Item Response Theory and Rasch Measurement". Boston College Connell School of Nursing. Dr. Anne Norris,
November 8, 2005.
"Faculty course evaluations: What is really going on in your class?" St. Mary's College (Marino), Dublin, Republic of Ireland. May 30, 2005.
"Standardized and non-standardized forms of faculty evaluations". Mary Immaculate College, Limerick, Republic of Ireland. May 27, 2005.
"Value added modeling: Background, implementation issues, research areas". Commonwealth Education Deans
Council. Boston College. April 1, 2005.
"Power analysis: You did what to estimate your sample size!?". Counseling and Development Psychology
Dissertation Seminar, Lynch School of Education, Boston College. March 30, 2005.
"Value-added modeling: Where did it come from, where is it going?". American Association of Colleges for Teacher
Education, Winter Institute. Cancun, Mexico. January 2, 2005.
"Item Response Theory and Rasch Measurement". Boston College Connell School of Nursing. Dr. Anne Norris,
October 12, 2004.
"Irish Primary Teachers' Lives: A Longitudinal Research Project". Colleges of Education Research Consortium, St.
Patrick's College of Education, Dublin, Ireland. May 26, 2004.
"Item Response Theory". Boston College Connell School of Nursing. PhD. Colloquia. February 24, 2004.
"Rasch measurement: Transforming theory into practice". Boston University, Sargent College of Health and
Rehabilitation Sciences. Dr. Alan Jette, December 5, 2003.
"Item Response Theory and Rasch Measurement". Boston College Connell School of Nursing. Dr. Anne Norris,

October 7, 2003.
"Faculty course evaluations: Is a picture worth a 1000 numbers?" Michigan State University, Center for the
Scholarship of Teaching. Dr. Suzanne Wilson, March 27, 2003.
"Rasch Measurement: So what, Who cares?". University of Massachusetts at Amherst: Dr. Ron Hambleton,
March 3, 2003.
"The Massachusetts Test for Educator Licensure: Where did we start, and where are we now?" Keynote Address.
Third Annual Conference on Massachusetts Teacher Test Preparation. Framingham State College, MA. Jan. 10, 2003.
"CTT and IRT For Health Care Measurement". Boston College Connell School of Nursing. Dr. Anne Norris,
October 29, 2002.
"Rasch Models for Measuring Readiness-to-change". University of Rhode Island Cancer Prevention Institute: Director O. Prochaska. URI: March 13, 2002.
"Teacher Tests: Who's Accountable? Who's Not? Who Decides?". AACTE annual meeting. NY,NY. February 25, 2002.
"MTEL Preparation: Who is doing what and how successful is it". Colleges of Worcester Consortium. Holy Cross, MA. January 11, 2002.
"Course Evaluations: Making Sense Out of the Ratings We Receive." Committee for Teaching and Learning, LSOE, Boston College. November 7, 2001.
"MTEL Results and Title II Report Card," Association of Independent Colleges and Universities in Massachusetts: President's Only Meeting. Wellesley College. March 15, 2001.
"On the Simulation and Analysis of Measurement Model Residuals". Educational Testing Service, Princeton, NJ. January 11, 1984. ERIC Document: ED 241-586.
"Patterns in Attitudes Toward Blindness". Rehabilitation Research & Development Research Series. USVA Hospital, Hines, IL. February 26, 1982.

Reviewer:

a) Books

Pearson, Merrill, Prentice Hall. *Educational Research: Competencies for Analysis and Applications* (8th ed.) Gay, L.R., Mills, G. & Airasian, P. 2006
Prentice Hall. *Understanding Research: Becoming a Competent and Critical Consumer*. Jones and Kottler. 2005.
Allyn & Bacon. *Using Multivariate Statistics*. (4th ed.). Tabachnick & Fidell. 2003.
Prentice Hall. *Research Methods in Education: How To Find Stuff You Need To Know*. Blind review, 2002.
Houghton Mifflin. *Applied Statistics for the Behavioral Sciences* (5th ed.) Hinkle, Wiersma & Jurs. 2001.
John Wiley & Sons. *Elements of Statistical Reasoning*. Minium,E., Clarke,R.C., & Coladarci,T. 2001.
Prentice Hall. *Quantitative Methods for the Social Sciences*. 2000.
Lawrence Erlbaum Associates. *Prospectus for Rasch Measurement: Theory, Models, and Applications,* 1999.
Prentice Hall. *Multivariate Data Analysis* (5th ed.) Hair, Anderson, Tatham & Black, 1999.
Lawrence Erlbaum Associates. *Prospectus for Design and Analysis of Questionnaires and Rating Scales*. Everitt,R. & Wolfe,E., 1999.
Addison Wesley Longman. *Test Theory: A Unified Treatment.* McDonald, R., 1997.

Longman Publishing. *Essentials for Statistical Analysis in the Behavioral Sciences* (2nd. ed.) Glasnapp,D.R. & Poggio,J.P., 1992.
McGraw Hill. *Psychometric Theory* (3rd. ed.) Nunnally,J. & Bernstein,I., 1991.
State University of New York Press.

b) Journals

American Journal of Economics and Sociology
Archives of Physical Medicine and Rehabilitation
Consciousness and Cognition
Developmental Psychology
Educational Measurement: Issues and Practice
Educational Policy
Educational and Psychological Measurement
Educational Research Quarterly
Field Methods
Journal of Applied Measurement
Journal of Higher Education
Journal of Outcome Measurement
Journal of Statistics Education
Journal of Teacher Education
Management Science
Medical Care
Psychology & Marketing
Psychological Methods
Psychological Reports: Perceptual and Motor Skills
Research Quarterly for Exercise and Sport
Structural Equation Modeling
Applied Measurement in Education
Journal of Technology, Learning, and Assessment

c) Conferences

National Council on Measurement in Education
American Educational Research Association
American Educational Research Association-SIG: Rasch Measurement
New England Educational Research Organization
International Council on Education for Teaching

d) Other

University of Miami, Educational and Psychological Studies, Research, Measurement and Evaluation Program. Program Review. Co-Directors: Dr. Guerda Nicholas and Dr. Randall Penfield. October 7-9, 2009.
Outside reader: Seng, Lee See (2008). *The Construction and Validation of a School Quality Management Scale for Quality Improvement in School Management*. University of Malaya.
"What more can schools do? Identifying school building level characteristics and policies that affect student progress in mathematics and reading classrooms". Teacher Quality Partnership: Seidel K, Franco S, Wang L, Stock R & Newman I. Invited reviewer (2006).
Promotion and Tenure: Rutgers University (2003), Boston University (2003), University of Virginia (2003), Tel Aviv University (2004), George Mason University (2006), University of Kentucky (2007), National Institute of Education—Singapore (2007), Emerson College (2008), National Institute of Education—Singapore (2009)
The Retirement Research Foundation. RFP submitted for funding. 1991.
Best Paper Competition, New England Educational Research Organization.
Outside reader: Kalinowski, A. (1992). *The Latent Structure of Pain Intensity*. Clark University.

External Grants:

Co-Principal Investigator (with Marilyn Cochran-Smith). *Teacher development and teacher retention: Unraveling complex issues*. The Ford Foundation. Academy of Educational Development Project No. 3086. $349,996. July 14, 2008.

Boston College Grants:
Teaching, Advising, and Mentoring Grant. Summer 2005.*Training Better Consumers and Producers of Research.*
Teaching, Advising, and Mentoring Grant. Summer 2004. *Classroom Assessment Integration Project*
Teaching, Advising, and Mentoring Grant. Summer 2002. *Instructor Evaluation Ratings: How to Interpret and Use Them Effectively.*
Research Incentive Grant: Summer 1998. *Multivariate Statistics: Creative Design and Analysis*.
Research Expense Grant: Spring 2002, Fall 1998, Fall 1986, Spring 1986, Fall 1985, Summer 1985, Spring 1985.
Summer Research Grant: 1984. *Residual analysis as a diagnostic technique for revealing surprising responses on a test*.

Professional Affiliations:
American Educational Research Association and Special Interest Groups:
Rasch Measurement, Educational Statisticians, Professors of Educational Research, Faculty Evaluation
American Statistical Association
Classification Society of North America
National Council on Measurement in Education
Northeastern Educational Research Association
Western Psychological Association
Institute for Objective Measurement
Eastern Educational Research Association
New England Educational Research Organization
Psychometric Society

University Service:
Dissertations Completed as Chair or Reader: 65
Search committee, Psychology Arts & Science, statistics professor (open), 2007-2008.
Provost Advisory Committee, 2006-present.
Search committee, LSOE/ERME, statistics/measurement assistant professor, 2005-2006.
Search committee, LSOE/ERME, Boisi Chair, 2005-2006.
University Council on Teaching: Teaching, Advising and Mentoring Session Leader: AnDover Retreat. Natick, MA.
Dec 2-3, 2005
Instructional Technology Services (Elizabeth Clark), *Teaching with New Media Award* committee. 2005.
Search committee, LSOE, Dean, 2003-2004.
Subcommittee on Atrium Classroom Design. LSOE, 2003-2004.
University Council on Teaching, 2002-2006.
- Teaching, Advising, and Mentoring Grant review subcommittee (Chair 2006).
- Online Faculty Evaluations subcommittee.

Chair: Educational Research, Measurement, and Evaluation Department. LSOE, 2001-present.
Committee on Teaching and Learning. LSOE, 2001-2004.
Co-coordinator, Brown Bag Research Seminar for Departments of Teacher Education, and Educational Research,
Measurement and Evaluation. LSOE, 2000-present.
Statistics and Methodology Committee. 2000-2002.

Educational Policy Council, Chair. LSOE, 2000-2001.
Faculty Statistical Coordinator for Graduate Statistical Assistant program. LSOE and School of Nursing. 2000-
2002.
Faculty Fellowship Committee. LSOE, 2000-2001.
Certification and Regulations Committee. LSOE, 2000-2002.
Dissertation Awards Committee. LSOE, Spring 2000.
Massachusetts Test of Educator Licensure Ad Hoc Task Force, Convener. LSOE, 1999-2002.
Teacher Education Council. LSOE, 1999-2004.
Educational Policy Council, Chair: Subcommittee-A. LSOE, 1999-2000.
Educational Policy Council, Chair: Governance Subcommittee. LSOE, Spring 1999.
Program Director: Educational Research, Measurement, and Evaluation Program. SOE, 1995-1997.
Educational Policy Committee-Policy Subcommittee (invited member). SOE, 1996-1997.
Faculty Fellowship Committee. SOE, 1994-1995.
Chair/advisor/reviewer, SOE Graduate Student Forum. 1993-1996, 2002-present.
Search Committee, Associate Dean for Graduate Programs. School of Nursing, 1993-1994.
Faculty Fellowship Committee, Chair. SOE, 1991-1992.
Tenure and Promotion Committee. SOE, 1990-1991.
Steering Committee, Andover Workshop for New Faculty. 1990.
Facilities Planning Committee. SOE (ad hoc), 1989.
Academic Vice President Computer Advisory Committee. 1987-1989.
Media Center Advisory Committee. SOE, 1986-1987.
Compensation Committee Health Issues Task Force. 1986-1987.
Student Affairs Goals for the Nineties Task Force on Technology and Student Life. 1986-1987.
Faculty Marshall for Graduate School of Arts and Sciences Commencement Ceremony. 1986.
Search Committee for Burns Library Conservationist-Archivist. 1986.
Nomination Committee. SOE, 1986-1987.
United Way Solicitor. SOE, 1986,1992.
Search Committee for University Librarian. 1986.
Search Committee for Assistant University Librarian for Access Services. 1985-1986.
Faculty Representative, Interview of Finalists for Baptist Librarian Position. 1985.
Faculty Advisor, SOE Freshman Advisement Program, 1984-1986.
Faculty Supervisor, SOE Student Teaching. Fall 1984.
Committee on Computer Courses/Resources. SOE, 1984-1986.
Educational Policy Committee. SOE, 1984-1986.
University Academic Council. 1984-1985.
University Library Committee. 1984-1986.

Professional Service

Westat, Expert Advisory Panel. "Evaluation of the Regional Educational Laboratories". PI: Babette Gutmann. 2009-present.
Pontifica Universidad Catolica De Valparaiso, "Design and Implementation of University Faculty Evaluation System." PI: Gloria Contreras Perez. Valparaiso, Chile. 2009-present.
University of Diego Portales. "Design of a Monitoring and Evaluation System for Learning Methods of Undergraduate Students of the Faculty of Education." PI: Horacio Walker. Santiago, Chile. 2009-present.

Brandeis University, The Heller School for Social Policy and Management. “Assessing a Participant Directed Service System for Low Income Children with ASD”. PI: Dr. Marji Warfield. 2009-present.
University of Massachusetts Donahue Institute, Technical Assistance to the National Cancer Institute’s Cancer Information Service. PI: Dr. Sue Liebowitz, 2009.
Boston College, Connell School of Nursing. “Binge drinking and eating disorder detection”. PI: Susan Kelly-Weeder, 2008.
Boston College, Department of Biology. “Evaluation of Advanced Biology Module”. PI: Clare O’Connor, 2007.
Hudson Public Schools, Hudson, MA. “Teaching American History”. PI: Oscar Loureiro/Todd Wallingford. 2007-present.
Boston University, Programs in Deaf Studies. “The Development of a Signed Language Measurement Instrument”. PI: Robert J. Hoffmeister, PhD. 2007-present.
California State University, Panel of External Advisors, CSU Evaluation Plan. Evaluation Director: David Wright, 2006-present.
Boston College, Connell School of Nursing. “Cognitive Recovery from Mild Head Injury”. PI: Sr. Callista Roy, 2006-2007.
National Science Foundation; Education and Human Resources Directorate; Research, Evaluation and Communication Division; Research and Evaluation on Education in Science and Engineering (REESE) proposal review panel. June, 2006.
Boston College, LSOE. “Treatment Development for Ethnic Minority Adolescents”. PI: Guerda Nicolas. 2006.
Boston College, LSOE. “Professional Development in ‘Accountable Talk’: Quality Book Discussions as Means of Improving Student Reading Comprehension”. Co-PI’s Curt Dudley-Marling, Sara Michaels, 2006.
Boston College, LSOE. “Teacher Autonomy Scale”. Director Curt Dudley-Marling. 2006.
McLean Hospital, Department of Mental Health Services Evaluation. “BASIS-24 Sensitivity Model”. Director: Alex Speredelozzi. Belmont, MA. 2005-2007.
Boston College, LSOE. “Boston College Evaluative Research Project on Children’s Hospital Social and Academic Discourse Program”. PI: David Scanlon. 2005-2007.
Franklin W. Olin College of Engineering, Center for the Advancement of Engineering Education. “Academic Pathways Study”. Director: Ozgur Eris. Needham, MA. 2005-present.
National Science Foundation; Education and Human Resources Directorate; Research, Evaluation and Communication Division: Faculty Early Career Development (CAREER) proposal review panel. October 27-28, 2005.
University of Massachusetts, Amherst, School of Education, Education Policy, Research and Administration: “Academic Quality Assessment and Development: Departmental Self-Study”. Chair: Joseph Berger, 2005.
Boston College, LSOE, Carnegie Foundation Project: “Teachers for a New Era”. Member of the Leadership Team and the Evidence Team. Project Manager Gilda Morelli. 2004-present.
St. Patrick’s College, Dublin, Republic of Ireland, Colleges of Education Research Consortium. “Teacher’s Lives”. Co-directors Mark Morgan and Michael O’Leary. 2004-present.
American Education Research Association, Professional Development and Training Committee. 2004-2007.
EDC/Center for Children and Technology, Advisory Board. “Evaluation of outcomes for Intel Teach to the Future”. Director: Katie McMillan Culp. Newton, MA. 2003-2007.
Boston College, LSOE. “Culturally Meaningful Adventure Stories”. PI: Beth Casey. 2003-present.

Franciscan Children's Hospital & Rehabilitation Center. Various projects around "Physical Therapy Prognosis for Recovery". Director: Helene Dumas. Brighton, MA. 2003-present.
Boston College, LSOE. "Orphans of Eritrea". Co-director: Irving Hurwitz. 2003-2006.
EDC/Center for Online Professional Education, Advisory Board: "Optimizing Online Professional Development", Co-PI: Mike Russell. Newton, MA. 2003-2007.
U.S. Department of Education, Office of Indian Education and the National Center for Education Statistics Technical Review Panel: "National Indian Education Survey". Chair: Henry Braun. 2003-present.
Boston College, LSOE. "Study of Generations of American Rhoades Scholars". Co-directors Ted Youn and Karen Arnold. 2002-2004.
American Association of College Teachers of Education Institutional Representative (Boston College—Lynch School of Education). 2001-present.
National Education Association, The Massachusetts Center for Teaching and Learning at University of Massachusetts, Boston, "Massachusetts Teacher Preparation and Induction Study". Director: Karen O'Connor. 2001-2004.
World Health Organization, "Statistical strategies for cross-population comparability". Harvard, MA. 2001.
Massachusetts Department of Education, Title II Advisory Committee. Chair: Carol Gilbert. Boston, MA. 2000-2004..
Association of Independent Colleges and Universities in Massachusetts (AICUM). President: Clare Cotton. Boston, MA. 2000-2004.
Rehabilitation and Research Training Center on Measuring Rehabilitation Outcomes (RRTC). Boston University, Sargent College. PI: Stephen Haley. 2000-present.
Guliano, et al. vs. The New York State Department of Education and the New York City Board of Education. 96 Civ. 8414 (CBM). 2000-2004.
University of Massachusetts, Donahue Institute for Governmental Services. "Family Trusts Project: Brightside for Children and Families". Director: Pamela Lamlein, 2000-2002.
Massachusetts Public School Districts: Malden, Revere, Fall River, Natick, Montachusetts, Lunenburg, Quincy, Scituate, Chatham. Analysis of student performance on the Spring 1998 Comprehensive Assessment System (MCAS) tests. 1998-1999.
Wheelock College. "Wheelock College MECT Results: 1998-1999". Laurie Crumpacker, Undergraduate Dean Boston, MA. 1999. "Wheelock College MECT Results: 1998-2000". Vice-President Finance: Stanley Rumbaugh. 2000.
Salem State College. "Salem State College MECT Results: 1999". Professor: Jamie Wurzel. Salem, MA. 1999.
Lesley College. "Lesley College MECT Results: 1998". Provost: Janet Schulte. Cambridge, MA. 1998-1999.
Northeast Rehabilitation Health Network. "Computation of Recipient Sample Extrapolation"--expert witness. John Prochilo, Salem, NH. 1998.
NAACP. "Sarah P. Wessman v. Boston School Committee, et al." Civil Action No. 97 CV 11923 JLT. Boston, MA. 1998.
National Institute of Disability Management and Research. (a) "Disability Management: Return-to-Work Coordinator Occupational Standards Survey", (b) "Homeless Job Retention and Sustained Employment Study". Director: Norm Hursh, 1998-present.
Center for Psychiatric Rehabilitation, Boston University. "WFD Survey" and other projects. E. Sally Rogers, 1998-2004.
Center for Rehabilitation Effectiveness (CRE), Boston University, Sargent College. PI: Stephen Haley. 1998-2000.
ASTRA Hassle (through CRE)."Quality of Life Indicators". Ingela Wiklund, 1998.

Carroll Center for the Blind. “Staff salary and benefits survey”. President: Rachel Rosenbaum. Newton, MA, 1997-1999.
Commission on Institutions of Higher Education, New England Association of Schools and Colleges. “Institutional Effectiveness”. Bedford, MA, 1996.
New England Medical Center: Department of Pediatrics, Floating Hospital for Infants and Children. “Outcome in Developmentally Delayed Internationally Adopted Children”. Boston, MA, 1993.
University of Massachusetts Medical Center: Department of Pediatrics. (1) “Social Influences on Development in Toddlers with Asthma”, (2) “Early Intervention Collaborative Study: Age 10 Follow-Up”. Penny Hauser-Cram. Boston, MA, 1992-1998.
Center for the Study of Testing, Evaluation and Educational Policy (CSTEEP): Third International
Mathematics and Science Study (TIMSS.) Boston College, Albert Beaton, Ph.D. 1994-1996.
Fitchburg State College: Special Education Department, Fitchburg, MA, 1993.
Education Development Center, Inc., Newton, MA, 1990.
Cox Educational Services, Cambridge, MA, 1990.
Phillips Exeter Academy, N H. “Student Affairs Longitudinal Study”. 1989-1993.
Tufts University School of Medicine, and New England Medical Center: Department of Rehabilitation Medicine. 1) Tufts Assessment of Motor Performance, and 2) Pediatric Evaluation of Disability Inventory. PI: Stephen Haley. Boston, MA., 1987-1992.
S. Gruner vs. Bolton School Committee, 1987.
Georgia Association of Educators, (G.A.E., et al., vs. Georgia State Board of Education, et al.), Civil Action No. C86-2234A, 1987.
McSherry Associates, Inc., Southborough, MA, 1985.
Northeastern University, Office of Learning Resources. “Faculty Evaluation Systems”. Director: Michael Theall, 1985-1989.
Texas State Teachers Association, 1986.
Arkansas Education Association, (Arkansas Education Association Teacher Testing Law), 1985.
Alabama Education Association, (Allen, et al., vs. Alabama State Board of Education, et al.), Civil Action No. 81-697-N, 1984-1986.
Connell School of Nursing, Boston College, 1984-present.
The Center for the Study of Testing, Evaluation, and Educational Policy. Boston College, 1984-present.
Ball Foundation, Glen Ellyn, IL, 1982.
DIAL, Inc.. “Predictors of Early Childhood Development”. DeKalb, IL, 1981.
Department of Immigration, Quebec, Canada, 1980.
California State Department of Education. “Applicability of the Rasch Model for Achievement Tests”. Director: Tej Pandey. Sacramento, CA, 1979.
California Black Commission on Alcoholism, Sacramento, CA, 1977.
Golden Bear Raceway. “Prediction System for Harness Races”. Sacramento, CA, 1976.

Board Memberships:
Associate Editor, *Journal of Teacher Education*. 2002-2006.
Editorial Board, *Journal of Applied Measurement*. 2000-present.
Board of Cooperating Editors, *Educational and Psychological Measurement.* 1989-2006.
Board of Registry, American Society of Clinical Pathologists, Research and Development Committee, 1989-1996.
Board of Directors, Mead Measurement Resources, Inc., Minneapolis, MN, 1985-present.

Other:
Hudson Public Schools, Hudson, MA. "Teaching American History: Year Two Annual Report". PI: Todd Wallingford. 2009.
Tenure and promotion review for assistant professor. University of Kentucky. Fall 2007.
Tenure and promotion review for associate professor. Nanyang Institute of technology, Singapore. Summer 2007.
Tenure and promotion review for full professor. George Mason University, Virginia. Spring 2007.
Co-evaluator (with David Scanlon). Project NewBCTech: Newton MA Public Schools, Co-Director Alec Peck, Boston College, 1999.
Evaluator. Boston College Collaborative: College Bound, 1989-1992.
Past President, Immediate Past President, President, Vice-president and Program Chair, Regional Director. New England Educational Research Organization, 1986-1992.
Project Coordinator (local). Educational Testing Service, GRE General Test, 1986-1987.
Evaluator (Chair). Chenery Middle School Computer Education Program, Belmont, MA, 1986.

Honors:
Distinguished Service Award, New England Educational Research Organization. April 22, 2004.
Nominee, Boston College Distinguished Teaching Award. May 2002.
Feature article, "The Spillover Effect", Boston College Chronicle, January 7, 2002.
Nominee, Boston College Distinguished Teaching Award. May 2000.
Ludlow, L.H. & Bell, K. Co-recipients of the 1995 New England Educational Research Organization Best Paper Award: "Psychometric Characteristics of the Attitudes Towards Mathematics Scale."
Runner-up, 1994 New England Educational Research Organization Best Paper Award: "Effect of Context in Rating of Functional Performance".
Faculty Fellowship, Fall 1994. Title: "Graphical Comparisons of Three-Dimensional Data Structures".
Feature article, "The Statistical Gunslinger", Boston College Biweekly, May 7, 1992.
Runner-up, 1990 New England Educational Research Organization Best Paper Award: "The Test Anxiety Scale for Children: A Measure of Academic Anxiety?".
Runner-up, 1989 New England Educational Research Organization Best Paper Award: "The Tufts Assessment of Motor Performance: An Application of the Rasch Rating Scale Model".
Ludlow, L.H. "Analyzing the Tufts Assessment of Motor Performance with the Partial Credit Model". Invited paper for the Midwest Objective Measurement Seminar, December 2, 1988.
Runner-up, 1988 New England Educational Research Organization Best Paper Award: "The Family Map: A Graphical Representation of Family Systems Theory".
Ludlow, L.H. "On the Simulation and Analysis of Measurement Model Residuals". Seminar conducted for Educational Testing Service, Princeton, NJ, January 11, 1984. ERIC Document: ED 241-586.
Ludlow, L.H. "Patterns in Attitudes Toward Blindness". Invited address for the Rehabilitation Research & Development Research Series USVA Hospital, Hines, IL, February 26, 1982.

Professional Development:
Hosmer, D. Applied logistic regression. Natick, MA, October 25, 2008.
Goldstein, H. Learning environment for multilevel methodology and applications. Coventry, England (British Educational Research Association). September 7, 2006.

Beretvas,T. & Roberts,K. Multilevel and hierarchical linear modeling. Montreal, Canada (AERA). April 11, 2005.
Habing,B. & Froelich,A. Multidimensional item response theory. Montreal, Canada (AERA). April 10, 2005.
University of Massachusetts, Amherst. (Ron Hambleton). One semester course in Item Response Theory. September-December, 2004.
University of Maryland (Directed by Robert Lissitz). Value-Added Modeling: Issues with Theory and Application. College Park, MD. October 21-22, 2004.
National Institute of Health. Advances in Health Outcomes Measurement. Bethesda, MD. June 23-25, 2004.
Mueller, R.O. & Hancock, G.R. Structural Equation Modeling. San Diego, CA (AERA). April 11-12, 2004.
Mislevy, B. Graphical Techniques for Educational Assessment. New Orleans, LA (AERA). April 1, 2002.
Conley, S. Conference on Massachusetts Teacher Test Preparation. Framingham State College, MA. January 7, 2002.
Prochaska, J.O. & Levesque, D. Transtheoretical Model of Behavioral Change: Implications for Behavioral Health Practice. Brightside Series. Northampton, MA. December 7, 2001.
Thompson, B. & Huberty, C. Multivariate Statistics II: Canonical Correlation Analysis. Seattle, WA. (AERA). April 24, 2001.
Sayer, A. An Introduction to Linear Multilevel Models. Boston College, November 10-13, 2000.
Huynh-Huynh. Robust Regression. New Orleans, LA (AERA). April 25, 2000.
Thompson B. & Huberty, C. Multivariate Statistics II: Predictive Discriminant Analysis. Montreal, Canada (AERA). April 22, 1999.
Bentler, P. & Schumacker, R. A Second Course in Structural Equation Modeling: Theory and EQS 6.0 Implementation. Montreal, Canada (AERA). April 19, 1999.
McNeil, K. & Newman, I. Advanced General Linear Model: Policy Decisions. San Diego (AERA), April 13, 1998.
Brennan,R. Generalizability Theory and Applications. San Francisco (AERA), March 27, 1989.
Masters,G.N. & Wilson,M. Understanding and Using Partial Credit: An IRT Model for Ordered Response Categories. New Orleans (AERA), April 6, 1988.
Bock,R.D., Mislevy,R. & Thissen,D. Applications of Item Response Theory. Chicago: Palmer House, March 29-30, 1985.
Joreskog,K.G. & Sorbom,D. Factor Analysis, Covariance Structure Analysis, Path Analysis and Structural Equation Models. Chicago: University of Chicago, September 8-12, 1980.

SELECTED PUBLICATIONS FROM 2000 TO PRESENT PROFESSOR LARRY LUDLOW, BOSTON COLLEGE
OCTOBER 10, 2009

Morgan M, **Ludlow LH**, Kitching K, O'Leary M & Clarke A. (2009). What makes teachers tick? Sustaining events in new teachers lives. *British Educational Research Journal*, iFirst Article, 1-18. http://dx.doi.org/10.1080/01411920902780972

Vasilyeva M, **Ludlow LH**, Casey B & St. Onge C (2009). Examination of the psychometric properties of the Measurement Skills Assessment (MeSA). *Educational and Psychological Measurement*, 69, 106-130.

Enterline S. Cochran-Smith M, **Ludlow LH** & Mitescu E. (2008). Learning to teach for social justice: Measuring change in the beliefs of teacher candidates. *The New Educator*, 4, 267-290.

Ludlow LH, Pedulla J, Enterline SE, Cochran-Smith M, Loftus F, Salomon-Fernandez Y & Mitescu E. (2008). From students to teachers: Using surveys to build a culture of evidence and inquiry. *European Journal of Teacher Education*, 31(4), 319-337.

Ludlow LH, Enterline S & Cochran-Smith M (2008) Learning to Teach for Social Justice—Beliefs scale: An application of Rasch measurement principles. *Measurement and Evaluation in Counseling and Development*, 20, 194-214.

Burns S. & **Ludlow LH**. (2006). Understanding student evaluations of teaching quality: The unique contributions of class attendance. *Journal of Personnel Evaluation in Education*. http://dx.doi.org/10.1007/s11092-006-9002-7

Maloy RW, Pine GJ, Seidman I & **Ludlow LH**. (2006). Arriving on a fast Track: Perceptions of teachers from alternative, campus-based and PDS teacher preparation programs about their first four years in the classroom. *The Teacher Educator,42*, 106-121.

Ludlow LH (2005). A longitudinal approach to understanding course evaluations. *Practical Assessment Research and Evaluation*, http://pareonline.net/pdf/v10n1.pdf.

Mahalik JR, Locke B, **Ludlow LH**, Diemer M, Scott PJ, Gottfried M, & Freitas G. (2003). Development of the Conformity to Masculine Norms Inventory. *Psychology of Men and Masculinity, 4*, 3-25.

Ludlow LH, Shirley D & Rosca C. (2002). The case that won't go away: Besieged institutions and the Massachusetts teacher tests. *Educational Policy Analysis Archives*, 10(50). http://epaa.asu.edu/epaa/v10n50.html.

Ludlow LH (2002). Rethinking practice: Using faculty evaluations to teach statistics *Journal of Statistics Education*, 10(3). www.amstat.org/publications/jse/v10n3/ludlow.html.

Ludlow LH & Alvarez-Salvat R. (2001). Spillover in the academy: Marriage stability and faculty evaluations. *Journal of Personnel Evaluation in Education, 15:2*, 111-119.

Ludlow LH (2001). Teacher test accountability: From Alabama to Massachusetts. *Education Policy Analysis Archives, 9* (6). http://epaa.asu.edu/epaa/v9n6.html.

Ludlow LH & Haley SM. (2000). New directions in pediatric rehabilitation measurement: The growing challenge. *Journal of Outcome Measurement, 4*, 482-490.

Computer Software:

Masters,G.N., Wright,B.D. & Ludlow,L.H. CREDIT: A Rasch program for Partial Credit Data. MESA Psychometric Laboratory, University of Chicago, 1983.

Masters,G.N., Wright,B.D. & Ludlow,L.H. SCALE: A Rasch program for Rating Scale Data. MESA Psychometric Laboratory, University of Chicago, 1981/2000.

Wright,B.D., Mead,R.J. & Ludlow,L.H. KIDMAP: A graphical report of individual performance. MESA Psychometric Laboratory, University of Chicago, 1980.

ARTICLES

Learning to Teach for Social Justice–Beliefs Scale: An Application of Rasch Measurement Principles

Larry H. Ludlow, Sarah E. Enterline, and Marilyn Cochran-Smith

The authors illustrate how a Rasch model can guide the development of a new affective measurement instrument—the Learning to Teach for Social Justice–Beliefs scale. The results provide strong evidence of a meaningful continuum of attitudes about teaching for social justice ranging from those easier to endorse to those more difficult to endorse.

The idea of preparing teachers to teach for social justice is prevalent in a loosely related collection of teacher preparation programs, partnerships, grassroots teacher and community groups, and other initiatives in the United States and elsewhere. Despite national attention, however, there is considerable variation in meanings of the phrase *teacher education for social justice*, and, in general, this has not been a well-theorized term (North, 2006). Very generally speaking, however, most definitions (e.g., Adams, Bell, & Griffin, 1997; Cochran-Smith, 1999, 2004; Michelli & Keiser, 2005; Oakes & Lipton, 1999; Sleeter, 1996; Villegas & Lucas, 2002; Zeichner, 2003) have in common explicit recognition of the marked disparities in educational opportunities, resources, achievement, and long-term outcomes between minority and low-income pupil groups and their White, middle-class peers. This is coupled with the position that teachers have the potential to be both educators and activists committed to the democratic ideal and to reducing the inequities in American society. Teacher education for social justice, then, is teacher preparation deliberately designed to provide the social, intellectual, and organizational contexts to foster teaching for social justice in schools accommodating students in kindergarten through 12th grade (K–12).

Teaching for social justice in K–12 schools has as its primary consideration promoting pupils' learning (academic, social, emotional, and civic) and enhancing pupils' life chances, including challenging the structures, curriculum, labels, and school arrangements that limit or inhibit life chances. This agenda builds on a wide-ranging body of scholarship, practice, and grassroots efforts, including multicultural theory and pedagogy; research on effective practices in diverse classrooms; critical analyses of education and society; research on culture, language, and identity; organization at the grassroots community level to change schools; and theories related to the role of education in democratic societies. Teaching for social justice builds on and requires knowledge (i.e., knowledge of content, pedagogy, learners, cultures, schooling, communities, as well as knowledge of self), interpretive frameworks (i.e., ways of understanding and acting on the events and processes of schooling based on the integration of knowledge with beliefs, values, ethics, moral commitments, and attitudes), and practices (including subject-specific pedagogies and strategies for supporting the learning process of English language learners [ELL], pupils with special needs, and pupils from a range of socioeconomic backgrounds). Teaching for social justice also involves teacher commitment to being part of larger social movements by working as advocates and activists for their pupils.

In this article, we assume that teaching for social justice is a legitimate and measurable outcome of teacher education. The purpose of this article is to present evidence of the extent to which this assumption has been met. Specifically, we present the psychometric characteristics of the Learning to Teach for Social Justice–Beliefs scale. This includes the operational definition of the construct, the item development and pilot testing procedures, item analysis results of both classical test theory (CTT) and item response theory (IRT; Rasch) procedures, and evidence of discriminant validity.

Larry H. Ludlow and Sarah E. Enterline, Department of Educational Research, Measurement, and Evaluation; Marilyn Cochran-Smith, Department of Teacher Education, Special Education, Curriculum and Instruction, all at Boston College. Correspondence concerning this article should be addressed to Larry H. Ludlow, Department of Educational Research, Measurement, and Evaluation, Lynch School of Education, Boston College, 140 Commonwealth Avenue, Campion Hall 336C, Chestnut Hill, MA 02467 (e-mail: ludlow@bc.edu).

METHOD

Variable Definition

Learning to teach for social justice is conceptualized in terms of six core components: teachers' knowledge, skill, and interpretive frameworks; teachers' beliefs, attitudes, and values; classroom practice and pedagogy; community participation; teachers' learning in inquiry communities; and promoting pupils' academic, social–emotional, and civic learning. In developing the scale to measure this variable, the Survey Team of the Boston College–Teachers for a New Era (BC–TNE) project (Ludlow et al., 2007) came to a common understanding that any variable, even one as complex as learning to teach for social justice, may be conceptualized as a continuum along which people differ. In the academic setting addressed in this study, this means that teachers would differ in the extent to which they understand, accept, and are prepared to teach in ways consistent with the social justice principles just described.

Measurement Models

The current field of psychometrics relies on two primary measurement models, CTT and IRT. CTT is based on the simple yet powerful concept that an individual's *observed score*, defined as the *total score* on some measurement instrument, is made up of two unobservable, theoretical components: a *true score* and an *error score* (X = True + Error; Gulliksen, 1950; Lord & Novick, 1968; Spearman, 1904). Although the true score is never actually known, it is possible to generate estimates of the extent to which measurement error affected the observed score, thereby reducing the extent to which the true score is captured by the observed score. Hence, great effort is expended to estimate and reduce measurement error because the more measurement error can be reduced, the more confidence there is that the observed score accurately represents the true score.

The basic psychometric tools of CTT include factor analysis, reliability analyses (e.g., test–retest, internal consistency, and inter- and intrarater reliability), and validity analyses (e.g., content, variable, discriminant, divergent, predictive, consequential).

The limitations of CTT (e.g., the ability estimate of a person is dependent on the difficulty of the items, the standard error applies equally to all ability levels, item discrimination can be too high) have been widely recognized (Brennan, 2001; Hattie, Jaeger, & Bond, 1999; Masters, 1988; Traub, 1997; Wainer, 1986) and have led many investigators to use the principles underlying IRT. The IRT models differ, however, in how the probability of a specific response to a specific item is estimated (Hambleton, Swaminathan, & Rogers, 1991; Lord & Novick, 1968).

In general, IRT models are differentiated by the number of parameters associated with various item-specific characteristics. These characteristics are generally referred to as *item difficulty*, *item discrimination*, and *item pseudoguessing parameters*. IRT models that take each of these characteristics into account are referred to as one-parameter (1-PL, or the Rasch model), two-parameter (2-PL), and three-parameter (3-PL) logistic models, respectively (van der Linden & Hambleton, 1997).

The more significant distinction between these models, however, is that they differ in their fundamental purposes. Rasch models are used as confirmatory tests of the extent to which scales have been successfully developed according to explicit a priori measurement criteria. These criteria include the requirements that (a) items define a unidimensional continuum in the domain and (b) items follow a strictly hierarchical ordering in their definition of the domain. If the responses of study participants to the scale items suggest misfit to these criteria, particularly regarding hierarchical ordering, then the items are examined for the purpose of strengthening them—the Rasch measurement model is not discarded or modified. In contrast, 2-PL and 3-PL models are designed to maximize the extent to which item response variation can be accounted for—they are *statistical* models subject to reexpression in any way that reduces residual variation. Hence, 2-PL and 3-PL models will always fit any data set better than a Rasch model.

From a Rasch measurement perspective, better fit is not a sufficient reason for choosing an IRT model. Rasch models are preferred because they dictate the way analysts think about and subsequently construct measurement instruments. When the data fit the model, the continuous scale is analogous to a linear ruler that is invariant in terms of level of ease or difficulty of accomplishing the task for any individual appropriate for testing. With regard to measuring the variable learning to teach for social justice, a Rasch model was used, not because it fit the data better than any other model, but because when the data fit the model, teacher candidates could be ordered along a continuum based on their endorsements of simpler to more complex beliefs.

Although this notion may seem obvious to some, to measure a human characteristic well, one must know a great deal about it. This is true whether one's interests are in functional ability (Coster, Haley, Ludlow, Andres, & Ni, 2004; Coster, Ludlow, & Mancini, 1999; Haley, Ludlow, & Coster, 1993; Ludlow & Haley, 1992), test anxiety (Ludlow & Guida, 1991), reading ability (Ludlow & Hillocks, 1985), attitude toward math (Ludlow & Bell, 1996), rater effects (Ludlow & Haley, 1996), self-efficacy (Gable, Ludlow, & Wolf, 1990), job satisfaction (Ludlow & Lunz, 1998), or masculine stereotypes (Ludlow & Mahalik, 2001). To speak about measuring a variable is really to question what it means to be characterized as high, low, or moderate on a variable. Specifically, what kinds of tasks separate a person with high math ability from a person with low math ability, what kinds of behaviors characterize a high functioning person from a low functioning person, what distinguishes a teacher who possesses a high degree of commitment to teaching for social justice from one who does not have this commitment?

If one believes that a variable can be operationally hypothesized a priori in a continuous and hierarchical manner, then one can assert that tasks (items) can be created that represent those levels of the variable. Furthermore, if these items can successfully be constructed, then there is an opportunity to both conceptually and literally locate and position a person in relation to that variable and then describe the types of items, or tasks, that are most closely associated with that person's score (which defines that person's position or location) on the instrument (the variable being measured).

Underlying this concept of a variable is the assumption that the variable is unidimensional. This assumption means that there is an attempt to measure one characteristic of a person at a time. This is done not because humans are unidimensional but because to understand behavior well often means that the behavior first must be broken down into simple constructs. Once these simple constructs are clearly understood, it may then be possible to

build more complex models for the purpose of generating a more comprehensive picture of the person. Multidimensional Rasch models exist, but they have not yet made their way into standard practice (Embretson, 1991). They are, however, becoming more frequently used in computer-adaptive testing in the medical realm; for example, using a multidimensional computer-adaptive test to achieve greater levels of precision and efficiency than are possible with separate unidimensional assessments (Haley, Pengsheng, Ludlow, & Fragala-Pinkham, 2006).

The assumption is made that every variable stretches across a continuum of simple-to-complex tasks, or levels of knowledge, or affective characteristics, or cognitive abilities. Based on this assumption, a very deliberate hierarchical arrangement of tasks (like a ladder) can be constructed along which a person is located. This person could also be expected to advance along that continuum (or up the ladder). In a very real sense, a hypothesized structural model of behavior (or perception, attitude, or ability) is being constructed.

After this time-consuming, literature-exhausting effort to define a variable, which commences in item development, the instrument is administered to an appropriate sample of study participants (usually it is first administered as a series of pilot studies). The item responses are then analyzed to determine the extent to which the instrument functioned as intended, or the extent to which the data fit the model. This means the solution is examined to see if there is evidence to support the existence of a unidimensional variable stretching across the a priori defined continuum. It also means examining the extent to which study participants have given responses that are consistent with their estimated locations on the variable, and it means examining items to determine the extent to which these items provoked responses consistent with the locations of these same participants on the variable. This issue of person and item consistency is addressed through various goodness-of-fit statistics.

In this article, we do not argue that there is one and only one way to statistically model a vector of person-level item responses, nor that the best way to measure the variable learning to teach for social justice is through use of these statistical techniques. Our purpose is to illustrate that Rasch measurement principles offer an opportunity to construct measurement instruments that can be useful for understanding skills, abilities, attitudes, perceptions, and behaviors in ways that are fundamentally different from those of instruments designed to be fit by 2-PL IRT models.

Pilot Analyses

From a set of more than 200 potential items related to social justice that we gathered from the literature on teaching for social justice, a subset of 25 items that seemed to possess some degree of content validity was administered to 284 graduating teacher education students. The CTT analyses produced low reliabilities, low variances in the item and total scores, and poor item–total correlations. The Rasch analyses produced poor data-to-model fits and no indication of a meaningful continuum that made any theoretical sense with regard to the item or teacher candidate estimates (i.e., location estimates; see The Model section). These results were not entirely unexpected. The 5-year BC–TNE project had a short window of time available for the development of its first exit survey; therefore, either a rough survey with recognizable flaws would be administered or no survey would be administered to the 1st year's graduating class.

As a result of the poor performance of the 25 items on this first exploratory exercise, a more serious, conceptual, theory-driven approach was taken in the development of the Learning to Teach for Social Justice–Beliefs scale. For example, focus group exercises were conducted with students in our undergraduate and graduate classes. These exercises consisted of prompts such as, "What would be the characteristics of a teacher who is (or, is not) effective at teaching for social justice?" "How would you construct a scale to measure social justice?" and "What is a social justice item that anyone with even the slightest sense

of social justice would agree to?" We conducted these exercises in research methods, test construction, and psychometrics classes. They were introduced specifically during lectures addressing issues of operational definitions, item writing, reliability, construct and discriminant validity, instrument development, pilot testing, and data analysis. We also conducted them in graduate classes addressing philosophies and practice of education.

Those exercises produced a new set of 20 potential items that were subsequently piloted in different research methods classes. Students responded to each item using a 5-point, Likert-type scale that ranged from *strongly disagree* (1) to *strongly agree* (5). More importantly, they were asked to comment on the clarity of each statement (e.g., were any terms or phrases confusing, did double-barreled stems exist, did the scale responses seem appropriate, did the item itself seem related to social justice?).

These 20 items were analyzed using both CTT and Rasch measurement models. This second round of analysis led to additional changes in the items. Specifically, 8 of those 20 items revealed flaws (e.g., low item–total correlations, confusion in wording, poor fit) that eliminated them from future surveys.

The analyses were also discussed in interdisciplinary BC–TNE team meetings, which provided an additional layer of complexity to our task of attempting to measure the variable learning to teach for social justice. The most interesting aspects in the development of the instrument at this point were discussions that addressed the desired direction of scoring on the items. For example, would a teacher who is better prepared to teach for social justice *strongly agree* or *strongly disagree* with the statement, "Teachers should be 'color-blind' when it comes to working with students in the classroom"?

Current Operational Version of the Learning to Teach for Social Justice Scale

The final set of 12 social justice items resulting from these pilot analyses has been integrated into the BC–TNE Entry Survey, Exit Survey, One-Year Out Survey, Two-Year Out Survey, and Three-Year Out Survey system as the Learning to Teach for Social Justice scale (for details of other measures included in the surveys, see Ludlow et al., 2007). These surveys are completed by the teacher candidates and graduates at crucial times in their education and practice. The 12 items are identical on the Entry and the Exit Surveys, and they are worded such that teacher candidates are asked about their beliefs regarding teaching for social justice. This version of the scale is the Learning to Teach for Social Justice–Beliefs scale. The concepts and subjects addressed in the 12 items remain the same for the One- Two- and Three-Year Out Surveys; however, the items are reworded to ask teachers about their specific classroom practices regarding social justice. This version of the scale is the Learning to Teach for Social Justice–Practices scale. The analysis in this article addresses only the Learning to Teach for Social Justice–Beliefs scale used in the Entry and Exit Surveys.

The Learning to Teach for Social Justice–Beliefs scale (SJ) items are presented in Appendix A. For each item, respondents are asked to answer using a 5-point, Likert-type rating scale in which 1 = *strongly disagree*, 2 = *disagree*, 3 = *uncertain*, 4 = *agree*, and 5 = *strongly agree*. These items are positively and negatively worded. The positively worded items (Items SJ1, SJ2, SJ4, SJ7, and SJ8) were conceptualized and written as statements that address relatively less controversial aspects of teaching for social justice. We expected that students would agree with these items.

The inclusion of negatively worded items (Items SJ3, SJ5, SJ6, and SJ9–12) was deliberate. That is, we expected that it would be relatively easy to endorse positively worded statements with which most students should have some minimal level of experience, even in their first semesters of college. The negatively worded items, however, were intended to address concepts and experiences that only an experienced teacher candidate would have encountered. Moreover, to prevent response bias resulting from students responding in a socially desirable positive manner, these items were written in such a way that careful consideration of each item's meaning and intent would be required. Students were expected to disagree with these items. These items were reverse scored (R) so that higher scores (e.g., scores closer to 5) would correspond with a stronger commitment to teaching for social justice.

Once the items are coded in the appropriate direction, a higher total score corresponds to a higher level of commitment to teaching for social justice. We expected these scoring categories to generate wide variability in participant responses. Furthermore, we included a middle category choice of *uncertain* because we expected to see changes in the use of this category as teacher candidates progressed through their preparation. In other words, we expected that entering students would not fully understand some of the ideas and concepts that these items in the scale addressed because of their presumed lack of exposure to the concepts associated with the variable learning to teach for social justice. Hence, one measure of program effectiveness would be a reduction in the frequency with which the *uncertain* category was chosen by graduating candidates on the Exit Survey.

The first administration of these two surveys (Entry and Exit) to the two cohorts of student candidates and graduates now define a baseline for the remaining years of the BC–TNE project. Because an underlying goal of our collective research at Boston College is to transform the teacher preparation program, the analyses will be longitudinal and will provide evidence that feeds back into the program. Hence, the psychometric analyses reported in the current study were intended to ensure that the Learning to Teach for Social Justice–Beliefs scale was invariant across time of testing and experience with teaching. This property of invariance is a necessary condition in subsequent efforts to measure change within student candidates and graduates in the teacher preparation program.

In addition, we expected to find that the mean level of participants' scores would be relatively low for the incoming class (and all incoming classes) as measured by the Learning to Teach for Social Justice–Beliefs scale in the Entry Survey. We hypothesized that the graduating class (and all graduating classes) would score higher than the entering class as measured by the Learning to Teach for Social Justice–Beliefs scale in the Exit Survey and that as the BC–TNE effect would become more pronounced over time, the mean score of the graduating class on this scale would rise. This pattern would establish the discriminant validity of the scale.

Chronologically, the 2005 Exit Survey was completed by 224 graduating teacher candidates in late spring 2005, and the 2005 Entry Survey was completed by 268 incoming undergraduate and graduate students in early fall 2005. The overall response rate was 88.2% and 90.0% for the Exit and Entry surveys, respectively.

CTT RESULTS

Entry Survey

For those in the incoming class who completed the Entry Survey, the responses to the 12 items of the Learning to Teach for Social Justice–Beliefs scale generated a Cronbach's alpha of .77, and the item analysis revealed no negative point-biserial correlations. In addition, the items were factor analyzed for the purpose of providing one form of construct validity. Prior to the factor analysis, we checked the appropriateness of the correlation matrix for factoring. The Kaiser–Meyer–Olkin (KMO) statistic was high (.809), the determinant was nonzero, and Bartlett's test of sphericity was significant, all of which were encouraging results.

Using principle axis factoring (thereby choosing to analyze only common variance) with a varimax rotation, two factors were extracted. The two factors accounted for 34.1% of the total variance. Factor 1 accounted for 17.7%; Factor 2 added 16.4%. These factors were significantly correlated with one another. Specifically, when an oblique rotation was performed, the factor correlation was .36.

The seven negatively worded items define Factor 1, the five positively worded items define Factor 2. Given the scale development specifications, this result was expected. The plot of

the varimax loadings is seen in Figure 1, Box 1. Does this result suggest that two scores should be reported for each person—one positive and one negative score? Or, based on the oblique rotation and the correlation between the two factors, is it possible that the solution can be interpreted as a single variable composed of two clusters of similarly worded items addressing relatively easy and more difficult to endorse aspects of that variable?

Recall that the factors in principal axis factoring are extracted in such a way that each eigenvector (or factor) accounts for the maximum orthogonal variance possible. Hence, the first factor mathematically defines a solution along which the items share the greatest common variance. The second extracted factor then defines a projection through the variable space that accounts for the next greatest amount of common variance. The unrotated factors are customarily rotated because the factor loadings are usually too numerous and too variable to comprehend when presented in a table, and plots of the unrotated factors are usually too indistinct to distinguish any meaningful pattern. That is not the case, however, with the data in this study.

Figure 2, Box 1 contains the unrotated solution for the Entry Survey data. Inspection of the unrotated factor loadings reveals that the three highest loadings on Factor 1 are Items SJ3R, SJ4, and SJ10R—all addressing diversity. (See Appendix A for wording of all items.) The next two items address English language issues (Items SJ5R and SJ6R), and the next two items address race issues (Items SJ1 and SJ2). These items consist of both positive and negative statements. This first unrotated factor provides evidence for the existence of a common Learning to Teach for Social Justice–Beliefs factor. The second unrotated factor then separates out the positive and negative statements (the small circle above the horizontal axis and the larger circle below the horizontal axis, respectively).

The question remains, however—do the circles represent substantive differences or trivial differences that are due to the direction of the phrasing? To answer this, the seven negatively worded items were rephrased as positively worded items. For example, Item SJ3R, "For the most part, covering multicultural topics is only relevant to certain subject areas, such as social studies and literature," became "Covering multicultural topics is relevant to all subject areas, including math and science." If responding to the negatively phrased items were simply a more difficult cognitive task, then analysis of this new scale would not show two factors, or clusters, of positive and negative items. Conversely, if the positive and negative items reflected substantive differences, as intended, then factors and clusters similar to the initial solution should result.

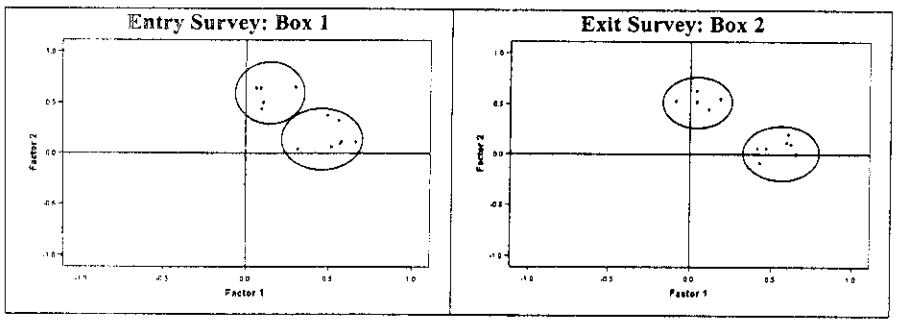


FIGURE 1

Varimax Solutions for the Data From the Entry and Exit Survey Learning to Teach for Social Justice–Beliefs Scale

Note. Factor 1 = negatively worded items; Factor 2 = positively worded items.

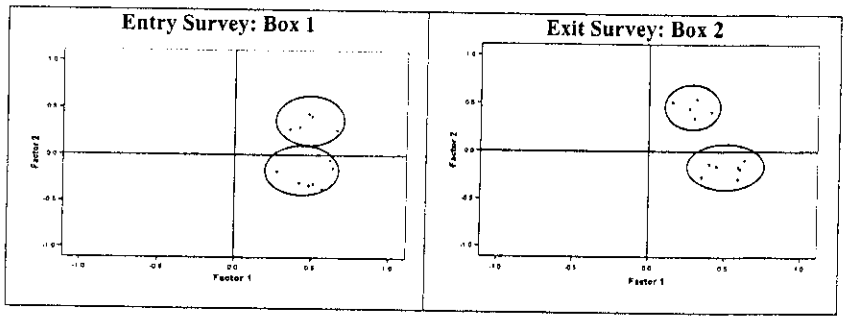


FIGURE 2

Unrotated Solutions for the Data From the Entry and Exit Survey Learning to Teach for Social Justice–Beliefs Scale

Note. Factor 1 = Learning to Teach for Social Justice–Beliefs scale; Factor 2 = less complex and more complex clusters.

Accordingly, the new set consisting only of 12 positively worded items was administered to two graduate-level classes—one studying psychometrics, the other studying curriculum controversies. These students were not comparable to those who responded to the Entry and Exit Surveys. The data were factored as previously described and nearly identical clusters of items emerged. These results provide evidence that the two clusters of items share a common factor addressing learning to teach for social justice but are distinguishable in that they address different aspects of this variable.

Exit Survey

A similar analysis was conducted on the Exit Survey data for the spring 2005 graduating teacher candidates. These students were exposed to the inherent emphasis placed on social justice that characterizes Boston College, but they experienced minimum exposure to any demonstrable BC–TNE effect because the teacher education program was undergoing changes in curriculum and practica at that time. Their results serve as a baseline against which the results from all future graduating classes will be compared.

The responses of the graduating teacher candidates had a lower reliability estimate compared with the Entry Survey results ($\alpha = .71$). Interestingly, the scale variance on the Exit Survey (28.9) was considerably less than the variance on the Entry Survey (37.9). Recalling that the magnitude of the Cronbach's alpha is a function of the scale variance, this slight decline in alpha is understandable and meaningful. This reduction in scale variance may itself be understood as a program effect of having brought a degree of homogeneity to the students after 4 years of study.

The KMO statistic was high (.755), the determinant was nonzero, and Bartlett's test of sphericity was significant. Principal axis factoring with a varimax rotation extracted two factors explaining 29.6% of the total variance. Factor 1 (negatively worded items) accounted for 17.5% of the variance; Factor 2 (positively worded items) accounted for 12.1%. The oblique rotation generated a factor correlation of .15. Similar rotated and unrotated factor loadings were found in comparison to the Entry Survey. The rotated plot is presented in Figure 1, Box 2; the unrotated plot is presented in Figure 2, Box 2. In Box 2, the unrotated Factor 1 defines the Learning to Teach for Social Justice–Beliefs scale, while the unrotated Factor 2 again separates the positive (less complex) and negative (more complex) items.

The factor structures for the Entry and Exit Surveys are similar in appearance (whether one chooses to use the rotated or unrotated solutions), with the negatively worded items

loading highest on one factor and the positively worded items loading highest on the other factor. However, it is important to test whether these results are statistically equivalent across the two administrations. For this purpose, coefficients of congruence (CC) were computed. The CC is a measure of the degree to which the factor loadings on corresponding administrations are similar (Cureton & D'Agostino, 1983). Because these factors were derived using the same factoring procedures and were guided by the simple-structure criterion, this comparison is appropriate. The CC will equal 1.0 if the factor loadings are identical, and Cureton and D'Agostino suggested that the CC should be at least .90 if the two administrations produce the same factor. The coefficient is expressed as

$$CC = \frac{\sum I_1 I_2}{\sqrt{\left(\sum I_1^2 \sum I_2^2\right)}},$$

where I_1 corresponds to the factor loadings from the first solution, I_2 corresponds to the factor loadings on the same factor from the second solution, and n represents the number of items.

The CC for the positively worded factor was .95; for the negatively worded factor, it was .97. These comparisons meet the .90 criterion—the same factors were extracted across samples and levels of experience.

Having satisfied ourselves that the classical item analysis results suggest similar levels of stability and consistency within the scale across the two groups of teacher candidates, we next performed a simple, independent means t test on the total scores. The two means differed significantly—graduating students obtained statistically significantly higher scores than did entering students. These findings are consistent with our expectations and serve as evidence of the scale's discriminant validity.

RASCH MODEL RESULTS

The Model

The model used in this study is known as the Rasch rating scale (Andrich, 1988; Wright & Masters, 1982, p. 49). This model is appropriate when the scoring categories have rigorously defined scoring rubrics that transcend, or do not depend on, the characteristics of the specific items. The same analysis was performed on the two survey administrations. The WINSTEPS (Version 3.64.2) software package was used (Wright & Linacre, 1998, see Appendix B).

$$\pi_{nix} = \frac{e^{\sum_{j=0}^{x}[\beta_n - (\delta_i + \tau_j)]}}{\sum_{k=0}^{m} e^{\sum_{j=0}^{k}[\beta_n - (\delta_i + \tau_j)]}}$$

The π_{nix} is the probability of person n responding in category x to item i where δ_i is the location (generically referred to as item difficulty) of item i on the variable, τ_j is the location (threshold parameter) of the k^{th} transition from one scoring category to the next for the $m + 1$ rating categories, and β_n is the parameter for an individual's level of commitment to teach for social justice. These location estimates are reported in a metric referred to as *logits* (Ludlow & Haley, 1995). These estimates simultaneously portray the structure of the learning to teach for social justice–beliefs variable and the location of each teacher candidate along the variable.

Scale Structure

Figure 3 contains the separate *variable maps* for the 2005 Entry and 2005 Exit Surveys. In each map, the items are ordered from easiest to endorse (bottom of the map) to hardest to

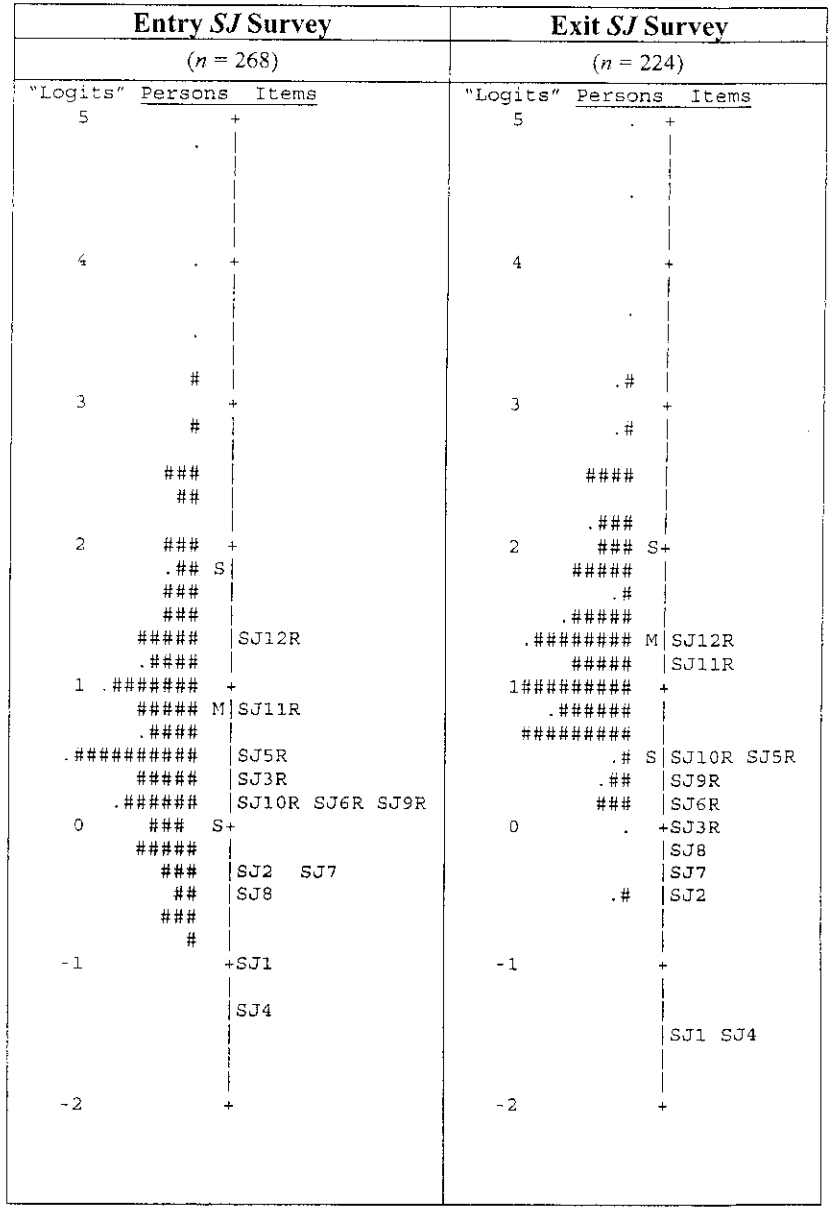


FIGURE 3

Variable Maps for the Data From the Entry and Exit Survey Learning to Teach for Social Justice–Beliefs Scale

Note. SJ = Learning to Teach for Social Justice–Beliefs scale; R = reverse-scored item; M = mean candidate estimate; S = standard deviation of estimates. Each "#" is 3 candidates. Each "." is 1 candidate.

endorse (top of the map). The teacher candidates are ordered from lowest scoring (bottom of the map) to highest scoring (top of the map) in terms of learning to teach for social justice beliefs. As demonstrated by the easiest to endorse items, it is easiest to *strongly agree* with Items SJ1 and SJ4 (examine one's own beliefs and incorporate diverse cultures), followed by Items SJ8 (think critically about government policies), SJ2 (racism and inequity openly discussed), and SJ7 (challenge school arrangements). Note that these are the positively worded items identified in the factor analyses. (See Appendix A for exact wording.)

At the top of each map are the hardest items to endorse (by disagreeing). Item SJ12R (prepare students for lives they are likely to lead) is harder to reject than Item SJ11R (success in school depends primarily on how hard students work). This is followed by Item SJ5R (assimilate immigrant children and ELL into American society) and Item SJ3R (covering multicultural topics is only relevant to certain subjects). In the middle of the distribution are Items SJ6R, SJ9R, and SJ10R (lower expectations for ELL, economically disadvantaged bring less and gain more, and it is not the teacher's job to change society, respectively). It is encouraging and consistent with our scale development expectations that proceeding up the variable means addressing increasingly more complex, controversial, and debatable choices.

The Entry Survey item separation is 8.6, and the person separation is 1.8. For the Exit Survey, item and person separation are 8.4 and 1.5, respectively. These statistics take into account the standard deviation of the parameter estimates relative to the mean measurement error in those estimates (Wright & Masters, 1982). Because these statistics are partly a function of the sample size and number of items, respectively, these values represent reasonable spread in the item and person estimates.

Variable maps show the mean location for each item based on the item's total score across all candidates. They do not reveal the location of different response categories or the level of response expected of a person to any item at a given location. The scale structure portraying the lowest and highest response category thresholds for the Entry Survey is presented in Figure 4.

In this variable map, the locations of the candidates are the same as in Figure 3. The column labeled "Mean Item Estimate" places the items in the same locations as shown in Figure 3. The column labeled "1st Threshold for Items," however, places the items at positions defined as mean item estimate plus the 1st threshold estimate, in this case −1.71. This set of locations shows how probable it is for the candidates to respond with a 2 on the items. The column labeled "4th Threshold for Items" places the items at positions defined as mean item estimate plus the 4th threshold estimate, in this case 2.17. This set of locations show how probable it is for the candidates to respond with a 5 on the items. This map clearly shows that most candidates, at the time they completed the Entry Survey, are in the *agree* to *strongly agree* range on the familiar items addressing broad-based issues of equity and diversity. But they are in the *uncertain* to *disagree* range on the more focused political action items (recall that a 5 on the reverse-scored items corresponds to a *strongly disagree* response). This map will serve as a baseline to chart not only progress over time but also the strength of commitment associated with that progress.

The next step was to determine the invariance of the Learning to Teach for Social Justice–Beliefs scale by comparing the results from the two administrations. A close inspection of Figure 3 for the Entry and Exit Surveys reveals a similar order to the item location estimates. The Entry Survey estimates are slightly more bunched, whereas there is a slightly more uniform spread in the Exit Survey estimates. This is reasonable because the graduating students have more experience with these ideas and can make finer distinctions between them. It is also evident that the person locations (measures) are relatively higher on the Exit Survey scale—note the location of the mean candidate estimate (M). Again, this is reasonable because these students have learned these principles as part of their teacher education program.

A plot of the 12 pairs of item estimates from the Entry and Exit Survey solutions provides additional evidence that the structure of the Learning to Teach for Social Justice–Beliefs

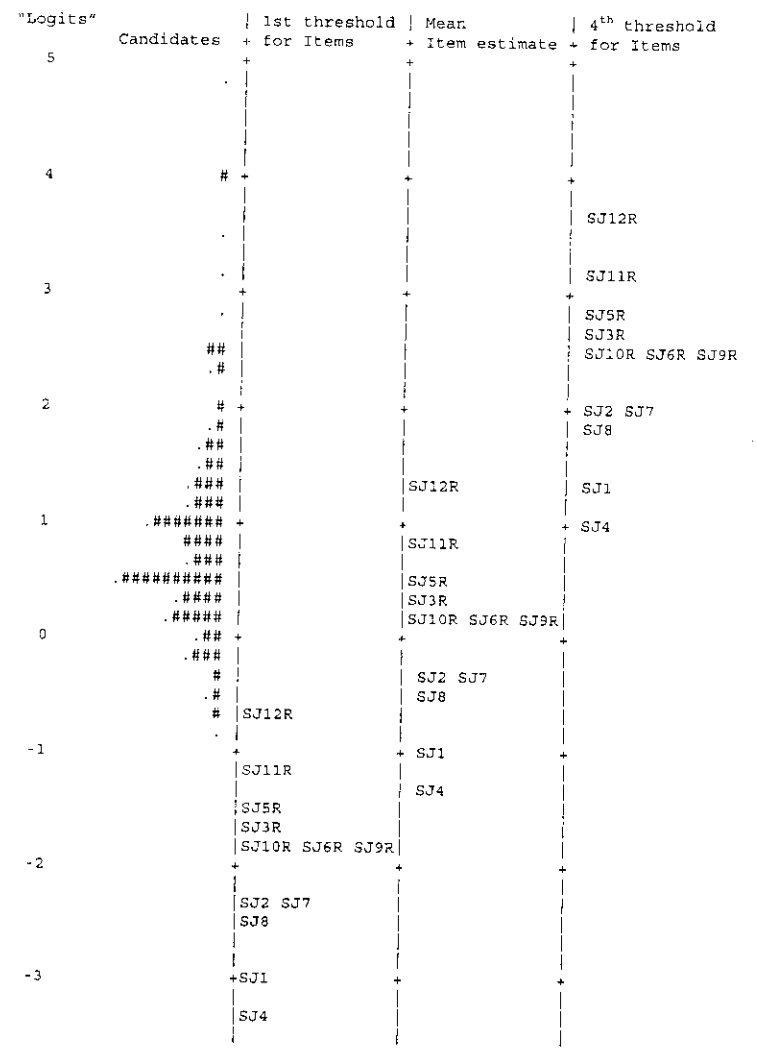


FIGURE 4

Variable Map of Candidates, Learning to Teach for Social Justice–Beliefs Scale Items, and Lowest and Highest Response Thresholds: Entry Survey

Note. SJ = Learning to Teach for Social Justice–Beliefs scale; R = reverse-scored item. Each "#" is 4 candidates. Each "." is 1 to 3 candidates.

scale is invariant across the two administrations (see Figure 5). The ranges and variances are comparable, with the Exit Survey estimates slightly more spread out. The Pearson product–moment correlation is .966. These results support the earlier factor analysis statistics, plots, and congruence coefficients.

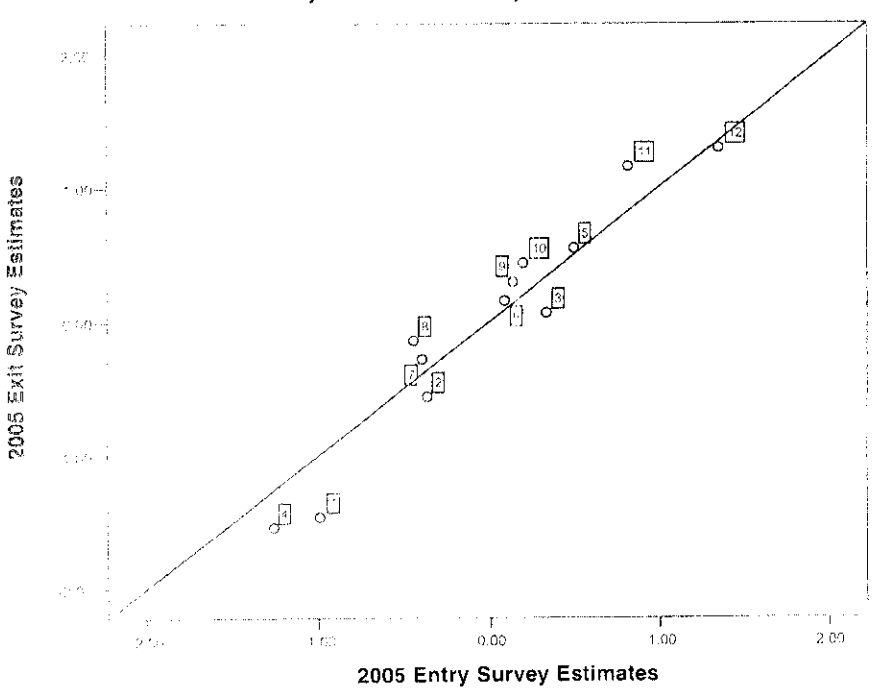


FIGURE 5

Entry Versus Exit Survey Learning to Teach for Social Justice–Beliefs Scale Item Estimates

Note. Entry item estimate range = 2.6, variance = .74. Exit item estimate range = 2.8, variance = .88. Pearson r = .966.

Category Characteristic Curves

Category characteristic curves (CCCs) display the probabilities of responding to each category for any person on any item. From a measurement and instrument development perspective, the ideal pattern in the CCCs occurs when each response category is the expected response for some level of person-by-item interaction. This pattern occurs when the threshold estimates follow a strictly increasing order. As seen in Figure 6, the spread in the CCCs and category threshold estimates (where the CCCs intersect) is excellent for the Entry Survey. That is, when a candidate with a relatively low estimate of commitment to teaching for social justice answers an item with a relatively high estimate of commitment to teaching for social justice and the difference is, for example –1.0, the expected response is 2. Likewise, when a candidate with a relatively high estimate responds to a relatively lower level item and the difference in the estimates is, for example 2.0, the expected response is 5.

The pattern for the Exit Survey threshold estimates in Figure 7 is noteworthy because it differs from the Entry Survey estimate's pattern. The Entry Survey threshold estimates were (–1.70, –0.50, 0.05, 2.20)—this is the desired pattern. The Exit Survey thresholds were (–1.30, –0.07, –0.40, 1.70)—here there is a disordinal pattern in the estimates. The middle

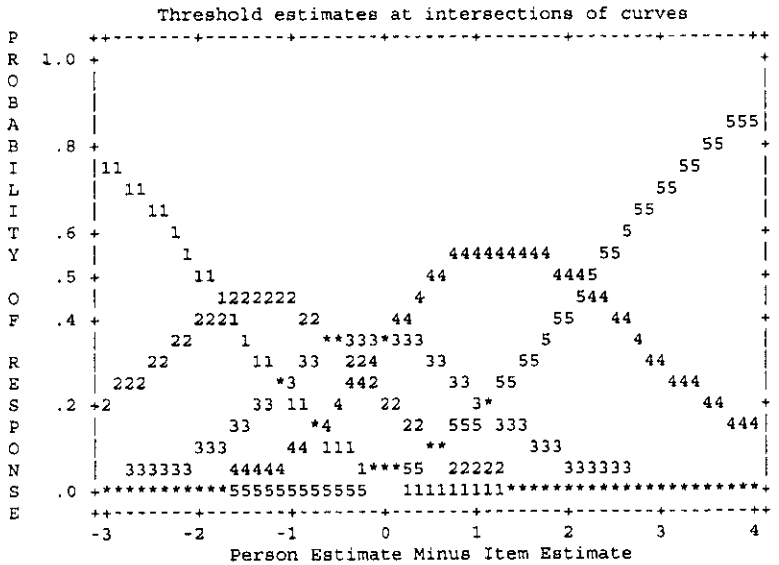


FIGURE 6

Category Characteristic Curves (CCCs) for the Data From the Learning to Teach for Social Justice–Beliefs Scale Entry Survey

Note. Threshold estimates: (–1.70, –0.50, 0.05, 2.20).

category of *uncertain* is not the expected response for any level of person-by-item interaction. This difference in the patterns is, however, a reasonable and expected result. It means that incoming students were relatively uncertain about some concepts and ideas when they entered the teacher education program. They may not have had much experience with these issues and may not have thought much about these specific issues. Graduates, in contrast, have spent up to 4 years in a program that is known for its commitment to social justice and have spent a great deal of time addressing these ideas. These graduating teachers are no longer as uncertain about their beliefs regarding learning to teach for social justice.

Fit Analysis

Rasch goodness-of-fit analyses rely principally on standardized residuals—the difference between the observed response and the response expected under the model (Wright & Stone, 1979). A positive residual results when a higher than expected response occurs; negative residuals result from lower than expected responses. Although a variety of statistical and graphical procedures are available for analyzing residuals (Ludlow, 1985, 1986), a standard first approach is to consider summary statistics in the form of standardized, weighted mean squares. Although these statistics do not have exact degrees of freedom and critical values, a rich history of experience has developed regarding their general properties and utility (see, for example, the work of Smith, 1991).

The Rasch goodness-of-fit analyses generally start with the standardized, weighted statistics because they are roughly analogous to t statistics and take into account the variance of the expected response (the so-called ZSTD INFIT in the WINSTEPS [Version 3.64.2] software). A very flexible criterion of +2 to +3 is often used initially. Because the version

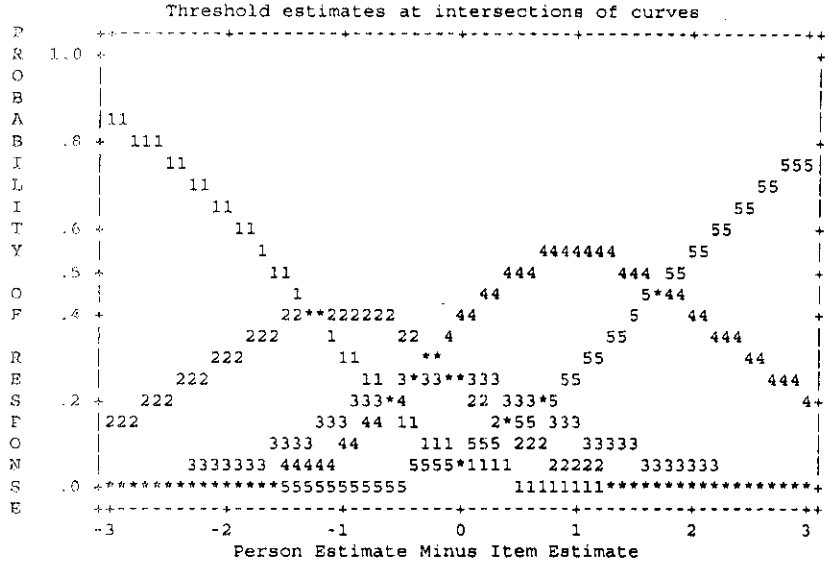


FIGURE 7

Category Characteristic Curves (CCCs) for the Data From the Learning to Teach for Social Justice–Beliefs Scale Exit Survey

Note. Threshold estimates: (–1.30, –0.07, –0.40, 1.70).

of this statistic for items is easily inflated as the sample size grows greater than 500 or so people, the unstandardized version (the so-called mean square INFIT) is also checked. Here a flexible criterion of +1.3 is often used to flag potential problems. These two criteria are generally sufficient to reveal consistent unexpected responses that are either made to an item (item fit) or made by a person (person fit).

A review of the flagged fit statistics and the largest standardized residuals for the data in this study revealed that roughly 12% of the candidates in each administration produced noticeable unexpected responses, whereas only a few items provoked consistent unexpected responses. Although there were a few clear instances of extremely unexpected responses (e.g., a *strongly disagree* when a *strongly agree* response was expected), most of the unexpected residual variation occurred as a matter of degree. That is, there were frequent instances in which a *strongly agree* was expected but only an *agree* was provided. This was particularly apparent on the Entry Survey, where an *agree* or *disagree* response would be expected, but an *uncertain* response was provided. These situations, however, do not warrant any concern about poor data-to-model fit.

Measuring Change

Given that the purpose of developing the Learning to Teach for Social Justice–Beliefs scale is to measure changes in beliefs, we used the Entry Survey item estimates as anchors to reestimate the candidates' Exit Survey responses. Although this is not strictly necessary because the scale estimates are independent of the person estimates when the data fit the model, it is useful to measure differences (growth) in the entering and graduating teachers by placing them on the baseline scale defined by the incoming students.

Figure 8 contains the anchored estimates for the Exit Survey responses based on the Entry Survey estimates. The difference in the two groups is still evident but the advantage of this procedure is that the scale structure is fixed and we have an unambiguous interpretation of the distributional differences in terms of progress up the scale for respondents. For example, the mean response for the Entry Survey corresponds roughly to *disagree* with SJ11R: "Whether students succeed in school depends primarily on how hard they work." The mean response for the Exit Survey corresponds roughly to *disagree* with SJ12R: "Realistically, the job of a teacher is to prepare students for the lives they are likely to lead."

When the anchoring is based on the Entry Survey estimates, the candidates' estimates when they graduate are measures of how far they have gone relative to the Entry Survey.

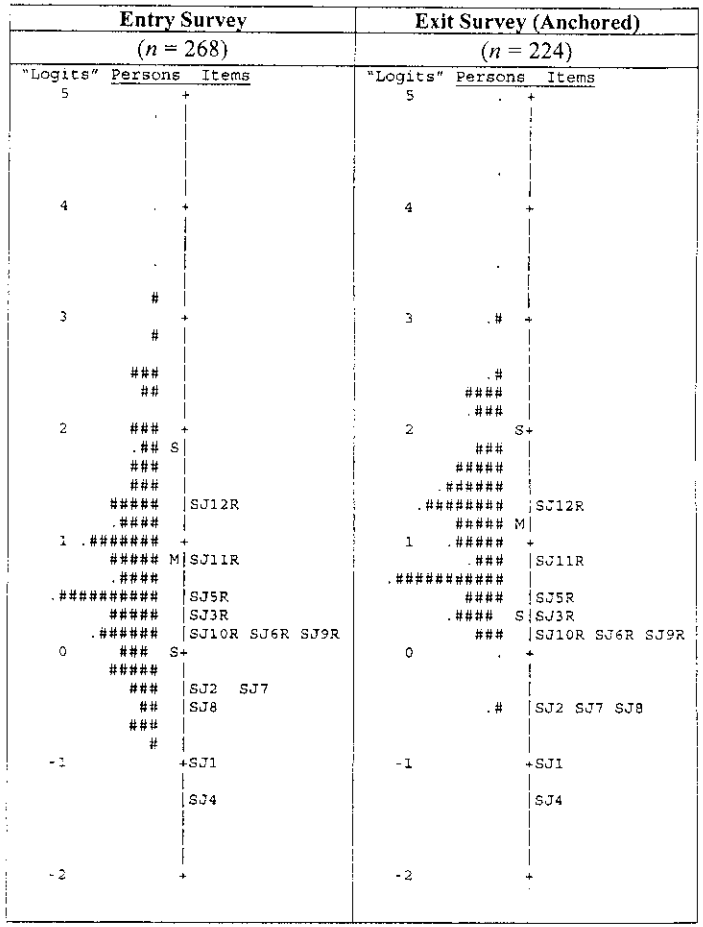


FIGURE 8

Anchored Variable Maps for the Data From the Entry and Exit Survey Learning to Teach for Social Justice–Beliefs Scale

Note. SJ = Learning to Teach for Social Justice–Beliefs scale; R = reverse-scored item; M = mean candidate estimate; S = standard deviation of estimates. Each "#" is 3 candidates. Each "." is 1 candidate.

On the other hand, a reasonable argument can be made that the anchoring should be performed based on the Exit Survey results. This approach would represent the desired outcome and the entering candidates would be measured in terms of how far they have to go (Ludlow, Haley, & Andres, 2005) or the extent to which they have to learn to teach for social justice. Using the Exit Survey estimates for anchoring on is also reasonable strictly from the scale definition perspective because the items at the Exit Survey form a clearer continuum. Both anchoring solutions will likely be applied over the course of the study with future administrations of these two surveys, because they each address a slightly different but meaningful question.

DISCUSSION

The purpose of this article was to explain the development of the Learning to Teach for Social Justice–Beliefs scale. Although the psychometric analyses could have stopped with the CTT results, we performed Rasch rating scale analyses for several important reasons. First, the items were intended to operationally define a unidimensional construct of learning to teach for social justice. This construct, in turn, was intended to define a continuum of questions corresponding to beliefs ranging from a weaker to a stronger commitment to teaching for social justice. Hence, the Rasch analyses served as a confirmatory test of the a priori hypothesized structure. Third, and most important, if the scale was invariant across the two surveys (as defined by similar item difficulty to endorse estimates), then it would be possible to measure how teacher candidates' beliefs develop and progress from the time they enter their teacher education program and to their subsequent graduation to the classroom.

The Rasch rating scale analyses were of paramount importance in establishing the invariant structure of the Learning to Teach for Social Justice–Beliefs scale over the two groups of teacher candidates. The subsequent establishment of this structure now provides an opportunity to track meaningful positive changes in the perceptions, attitudes, and experiences around the practice of teaching for social justice as Boston College teacher candidates progress through their teacher education program. Success in constructing such a scale not only contributes to the efforts of Boston College to gather evidence of the effectiveness of its BC–TNE initiatives but also constitutes a major contribution to the literature on social justice and its measurement.

Teaching for social justice has been described, for the purpose of the research at Boston College, as a pedagogy that is intended to foster pupil's learning and to help teachers to understand the social and institutional inequities that are embedded in our society. Through this scale, we have presented the idea that learning to teach for social justice is a measurable outcome of teacher preparation. The results provide evidence that this is indeed a reasonable task. These items are internally consistent, conceptually unidimensional, and define a meaningful and theoretically defensible continuum. Thus, future use of this Learning to Teach for Social Justice scale will provide valuable evidence of the impact of teacher education on learning to teach for social justice.

Based on the differences between the two surveys, we expect that entering students will always be at a similar point along this continuum because we are assuming that each class of incoming students will have similar entering characteristics. We then expect that graduating students will always be at a higher point because they have learned to teach for social justice as a result of their teacher preparation. Furthermore, we expect that as BC–TNE efforts are implemented in the coming years, graduating teachers will be exposed to a measurable BC–TNE program effect. This program effect will be observed through expected increases in each year's graduating class, as compared with former graduates, regarding their commitment to teaching for social justice.

REFERENCES

Adams, M., Bell, L. A., & Griffin, P. (Eds.). (1997). *Teaching for diversity and social justice*. New York: Routledge.

Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.

Brennan, R. L. (2001). Some problems, pitfalls, and paradoxes in educational measurement. *Educational Measurement: Issues and Practice, 20,* 6–18.

Cochran-Smith, M. (1999). Learning to teach for social justice. In G. Griffin (Ed.), *The education of teachers: Ninety-eighth yearbook of the National Society for the Study of Education* (pp. 114–144). Chicago: University of Chicago Press.

Cochran-Smith, M. (2004). *Walking the road: Race, diversity, and social justice in teacher education*. New York: Teachers College Press.

Coster, W. J., Haley, S. M., Ludlow, L. H., Andres, P. L., & Ni, P. S. (2004). Development of an applied cognition scale to measure rehabilitation outcomes. *Archives of Physical Medicine and Rehabilitation, 85,* 2030–2035.

Coster, W. J., Ludlow, L. H., & Mancini, M. (1999). Using IRT variable maps to enrich understanding of rehabilitation data. *Journal of Outcome Measurement, 3,* 123–133.

Cureton, E. E., D'Agostino, R. B. (1983). *Factor analysis: An applied approach*. Hillsdale, NJ: Erlbaum.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56,* 495–516.

Gable, R. K., Ludlow, L. H., & Wolf, M. B. (1990). The use of classical and Rasch latent trait models to enhance the validity of affective measures. *Educational and Psychological Measurement, 50,* 869–878.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Haley, S. M., Ludlow, L. H., & Coster, W. J. (1993). Pediatric Evaluation of Disability Inventory: Clinical interpretation of summary scores using Rasch rating scale methodology. *Physical Medicine and Rehabilitation Clinics of North America, 4,* 529–540.

Haley, S. M., Pengsheng, N., Ludlow, L. H., & Fragala-Pinkham, M. (2006). Measurement precision and efficiency of multidimensional computer adaptive testing of physical functioning using the Pediatric Evaluation of Disability Inventory. *Archives of Physical Medicine and Rehabilitation, 87,* 1223–1229.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hattie, J., Jaeger, R. M., & Bond, L. (1999). Chapter 11: Persistent methodological questions in educational testing. *Review of Research in Education, 24,* 393–446.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Ludlow, L. H. (1985). A strategy for the graphical representation of Rasch model residuals. *Educational and Psychological Measurement, 45,* 851–859.

Ludlow, L. H. (1986). Graphical analysis of item response theory residuals. *Applied Psychological Measurement, 10,* 217–229.

Ludlow, L. H., & Bell, K. N. (1996). Psychometric characteristics of the Attitudes Towards Mathematics and Its Teaching (ATMAT) scale. *Educational and Psychological Measurement, 56,* 864–880.

Ludlow, L. H., & Guida, F. V. (1991) The Test Anxiety Scale for Children as a generalized measure of academic anxiety. *Educational and Psychological Measurement, 51,* 1013–1022.

Ludlow, L. H., & Haley, S. M. (1992). Polytomous Rasch models for behavioral assessment: The Tufts Assessment of Motor Performance. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (Vol. 1, pp. 121–137). Norwood, NJ: Ablex.

Ludlow, L. H., & Haley, S. M. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement, 55,* 967–975.

Ludlow, L. H., & Haley, S. M. (1996). Effect of context in rating of mobility activities in children with disabilities. *Educational and Psychological Measurement, 56,* 122–129.

Ludlow, L. H., Haley, S. M., & Andres, P. (2005). Measuring change: Functional activity item calibrations at admission vs. discharge. In N. Bezruczko (Ed.), *Rasch measurement in health sciences*. Maple Grove, MN: JAM [Journal of Applied Measurement] Press.

Ludlow, L. H., & Hillocks, G , Jr. (1985). Psychometric considerations in the analysis of reading skill hierarchies. *Journal of Experimental Education, 54,* 15–21.

Ludlow, L. H., & Lunz, M. (1998). The Job Responsibilities Scale: Invariance in a longitudinal prospective study. *Journal of Outcome Measurement, 2,* 326–337.

Ludlow, L. H., & Mahalik, J. R. (2001). Congruence between a theoretical continuum of masculinity and the Rasch model: Examining the Conformity to Masculine Norms Inventory. *Journal of Applied Measurement, 2,* 205–221.

Ludlow, L. H., Pedulla, J., Enterline, S., Loftus, F., Kafka, A., Saloman-Fernandez, Y., et al. (2007, April). *From students to teachers: Teacher candidates' and graduates' perceptions and experiences.* Paper presented at the American Educational Research Association annual meeting, Chicago.

Masters, G. (1988). Item discrimination: When more is worse. *Journal of Educational Measurement, 25,* 15–29.

Michelli, N., & Keiser, D. (Eds.). (2005). *Teacher education for democracy and social justice.* New York: Routledge.

North, C. (2006). More than words? Delving into the substantive meaning(s) of "social justice" in education. *Review of Educational Research, 76,* 507–535.

Oakes, J., & Lipton, M. (1999). *Teaching to change the world.* Boston: McGraw-Hill.

Sleeter, C. (1996). *Multicultural education as social activism.* Albany: State University of New York Press.

Smith, R. M. (1991). *IPARM: Item and person analysis with the Rasch model.* Chicago: MESA Press.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15,* 72–101.

Traub, R. E. (1997). Classical test theory in historical perspective. *Educational Measurement: Issues and Practice, 16,* 8–14.

van der Linden, W., & Hambleton, R. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer.

Villegas, A. M., & Lucas, T. (2002). *Educating culturally responsive teachers: A coherent approach.* Albany: State University of New York Press.

Wainer, H. (1986). Can a test be too reliable? *Journal of Educational Measurement, 23,* 171–173.

Wright, B. D., & Linacre, M. (1998). WINSTEPS. Chicago: MESA Press.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis.* Chicago: MESA Press.

Wright, B. D., & Stone, M. (1979). *Best test design.* Chicago: MESA Press.

Zeichner, K. (2003). The adequacies and inadequacies of three current strategies to recruit, prepare, and retain the best teachers for all students. *Teachers College Record, 105,* 490–515.

APPENDIX A

Entry and Exit Survey Items of the Learning to Teach for Social Justice–Beliefs Scale

Respond to the following statements regarding your beliefs about teaching:

1 An important part of learning to be a teacher is examining one's own attitudes and beliefs about race, class, gender, disabilities, and sexual orientation.
2 Issues related to racism and inequity should be openly discussed in the classroom.
3R For the most part, covering multicultural topics is only relevant to certain subject areas, such as social studies and literature.
4 Good teaching incorporates diverse cultures and experiences into classroom lessons and discussions.
5R The most important goal in working with immigrant children and English language learners is that they assimilate into American society.
6R It's reasonable for teachers to have lower classroom expectations for students who don't speak English as their first language.
7 Part of the responsibilities of the teacher is to challenge school arrangements that maintain societal inequities.
8 Teachers should teach students to think critically about government positions and actions.
9R Economically disadvantaged students have more to gain in schools because they bring less to the classroom.
10R Although teachers have to appreciate diversity, it's not their job to change society.
11R Whether students succeed in school depends primarily on how hard they work.
12R Realistically, the job of a teacher is to prepare students for the lives they are likely to lead.

Note. Likert-type rating scale for responses: 1 = *strongly disagree;* 2 = *disagree;* 3 = *uncertain;* 4 = *agree;* 5 = *strongly agree.* R denotes reverse-scored items.

APPENDIX B

WINSTEPS Control File for the 2005 Entry Survey Learning to Teach for Social Justice–Beliefs Scale

```
; This is a WINSTEPS specification control file
; a semi-colon means a comment
&INST
TITLE = "Social Justice: Entry 2005"
        ;Input Data Format
NAME1 = 1               ; column of start of person information
NAMLEN = 3              ; maximum length of person information
ITEM1 = 4               ; column of first item-level response
NI = 12                 ; number of items = test length
XWIDE = 1               ; number of columns per response
PERSON = Person         ; Persons are called ...
ITEM = Item             ; Items are called ...
        ; Data Scoring
CODES = 12345           ; valid response codes
CLFILE = *              ; label the categories in Table 3
1 Weak                  ; 0 in the data means "Strongly Disagree"
5 Strong                ; 1 in the data means "Strongly Agree"
*
        ;User Scaling
UMEAN = 0               ; item mean - default is 0.00
USCALE = 1              ; measure units - default is 1.00
UDECIM = 2              ; reported decimal places - default is 2
MRANGE = 0              ; half-range on maps - default is 0 (auto-scaled)
&END                    ; enter the item names
sj1
sj2
sj3
sj4
sj5
sj6
sj7
sj8
sj9
sj10
sj11
sj12
END LABELS              ;enter the data
  1545532453411
  2554445442344
        .
        .
  283432432434111
  28454444544333
```

Note. WINSTEPS is a software package (Wright & Linacre, 1998).