

Index

- Accountability: changing standards for, 259; commonsense steps for, 307–8; decentralization and, 310–11; development of (Texas), 305–6; educators and academic results, 307–8; evaluation of Kentucky programs for, 297–99; individual program evaluation and, 320; Kentucky “thresholds” and, 252–53; leadership and, 306–7; methods for setting standards for, 259–60; need for more attention on high school performance, 319; opposition to (in Texas), 308; and the result of educational efforts, 311; school choice and, 20–21; special interests and, 20; and systemic change in Texas, 315–16; tests and, 19–20; Texas testing program, 308–9
- Advanced Placement courses, 41
- American Educational Research Association (AERA), 28–29, 30, 37, 38, 58
- Angoff method, 259
- Assessment: classroom teachers and, 101–2; instructional practices and, 228; need for earlier and wider use in Texas, 318; and raising the bar for all Texas schools, 318; results in Texas, 316–17; and social promotion, 318–19
- Assessment systems: determining test validity, 99–100; validity and, 100–102
- Assessment tests: criterion-referred instruments in Texas, 309; minority performance on multiple-choice, 286; multiple-choice versus other forms, 284–88; reasons for not using multiple-choice format, 285–86
- Authentic-testing movement, 13
- Authentic tests, 12–13
- Best Practices in Ungraded Classrooms*, 278
- Bishop, John, 52–53
- Blanchard, Paul, 247
- Bush, George W., and education reform in Texas, 306–7
- California Learning Assessment System (CLAS), and portfolio assessment, 205–6
- California Science Curriculum Framework (2002), 138
- Cannell, John J., 43–44
- Catterall Report*, 266
- Center for Research on Evaluation, Standards, and Student Testing (CRESST), 29
- Center for the Study of Testing, Evaluation, and Educational Policy (CSTEAP): and defining how higher-order thinking is tested, 50; study of commercially available tests, 47–48; and teaching to the test, 43
- Centralized funding and performance of U.S. schools, 7–8
- Change scores, 267–68
- Charter schools, 21, 22
- Chicago Charter School Foundation, 21–22
- Chicago Public High Schools, 134
- Chicago Public Schools: reading and math correlation in, 136–37; social promotion policy changes in, 12; summer program in, 12
- Citizen criticism and local schools, 312
- Classroom teachers: and assessments, 101–2; coaching versus assessing mode, 114
- Clay, Marie, 92
- Cliffs Notes book, 49
- Cody, Wilmer, 273–74
- Cognitive inflation in test frameworks, 131–33
- Cognitive science and accountability systems, 273–74
- Coleman, James, 19
- Commonwealth Accountability Testing System (CATS): administration of, 251–54; adoption of, 250; content standards and test development, 274; evaluation of, 297–99; restandardizing, 256–57, 258–64, 269; school performance under, 258–64; start of, 202; student and public response,

- 250–51; test formats and, 285–86; testing program under, 251–52
- Comprehensive Test of Basic Skills (CTBS-5), 251
- Consolidation and school performance, 8
- Constructed response items: cost of, 287; description, 273; information provided by, 287–88; versus multiple-choice, 285, 288
- Constructive tests for accountability, x–xi
- Constructivism, 42–43, 64–65
- Constructivist methods of teaching, 135
- Content standards: creation of, 275; in Kentucky, 275–76; and length of school day, 276; and test construction, 274
- Cooper, Harris, 12
- Corns, Judge Ray, 245
- CRESST estimate of test costs, 55, 56–59
- Criterion-referenced assessment (CRT), 278–79
- CSTEEP: estimate of test costs, 55, 56; testing survey by, 59
- CTB Bookmark method, 254, 260, 283
- Curriculum narrowing, 45–46
- Curriculum policy, U.S., 145–46
- Curriculum reform and portfolio assessment, 211–13
- Defective assessment examinations, 12–14
- Different Drummers* (Public Agenda), 64
- Disabled reader, characteristics of, 85–88
- Discovery-based learning: defined, 179; implementation and, 179; and Japanese teaching methods, 179–80; theory of, 180
- Distance learning, 18–19
- Dropout rate and high-stakes tests, 38–39
- Early education and the “wait and see” attitude, 83–84
- Early intervention: goal for, 117; impediments to, 83–89; and students with poor reading skills, 83
- Early Literacy Portfolio* (South Brunswick Schools), 93–94
- Educational Economic Policy Center (Texas), 305
- Educational philosophy, conflicting in Kentucky, 277–78
- Educational Testing Service, 230
- Education methods: biased evaluations and, 17–18; evaluating, 16–17; research on, 17; self-evaluation and, 17–18; technology and, 18–19; traditional versus progressive, 297–98
- Education reform: and content standards, 271; and higher-level thinking, 272–74; portfolio assessment and, 197–98; progressive education philosophy and, 292; test construction methods and, 271–72
- “The Effects of High-Stakes Testing on Achievement”* (Koretz, Shepherd, et al.), 43
- Evaluation of the Development and Implementation of KIRIS through December 1994* (KIER report), 266
- External pressure and education reform (Texas), 312–15
- FairTest. *See* National Center for Fair and Open Testing (FairTest)
- Fallout from the Testing Explosion: How 100 Million Standardized Exams Undermine Equity and Excellence in America’s Public Schools* (FairTest), 30
- Federal government and funding of schools, 16
- Florida, dropouts and high-stakes testing in, 38–39
- Getting By* (Public Agenda), 65
- Glaser, Robert, 278
- Governance of schools: citizens and, 14; federal government and, 16; taxes and, 14–16
- Governor’s Business Council (GBC) in Texas, 314
- Griffin, Brian, 38
- Gutterrez, Robert, 277
- Hands-on science investigations, value of in the classroom, 137
- Haney, Walter, 37, 40, 55
- Harmon, Maryellen, 48
- Heidorn, Mark, 38
- Herman, J., 233
- Higher-order thinking: and SAT scores, 49; testing of, 49–50; value of, 50–51
- High School and Beyond (HSB), 53
- High-stakes testing: alleged harms of, 41–42; appraising criticisms of, 41–43; constructivism and, 42–43; cost of, 55–57; external multiple-choice, 42; lower-order thinking in instruction and, 46–47; lower-order thinking in test context and, 47–51; minorities

- and, 53–55; versus no-stakes testing, 43; poorly performing states and, 52; press coverage of, 60; solutions to problems raised by, 61–63; and student achievement, 51–53; studies of, 52–53; supporters of, 60; test score inflation and, 43–45; women and, 53–54
- High-stakes tests and dropout rate, 38–39
- High-Stakes Tests Do Not Improve Student Learning* (FairTest), 51–52
- Hobby, Bill, 306
- Hoke County (North Carolina), 35–36
- Home schooling, rise in, 22
- Hoxby, Caroline, 10
- Illinois Goals Assessment Program (IGAP), 134
- Iowa Test of Basic Skills*, 233, 235
- Instructional practices and portfolio assessment, 212–13
- IQ scores and reading achievement, 84–85, 86
- Isle of Wight studies, 84
- Jaeger, Richard M., 37, 281–82
- Jaeger-Mills method, 260
- Japanese teaching methods: for algebra, 172–74; analysis of, 174–75, 187–88; analysis of Japanese algebra lesson, 174–75; summary, 161–62; videotape of geometry class, 162–174
- Johnson County (North Carolina), 36–37
- Juel, Connie, 82–83
- Kentucky: accountability in, 297–99; assessment program in, 274–75; assessment system and poor academic achievement in, 298–99; changing accountability standards in, 258–60; and claims about student performance, 270; Commonwealth Assessment Test System (CATS), 250–68; conflicting educational philosophies in, 277–78, 297–98; content standards, 275–76; content standards and test creation in, 274; evaluation of portfolio assessment in, 210–11; Kentucky Instructional Results Information System (KIRIS), 56, 202–3; philosophy and educational reform in, 264–65; political and business response to educational quality in, 264; portfolio assessment in, 202–4; portfolio assessment and curriculum changes in, 212–13; reform of public schools in, 245–48; school accountability in, 202–3; school accountability index, 252–56; school performance in, 256–64; test formats and, 284–86; testing in, 56; “thresholds” and schools in, 252–53; use of change scores for schools in, 267–68; use of graders, 282–83
- Kentucky Education Reform Act (KERA): formation of, 202, 245; and changes in Kentucky educational system, 247; and higher-level thinking, 272–74; and instructional philosophy, 276–78; requirement for ungraded primary classes, 277
- Kentucky Instructional Results Information System (KIRIS): cash bonuses for teachers and, 254; and the change to Commonwealth Assessment Test System (CATS), 250–52, 254–56; components of, 202–3; content standards and test development, 274; cost of, 246; description of, 56; design of, 265; Distinguished Educators, 254; effect of HB 53 on, 254–55; evaluating the technical qualities of, 265–66; evaluation of, 297–99; improvement under, 257; legislative changes to (1998), 249–51; matrix sampling and, 292; performance assessment and, 248–49; and portfolio assessment, 293–97; pressure to get system up and running, 265; reliability of, 266–68; teachers and, 254; validity of, 268–69
- Kentucky Instructional Results Information System: A Technical Review*, 266
- Kentucky “thresholds,” 252–53
- KIER Report*, 266
- Koretz, Daniel, 31, 43, 55
- Lake Wobegon Effect, 44–45, 227
- Large-scale assessments: authentic tests and, 13–14; traditional classroom examinations and, 14
- Lerner, Barbara, 39
- Linn, Robert L., 91
- Lower-order thinking: testing of, 50–51; value of, 50–51
- Madaus, George, 48, 55, 57
- Mastery learning, 278

- Matrix sampling, 291–92
- Mehrens, William A., 221–22
- Meisels, Samuel J., 96–98
- Miller, Charles, 305
- Minority students, quality of education and, 54
- Modified Angoff method, 282
- Multiple-choice tests: bias and, 224; versus constructed response items, 287; delimited domain and, 226–27; influence of cognitive psychologists and, 225–26; irrelevant content, 224; item difficulty and, 286–87; Lake Wobegon effects, 227–28; main criticisms of, 223–28; measurement by recognition only, 224–25; minority performance on, 286; versus performance assessment, 232–33, 237–38; procedural knowledge and, 225; reasons for not using, 285–86
- Murray, David W., 54
- National Academy of Education, 31
- National Assessment Governing Board (NAGB), 32, 282
- National Assessment of Educational Progress (NAEP): goal of, 291–92; history of, in U.S., 30–32; Kentucky scores, 247–48; math test and question construction, 130–31; math test and use of language, 132–33; and need for test redesign, 152–53; as performance-based assessment 90; portfolio assessment trial, 205; and school accountability in Kentucky, 248; science test and question construction, 129–30; science test and use of language, 132–33; science test correlation with reading, 133; score reporting of, 282; Texas performance, 34
- National Center for Education Statistics, 31, 139
- National Center for Fair and Open Testing (FairTest): rating of North Carolina testing system, 39; on TAAS program, 33
- National Commission on Excellence in Education, 12
- National Council of Teachers of Mathematics (NCTM): teacher knowledge of, 135; teaching standards of, versus Japanese, 182–83
- National Education Longitudinal Study, 53
- National Research Council, 81
- National Science Foundation (NSF) Systemic Initiative Grant, 137
- A Nation at Risk* (National Commission on Excellence in Education), 12
- Neill, Monte, 33
- “A New Accountability System for Texas Public Schools” (Educational Economic Policy Center), 305
- New Standards Project test, 90
- No Child Left Behind legislation, 117
- Norm-referenced measurement for school comparisons and student assessment, 279–81
- Norm-referenced tests (NRTs), 278–79
- North Carolina: help for poorly performing schools in, 35–36; high-stakes testing in, 37–38; Hoke County, 35–36; Johnson county, 36–37; raising the bar in, 61; testing in, 61–62; value-added rating system in, 35
- Observation Survey test battery, 92
- OEA Report, 266, 268
- OECD. *See* Organization for Economic Cooperation and Development
- Organization for Economic Cooperation and Development (OECD): 1991 survey, 57–58; pioneering value-added indicators and, 6; student progress in members of, 4
- Pearson, P. David, 99–100
- Peltzman, Samuel, 10
- Performance assessment: and accountability, 221–23, 229–30; accountability versus instruction and, 230; costs and, 230; described, 222–23; domains and, 232–33; equating scores and, 236–37; ethnic group differences and, 237; exam security and, 229; factors supporting, 223; in Kentucky, 249; lack of data regarding generalizability of, 234; and large-scale assessments, 288–91; legal issues and, 231; matrix sampling versus every-pupil testing, 229; versus multiple-choice testing, 232–33, 237–38; need for more research on, 237; professional credibility and, 231–32; and public acceptance, 230–31; and scaling of data, 236; subjectivity of scoring and,

- 235–36; threats to availability in, 234–37; and the unit of reporting, 236–37; use for accountability, 222; validity of, 232–34
- Performance Events: elimination of, 290–91; instability of, 291; and the Kentucky school index, 249; trial use of, 289–91
- Performance standards, setting, 281–84
- Perot, Ross, 304
- Phillips, Gary W., 31
- Phonological Awareness and Literacy Screening (PALS)*, 98–99
- Picus, Larry, 55
- Portfolio assessment: as an accountability tool, 206–11; assistance to students and, 293–94, 295; in California, 205–6; characteristics of implemental systems for, 200–201; costs and burdens of, 213–16; as a curriculum reform tool, 211–13; and demands on classroom time, 215; described, 198–200; ethical guidelines and, in Kentucky, 295–96; ethical guidelines and the teaching of writing in, 295–96; evidence about, 201–6; guidelines, 198; and instructional practices, 212–13; in Kentucky, 202–4, 208, 210, 249, 293–97; and large-scale assessments, 293; and the NAEP, 208–9; NAEP trial of, 205; and negative effects in high-stakes context, 213; OEA panel recommendation, 296; operational costs of, 214; in Pittsburgh, PA, 204–5, 209; pressure on teachers and, 294; professional response to, 216; purpose of, 199; reliability of, 207–9, 294; teachers and, 214–15, 216; use of the writing process method in, 296; utility of, 216–17; validity of, 209–10, 293, 294; in Vermont, 201–2, 207–8, 209
- Preventing Reading Difficulties in Young Children* (National Research Council), 81–82
- The Primary Assessment of Language Arts and Mathematics (PALM)*, 95–96
- The Primary Language Record*, 94–95
- Public Agenda national survey (1996): high school student responses, 10; and measuring standards, 11; survey of African-Americans, 54–55; teacher educator responses, 10–11
- Public Agenda poll, 28
- Ranking student performance, in U.S. versus other nations, 4–5
- Reading: and characteristics of disabled readers, 85–88; and early intervention for students with poor skills, 83; impediments to early assessment, 83–89; and oral language skills, 88–89; persistence of early problems with, 82–83, 86; phonological awareness and, 88
- Reading assessment: authentic, 91–92; early, 81–83; false negative rates and, 110–11; false positive rates and, 110–11, 116–17; formal, 89–91; impediments to early, 83–89; multiple-choice formats and, 91; Texas Primary Reading Inventory (TPRI), 102–13; usefulness of early, 116–17
- Reading disabilities: characteristics of 85–88; causes of, 86–87; heritability and, 87
- Reading Recovery program, 92–93
- Reform programs, setting up, 279–80
- “On the Relative Value of Multiple-Choice, Constructed-Response, and Examinee-Selected Items on Two Achievement Tests” (Lukhele, Thissen, and Wainer), 287
- Rose vs. Council for Better Education*, 245–46
- SAT, and impact of courses taken, 53–54
- SAT I: cost and benefit of, 40; customer response to, 41; impact of, 40; use of, in college admissions, 39–40
- Scholastic Assessment Test. *See* SAT I
- School accountability. *See* Accountability
- School assessments and setting cut-points, 256
- School boards and accountability, 8
- School choice: consumer preferences and, 21–22; examinations and surveys and, 20–22; funding for, 22; impact on public schools, 21; Milwaukee (Wisconsin) program, 20–21; scholarships and, 22; surveys about, 21
- School reform: in Texas, 304–6; and the Texas business community, 304–5; wide participation in, 304–5
- Science and math competency: correlation with reading competency, 133–35; and need for test redesign, 152–53
- Science curriculum (California), 138
- Shepard, Lorrie, 43, 45, 46
- Slavin, Robert, 277
- Smith, Mary Lee, 42

- Social promotion: as alternative to standardized testing, 63; changing policies for, 12; failing students and, 11–12; teacher effectiveness and, 11; in Texas, 318–19
- A Splintered Vision* (U.S. TIMSS National Research Center), critique of study behind, 140–46
- Stake, Robert, 58
- Standardized tests, reliability and, 266–67
- Standards and assessment, 114–15
- Standards-based assessment: described, 280; versus norm-referenced measurement, 280–81; setting standards prior to student performance, 283–84;
- Standards for Educational and Psychological Testing (Joint Standards)* (American Psychological Association), 231, 278–79
- Stanford Achievement Test-10, 90
- State NAEP: history of, 32; opposition to, 31; and standards-based reporting of scores, 32
- State standards and performance of U.S. schools, 7
- State testing policies, xi
- Student invention of solutions: claims for, 175–77; goals for, in U.S., 179; illusion of, 189; need for, 185–86; and problem solving, 186. (*See also* Discovery-based learning)
- Student performance: claims about, 270; U.S. versus other nations, 4–5; value-added comparisons, 5–6
- Summer slump, 12
- Systemic change in Texas, 315
- Teacher evaluations, reliability of, 63–64
- Teacher performance: lack of merit incentives and, 11; rare measurement and reward for good, 11
- Teachers, accountability of, 9
- Teachers' unions: membership, 9; performance of, 9–10; success of versus student results, 10
- Teacher training: and academic achievement, 276–77; importance of learning process to, 64
- Teaching methods: analysis of Japanese, 174–75, 187–88; analysis of Japanese algebra class, 174–75; curriculum standards and, 47; Japanese, 161–62; Japanese, for algebra, 172–74; Japanese, for math, 185–89; Japanese versus U.S., 180–81; U.S. reforms and Japanese, 181–82; U.S. versus Japanese, 180–81; videotape of Japanese teacher, 162–74
- Technology: distance learning, 18–19; identifying promising, 18–19
- Test format and level of thinking tested, 48–49
- Test frameworks, cognitive inflation in, 131–33
- Testing: accountability and, 19–20; alleged harms of high-stakes, 41–42; and cognitive inflation, 131–33; concern for children and, 58; constructive uses of, ix–x; high-stakes, 27–28; in Kentucky, 56, 202–3, 250–68, 284–86; level of U.S. versus other countries, 57–58; and need for test redesign, 152–53; in North Carolina, 61–62; opposition to, 27–30; opposition to formats for, 33; organizations opposing, 29–30; versus other options, 62–64; potential of, xii–xiii; professional response to, 27–28; and public policy, vii–viii; public response to, 27–28; purposes of, vii–viii; reasons for opposing, 29; reliability and, 101; misuse of results of, 60–61; surveys of attitudes toward, 58–60; in Vermont, 52
- “Testing America’s Schoolchildren” symposium, xiii
- Testing fairness: and changes in state education agencies, 61; for women and minorities, 53–55
- Tests: not all good, 128–30; productivity of U.S. schools and, 3–4; uses of, 3
- Test score inflation: idea behind, 43; school responses to, 44–45
- Test validity: determining, 99–100; and predictive validity, 100–102
- Texans for Education (TFE), 313
- Texas: accountability in, 307–10; annual campus report cards in, 309; assessment abuses in, 317; challenges in, 318–21; decentralization in, 310–11; education reform in, 304–6; encouraging results in, 316–17; external pressure in, 312; field response to assessment system in, 309–10; leadership in, 306–7; minority student performance in, 317; national test scores in, 303; promotion practices in, 306–7; pupil achievement in, 33;

- school choice in, 316; systemic changes in, 315–16; TAAS program, 32–35; tie between authority and responsibility in, 311
- Texas Assessment of Academic Skills (TAAS): benefits of, 34; case study of, 32–35; lawsuits and, 33; makeup of, 308–9; and NAEP, 34; political support for, 35
- Texas Assessment of Knowledge and Skills (TAKS) Test, 90, 309
- Texas Business and Education Coalition (TBEC), 313
- Texas Primary Reading Inventory (TPRI): administering, 106–7; development and structure, 103–6; and instruction, 114–16; professional development and, 113–14; recent changes to, 107–9; use of, 102; validation of, 109–13; validity and, 101
- Textbooks, U.S. versus other nations, 145–51
- Third International Mathematics and Science Study (TIMSS): Japanese performance on, 161–63; report by U.S. TIMSS National Research Center versus NCSE report, 140–41; reports about, 140–42; unrealistic U.S. views of performance, 139; U.S. performance, 138–39
- Thomas B. Fordham Foundation, 4
- Tierney, Robert J., 101
- TIMSS Repeat (TIMSS-R), 153–54
- TIMSS videotape classroom studies: analysis of Japanese teaching methods, 171–73, 174–75, 187–88; data coding and, 187; Japanese geometry class, 162–64; of Japanese teaching methods, 161–62
- U.S. General Accounting Office (GAO) survey of testing costs, 55–56
- U.S. school districts, number and size of, 8
- U.S. schools, poor performance of: and Board accountability, 8–9; centralized financing of, 7–8; governance and, 14–16; individual teacher performance and, 11; lack of incentives and, 10–11; lack of state standards and, 7; management and, 9–10
- U.S. schools, productivity of, 5
- U.S. TIMSS National Research Center, 140
- The Validity of Gains in Scores on the Kentucky Instructional Results Information System*, 266, 268–69
- Value-added orientation as basis of accountability, 19–20
- Value-added scores, 5–6
- Vellutino, Frank R., 88
- Vermont; mathematics portfolio assessment in, 212; portfolio assessment in, 209; portfolio assessment program in, 201–2; testing in, 52; Uniform Test, 202; Uniform Test scores versus portfolio assessment, 210
- Videotaped classroom teaching, 161–89
- The War on Testing* (Murray), 54
- Whole-language movement, 13
- Woodcock-Johnson achievement test, 90
- The Work Sampling System* (Meisels), 96–98
- Writing process method, in Kentucky, 296