

- Cartledge, G., & Milburn, J. F. (1996). *Cultural diversity and social skills instruction: Understanding ethnic and gender differences*. Champaign, IL: Research Press.
- Caspi, A., & Moffitt, T. E. (1995). The continuity of maladaptive behavior: From description to understanding in the study of antisocial behavior. In D. Chicchetti & D. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Risk, disorder, adaptation* (pp. 472-511). New York: Wiley.
- Cohen, M. K. (1994). *Children on the boundary: The challenge posed by children with conduct disorders*. Alexandria, VA: National Association of State Directors of Special Education, Project FORUM.
- Colvin, G., Sugai, G., & Kameenui, E. (1993). *Proactive school-wide discipline implementation manual*. Eugene: University of Oregon, Behavioral Research and Teaching, College of Education, Project PREPARE.
- Comer, J. P. (1997). *Waiting for a miracle: Why schools can't solve our problems—and how we can*. New York: Dutton.
- Costello, E. J., Angold, A., Burns, B. J., Erkanli, A., Stangl, D. K., & Tweed, D. L. (1996). The Great Smokey Mountains study of youth: Functional impairment and serious emotional disturbance. *Archives of General Psychiatry*, *53*, 1137-1143.
- Cuffe, S. P. (2000, October). *Missouri's approach to the collaboration between mental health and juvenile justice systems*. Paper presented at the annual meeting of the American Academy of Child and Adolescent Psychiatry, New York City.
- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, *54*, 755-764.
- Delquadri, J., Greenwood, C., Whorton, D., Carta, J., & Hall, R. V. (1986). Classwide peer tutoring. *Exceptional Children*, *52*, 535-542.
- Dodge, N., Keenan, S., & Lattanzi, T. (2002). Strengthening the capacity of schools and communities to serve students with serious emotional disturbance. *Journal of Child and Family Studies*, *11*, 23-34.
- Doren, B., Bullis, M., & Benz, M. (1996). Predicting arrest status of adolescents with disabilities in transition. *Journal of Special Education*, *29*, 363-380.
- Dumas, J. E., Prinz, R. J., Smith, E. P., & Laughlin, J. (1999). The Early Alliance prevention trial: An integrated set of interventions to promote competence and reduce risk for conduct disorder, substance abuse, and school failure. *Clinical Child and Family Psychology Review*, *2*, 37-53.
- Dwyer, K., & Osher, D. (2000). *Safeguarding our children: An action guide*. Washington, DC: U.S. Department of Education.
- Eber, L., & Nelson, C. M. (1997). School-based wraparound planning: Integrating services for students with emotional and behavioral needs. *American Journal of Orthopsychiatry*, *67*, 385-395.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, *37*(10), 15-24.
- Edmonds, R. (1986). Characteristics of effective schools. In U. Neisser (Ed.), *The school achievement of minority children* (pp. 93-104). Hillsdale, NJ: Erlbaum.
- Fessler, M. A., Rosenberg, M. S., & Rosenberg, L. A. (1991). Concomitant learning disabilities and learning problems among students with behavioral/emotional disorders. *Behavioral Disorders*, *16*(2), 97-106.

- Figuroa, R. A. (1983). Test bias and Hispanic children. *Journal of Special Education, 17*, 431-440.
- Ford, B. (1992). Multicultural education training for special educators working with African American youth. *Exceptional Children, 59*, 107-114.
- Foster, H. (1990). Ethnocentrism and racism: The disproportionate representation of minorities and poor in special education programs for the emotionally disturbed. *Perceptions, 25*(2), 16-19.
- Fox, N., Leone, P., Rubin, K., Oppenheim, J., Miller, M., & Friedman, K. (1999). *Final report on the linkages to learning program and evaluation at Broad Acres Elementary School*. College Park: University of Maryland.
- Friedman, R. M., Kutash, K., & Duchnowski, A. J. (1996). The population of concern: Defining the issues. In B. Stroul (Ed.), *Children's mental health: Creating systems of care in a changing society* (pp. 69-96). Baltimore: Paul H. Brookes.
- Gay, G. (2000). *Culturally responsive teaching: Theory, research and practice*. New York: Teachers College Press.
- Good, T. L., & Brophy, J. (1986). School effects. In M. C. Wittrock, *Handbook of research on teaching* (2nd ed., pp. 570-602), New York: MacMillan.
- Gunter, P. L., & Denny, R. K. (1998). Trends and issues in research regarding academic instruction of students with emotional and behavioral disorders. *Behavioral Disorders, 24*(1), 44-50.
- Harry, B. (1992). Restructuring the participation of African American parents in special education. *Exceptional Children, 59*, 123-131.
- Hilliard, A. G. (Ed.). (1991). *Testing African American students: Special re-issue of The Negro Educational Review*. Morristown, NJ: Aaron Press.
- Isaacs, M. (1992). Assessing the mental health needs of children and adolescents of color in the juvenile justice system: Overcoming institutionalized perceptions and barriers. In J. Cocozza (Ed.), *Responding to the mental health needs of youth in the juvenile justice system* (pp. 141-164). Seattle: National Coalition for the Mentally Ill in the Criminal Justice System.
- Johnson, G. (1994). An ecological framework for conceptualizing educational risk. *Urban Education, 29*, 34-49.
- Kea, C. (1997). Reconnecting with African-American families. *Reaching Today's Youth, 2*(1), 57-61.
- Kellam, S. G. (1990). Developmental epidemiological framework for family research on depression and aggression. In G. E. Patterson (Ed.), *Depression and aggression: Two facets of family interactions* (pp. 11-48). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kellam, S. G., & Ensminger, M. E. (1980). Theory and method in child psychiatric epidemiology. In F. Earls (Ed.), *Studying children epidemiologically* (pp. 145-180) [International Monograph Series in Psychosocial Epidemiology: Vol. 1]. New York: Neale Watson Academic.
- Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10*, 165-185.
- Kellam, S. G., & Rebok, G. W. (1992). Building developmental and etiological theory through epidemiologically based preventive intervention trials. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 162-195). New York: Guilford Press.

- Koppish, J. E., & Kirst, M. W. (1993). Editor's introduction. *Education and Urban Society*, 25, 123-128.
- Lantieri, L., & Patti, J. (1996). *Waging peace in our schools*. Boston: Beacon Press.
- Larson, K. (1995, April). *Redefining troublemakers: Creating an ethic of care and inclusion to improve outcomes for adolescents with disabilities and other highest risk adolescents*. Paper presented at annual Leadership Conference, Washington, DC.
- Lee, V. L., & Burkam, D. T. (2001, January). *Dropping out of high school: The role of school organization and structure*. Paper prepared for the conference Dropouts in America, Harvard Graduate School of Education, Cambridge, MA.
- Leone, P. E. (1991). *Alcohol and other drug use by adolescents with disabilities*. Reston, VA: Council for Exceptional Children. (ERIC Document Reproduction Service No. ED 340 150)
- Loeber, R., Farrington, D., Stouthamer-Loeber, M., & Van Kammen, W. (1998). Multiple risk factors for multiproblem boys: Co-occurrence of delinquency, substance use, attention deficit, conduct problems, physical aggression, covert behavior, depressed mood, and shy/withdrawn behavior. In R. Jessor (Ed.), *New perspectives on adolescent risk behavior* (pp. 90-149). Cambridge, Eng.: Cambridge University Press.
- Lomotey, K. (1990). An interview with Booker Peek. In K. Lomotey (Ed.), *Going to school: The African American experience* (pp. 14-29). Albany: State University of New York Press.
- Lyons, J. S., Baerger, D. R., Quigley, P., Erlich, J., & Griffin, E. (2001). Mental health service needs of juvenile offenders: A comparison of detention, incarceration, and treatment settings. *Children's Services: Social Policy, Research, and Practice*, 4(2), 69-85.
- Marsteller, F., Brogan, D., & Smith, I. (1997). *The prevalence of substance use disorders among juveniles admitted to regional youth detention centers operated by the Georgia Department of Children and Youth Services*. Washington, DC: U.S. Department of Health and Human Services, Center for Substance Abuse Treatment.
- Metz, M. H. (1997, April). *Keeping students in, gangs out, scores up alienation down, and the copy machine in working order: Pressures that make urban schools in poverty different*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, March, 1997.
- Morrison, G. M., Anthony, S., Storino, M. H., Cheng, J. J., Furlong, M. J., & Morrison, R. L. (2001). School expulsion as a process and an event: Before and after effects on children at risk for school discipline. *New Directions for Youth Development*, 92, 45-72.
- National Research Council. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.
- National Research Council & Institute of Medicine. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: Author.
- Nelson, C. M. (1992). Searching for the meaning in the behavior of antisocial pupils, public school educators, and lawmakers. *School Psychology Review*, 21(1), 35-39.
- Nelson, C. M., Jordan, D., & Rodrigues-Walling, M. (2002). Expand positive learning opportunities and results. *Journal of Child and Family Studies*, 11(1), 13-22.
- Nelson, C. M., & Rutherford, R. B., Jr. (1988). Behavioral interventions with behaviorally disordered students. In M. C. Wang, M. C. Reynolds, & H. J. Walberg (Eds.), *Handbook of special education: Research and practice: Vol. 2. Mildly handicapped conditions* (pp. 125-153). New York: Pergamon.

- Nettles, S. M., Mucherah, W., & Jones, D. S. (2000). Understanding resilience: The role of social resources. *Journal of Education for Students Placed at Risk*, 5(1, 2), 47–60.
- Noguera, P. (1997). Race, class, and the politics of discipline. *Motion Magazine* [On-line serial]. Available: <http://www.inmotionmagazine.com/pedro31.html>
- Office of Special Education Programs. (2000). Unpublished data. Available at [www.ideadata.org](http://www.ideadata.org)
- Ogbu, J. U. (1994). Understanding cultural diversity and learning. *Journal for the Education of the Gifted*, 17, 355–383.
- Osher, D., Dwyer, K., & Jackson, S. (in press). *Safe, supportive, and successful schools step by step*. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services.
- Osher, D., & Osher, T. (1996). The national agenda for children and youth with serious emotional disturbances. In M. Nelson, R. Rutherford, & B. Wolford (Eds.), *Comprehensive collaborative systems that work for troubled youth: A national agenda* (pp. 149–164). Richmond, KY: National Coalition for Juvenile Justice Services.
- Osher, D., & Hanley, T. (1995). The national agenda: Identifying promising practices. *SED Quarterly*, 2(1), 3–6.
- Osher, D., & Hanley, T. V. (1996). Implications of the national agenda to improve results for children and youth with or at risk of serious emotional disturbance. In R. J. Illback & C. M. Nelson (Eds.), *Emerging school-based approaches for children with emotional and behavioral problems: Research and practice in service integration* (pp. 7–36). Binghamton, NY: Haworth Press.
- Osher, D., & Hanley, T. V. (2001). Implementing the SED national agenda: Promising programs and policies for children and youth with emotional and behavioral problems. *Education and Treatment of Children*, 24, 374–403.
- Osher, D., Poduska, J., & Porter, G. (2001). *Multiple risk factors and the need for early intervention among Baltimore children and youth*. Unpublished research memorandum, American Institutes for Research.
- Osher, D., Rouse, J., Woodruff, D., Kendziora, K., & Quinn, M. (2002). *Addressing invisible barriers: Improving outcomes for youth with disabilities in the juvenile justice system*. Washington, DC: U.S. Department of Education.
- Osher, D., Sandler, S., & Nelson, C. (2002). The best approach to safety is to fix schools and support children and staff. *New Directions in Youth Development*, 92, 127–154.
- Quinn, M., Osher, D., Hoffman, C., & Hanley, T. (1998). *Safe, drug-free, and effective schools for ALL students: What works!* Washington, DC: American Institutes for Research, Center for Effective Collaboration and Practice.
- Reid, J. B., & Eddy, J. M. (1997). The prevention of antisocial behavior: Some considerations in the search for effective interventions. In D. M. Stoff, J. Breiling, & J. D. Maser (Eds.), *The handbook of antisocial behavior* (pp. 343–356). New York: John Wiley.
- Reynolds, C. R., Lowe, P. A., & Saenz, A. L. (1999). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 549–595). New York: John Wiley.
- Roberts, R. E., Attkisson, C. C., & Rosenblatt, A. (1998). Prevalence of psychopathology among children and adolescents. *American Journal of Psychiatry*, 155, 715–725.
- Rutherford, R. B., Bullis, M., Anderson, C. W., & Griller-Clark, H. M. (2002). *Youth with disabilities in the corrections system: Prevalence rates and identification issues*. Washington, DC: U.S. Department of Education, Office of Special Education Programs.

- Rutter, M., Giller, H., & Hagell, A. (1998). *Antisocial behavior by young people*. New York: Cambridge University Press.
- Rylance, B. J. (1997). Predictors of high school graduation or dropping out for youths with severe emotional disturbances. *Behavioral Disorders, 23*(1), 5-17.
- Sheppard, V. B., & Benjamin-Coleman, R. (2001). Determinants of service placement patterns for youth with serious emotional and behavioral disturbances. *Community Mental Health Journal, 37*(1), 53-65.
- Sims, A. (1996). *Individual and group characteristics associated with the disproportionate representation of African-American students classified as seriously emotionally disturbed*. Unpublished doctoral dissertation, Howard University, Washington, DC.
- Skiba, R., & Grizzle, K. (1992). Qualifications v. logic and data: Excluding conduct disorders from the SED definition. *School Psychology Review, 21*(1), 23-28.
- Skiba, R. J., Knesting, K., & Bush, L. D. (2002). Culturally competent assessment: More than nonbiased tests. *Journal of Child and Family Studies, 11*(1), 61-78.
- Skiba, R. J., Michael R., Nardo, A., & Peterson, R. (2000). *The color of discipline: Gender and racial disparities in school punishment*. Bloomington: Indiana Education Policy Center.
- Skrtic, T. (1991). The special education paradox: Equity as the way to excellence. *Harvard Education Review, 61*, 148-206.
- Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S., Smith, L., & Dianda, M. (1996). Success for all: A summary of research. *Journal of Education for Students Placed at Risk, 1*(1), 41-76.
- Snyder, H. N., & Sickmund, M. (1999). *Minorities in the juvenile justice system: 1999 national report series, juvenile justice bulletin*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
- Sprague, J., Walker, H., Golly, A., White, K., Myers, D. R., & Shannon, T. (2001). Translating research into effective practice: The effects of a universal staff and student intervention on indicators of discipline and school safety. *Education and Treatment of Children, 24*, 495-511.
- Stein, S., & Merrell, K. W. (1992). Differential perceptions of multidisciplinary team members: Seriously emotionally disturbed vs. socially maladjusted. *Psychology in the Schools, 29*, 320-330.
- Teplin, L. A. (2001, November 12). *Psychiatric disorders among jail detainees: Implications for treatment*. Paper presented at the National Conference on Health Care, Albuquerque, NM.
- U.S. Department of Education. (1994). *National agenda for achieving better results for children and youth with serious emotional disturbance*. Washington, DC: Author.
- U.S. Department of Education. (1999a). *Reducing class size: What do we know?* Washington, DC: Author.
- U.S. Department of Education. (1999b). *21st annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, DC: Author.
- U.S. Department of Education. (2000). *22nd Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. Washington, DC: Author.
- U.S. Department of Education. (2001). *Prevention research & the IDEA discipline provisions: A guide for school administrators*. Washington, DC: Author.
- U.S. Department of Education. (2002). IDEA data tables [On-line]. Available: [www.IDEAdata.org/tables24th](http://www.IDEAdata.org/tables24th)

- U.S. Public Health Service. (1999). *Mental health: A report of the surgeon general*. Washington, DC: Author.
- U.S. Public Health Service. (2000a). *Report on the surgeon general's conference on children's mental health*. Washington, DC: Author.
- U.S. Public Health Service. (2000b). *Youth violence: A report of the surgeon general*. Washington, DC: Author.
- U.S. Public Health Service. (2001). *Mental health: Culture, race, ethnicity. A supplement to the surgeon general's report on mental health*. Washington, DC: Author.
- Valdes, K. A., Williamson, C. L., & Wagner, M. M. (1990). *The national longitudinal transition study of special education students, statistical almanac: Vol. 3. Youth categorized as emotionally disturbed*. Menlo Park, CA: SRI International.
- Wagner, M. (1995). Outcomes for youths with serious emotional disturbance in secondary school and early adulthood. *Future of Children, 5*(4), 90-112.
- Walker, H. M., Colvin, G., & Ramsey, E. (1995). *Antisocial behavior in school: Strategies and best practices*. Pacific Grove, CA: Brooks/Cole.
- Wehby, J. H., Symons, F. J., Canale, J. A., & Go, F. J. (1998). Teaching practices in classrooms for students with emotional and behavioral disorders: Discrepancies between recommendations and observations. *Behavioral Disorders, 24*(1), 51-56.
- Woodruff, D., Osher, D., Hoffman, C., Gruner, A., King, M., Snow, S., & McIntire, J. (1999). The role of education in a system of care: Effectively serving children with emotional or behavioral disorders. In *Systems of care: Promising practices in children's mental health, 1998 series, vol. 3*. Washington, DC: American Institutes for Research, Center for Effective Collaboration and Practice.
- Yeh, M., McCabe, K., Hurlburt, M., Hough, R., Hazen, A., Culver, S., Garland, A., & Landsverk, J. (2002). Referral sources, diagnoses, and service types of youth in public outpatient mental health care: A focus on ethnic minorities. *Journal of Behavioral Health Sciences & Research, 29*(1), 45-60.
- Zabel, R. H., & Nigro, F. A. (1999). Juvenile offenders with behavioral disorders, learning disabilities, and no disabilities: Self-reports of personal, family, and school characteristics. *Behavioral Disorders, 25*(1), 2-40.

## CHAPTER SIX

---

# English-Language Learner Representation in Special Education in California Urban School Districts

---

ALFREDO J. ARTILES<sup>1</sup>

ROBERT RUEDA

JESÚS JOSÉ SALAZAR

IGNACIO HIGAREDA

### INTRODUCTION

Most studies of inequality in special education focus attention on the overrepresentation of African American students, particularly boys, in certain categories of special education. In these studies, Latinos are often said to be “underrepresented” in special education. There is, of course, great social and economic diversity within the Latino community. This study explores the difference in treatment on the basis of language in a large cluster of urban school districts where Latinos predominate. It finds that there is little difference in the early grades, but that students who are classified as limited English proficient (LEP) or English-language learners (ELLs) are seriously overrepresented in two special education categories in the later grades. The study also finds that ELLs with less primary language support are more likely to experience classroom segregation.

Because the number of Spanish-speaking students is growing rapidly and the knowledge base on the above issues is almost nonexistent, basic research is urgently needed. Even without new research, however, the patterns found in this study suggest a new focus for civil rights compliance offices and for staff development work within school districts. While there are clearly questions related to language development that the schools need to address, overassignment of students to special education seems a poor and possibly destructive answer.

Policies and programs that address the needs of English-language learners have been controversial throughout the history of education in the United

States. Although bilingual programs have existed for English-learner immigrants since the latter part of the nineteenth century, societal attitudes toward English learners have shifted over time, depending on economic, political, and historical forces (Baca & Cervantes, 1989). California is a good example of this swinging pendulum. This state has moved from pioneering programs for English learners in the 1970s to the passing of Proposition 227—which greatly restricted native-language support systems in the schools, such as bilingual education—in the late 1990s.

The intersections of English learners and special education are little understood. In most research that compares Latinos and other groups in special education, Latinos are not overrepresented the way that blacks are in certain categories of special education. When the issue of language is added in, however, the results change and a serious civil rights issue emerges. Urban school districts in California that have high minority immigrant student populations provide an important setting for study of this intersection, especially in the face of Proposition 227. Furthermore, in the political tension between bilingual education proponents and the English-only movement lies huge potential for educational benefit—or harm. English learners already experienced poor educational outcomes before any of the recent changes. What will happen to this group as California significantly reduces the language support these students need to enhance their opportunities to learn? Will assumptions about English learners' "difference" based on language become entangled with assumptions about their ability or learning differences, and thus lead to increased levels of special education placement? What will the impact of English immersion classes versus bilingual classes for English learners be in terms of overrepresentation in special education and/or increased isolation from regular classrooms and schools?

As a first effort to tackle some of these complex questions, we set out to obtain a descriptive profile of English-learner placements in special education in several large urban districts in California. We examined placement patterns according to special education programs and by disability categories. We also compared the restrictiveness of special education service (i.e., more or less segregated) for English learners placed in the new language programs that followed Proposition 227, specifically looking at restrictiveness from the regular classroom as language support was reduced. Given that English learners are expected to make a rapid transition to English-only classes, we also examined placement patterns by grade level.

In summary, in eleven urban school districts in California that have high proportions of English learners, high Latino enrollments, and high poverty levels, we found overrepresentation of English learners in special education emerging by grade five and increasingly visible through grade twelve. At the district

level, the English-learner population was overrepresented in the mental retardation (MR) and language and speech impairment (LAS) categories, especially at the secondary level. In terms of odds ratios, English learners were 27 percent more likely than English-proficient students to be placed in special education in elementary grades and almost twice as likely in secondary grades.

Looking at degree of isolation, we found that English learners who were placed in the straight English immersion classes—those with the least home language support—were more likely to be placed in resource specialist classes or the “special day class” option than were English learners in modified English immersion or bilingual classrooms. For example, English learners placed in straight English immersion classrooms were almost three times more likely to be placed in the special education resource specialist program than ELLs from bilingual classrooms.

We based the study on data for the 1998–1999 school year, the first year Proposition 227 was implemented. Although studies based on large databases have limitations, the descriptive profile represents a snapshot of English-learner representation in special education in California at a particular point in time, allowing examination of issues that can get obscured when data are aggregated at the national level. Specifically, most studies of special education placement focus on ethnicity or race, but few focus on English learners. English-learner placement is rarely “unpacked”—that is, variations by disability category, grade level, and type of language support are rarely examined—and few studies focus on older (secondary) students.

The school districts we report on have undergone significant policy changes over the past few years, including policy related to the instruction of English learners. Although it will be important to track these changes over time in a longitudinal fashion, it is also important to document the current situation. We begin by outlining the policy and reform contexts and sketching the evolution of the overrepresentation problem in California.

### Contemporary Policies and Reforms for English Learners in California

Immigrants and other linguistic and ethnic minorities lived in California under increasingly adverse circumstances in the 1990s, as a severe economic recession and an unprecedented immigration wave changed the state’s educational and political landscapes. These trends were followed by passage of Proposition 227 and its predecessors, Propositions 187 and 209.<sup>2</sup>

Proposition 227, which aimed to abolish bilingual education programs, represents the first time citizens were asked to cast their vote to impose an educational curriculum for schoolchildren. Some of the arguments advanced to

justify this proposition included the low efficacy of bilingual programs, as reflected in the persistently poor academic achievement of English learners and their low reclassification rate as fluent English proficient (FEP) (Gándara et al., 2000). Voters approved Proposition 227 in June 1998, even though only one-third of English learners were in bilingual programs at that time, many of them were taught by unqualified teachers, and there was little evidence on the quality of bilingual program implementation (Gándara et al., 2000).

It should be noted that Proposition 227 was approved when school districts were being bombarded with multiple and often contradictory reforms that included class-size reduction, elimination of social promotion, standards and accountability movements, new testing programs, reductionist reading methods reforms, and tougher disciplinary measures. Furthermore, teachers were expected to orchestrate the implementation of these competing reforms simultaneously. At the same time, the state's constructivist-oriented language arts framework had been replaced by a framework that emphasized basic skills, phonics, and direct instruction. These and other changes occurred within a relatively short period of time and in an overlapping fashion, without a great deal of attention to coherence or possible contradictions.

#### Overview of Minority Representation in Special Education in California

Minority overrepresentation in special education has been contested in California for the last thirty years, and litigation has played a major role. *Diana v. State Board of Education* (1973) and *Larry P. v. Riles* (1979) dealt with issues related to assessment bias, disproportionate placement, and the long-term consequences of special education placement. It has been argued that these cases have had a considerable impact on special education legislation (e.g., nondiscriminatory assessment requirements) and evaluations of the overrepresentation problem (Figueroa & Artiles, 1999).

*Diana* (1973) was a class action suit filed by nine Mexican American families before the advent of extensive bilingual education. Chief among the families' complaints was linguistic bias in assessment practices, since English learners were tested with English instruments and placement decisions in mild mental retardation (MMR, then called EMR) programs were based on such test results. The families further argued that these tests and the MMR curricula were oblivious to the students' culture. Another complaint in this suit was the overrepresentation of Mexican American students throughout the state's special education programs. In the settlement, California agreed to test students in their first language as well as in English and to use nonverbal intelligence tests.

In sum, Mexican American students were to be reevaluated using nonverbal intelligence tests, districts were required to report on their reevaluation efforts and on plans for transitioning students from MMR programs to general education, and districts had to monitor whether Mexican American students were overrepresented in MMR (*Diana v. State Board of Education*, 1973).

In contrast to *Diana*, *Larry P.* focused on overrepresentation of African American students in special education and argued that intelligence tests were culturally (rather than linguistically) biased. Coincidentally, this case was tried by the same judge. The California Department of Education refused to settle out of court, and the case went to trial in 1977. Two years later, Judge Peckham ruled that IQ tests were culturally biased against African American students and described special education programs as “dead end,” “inferior,” and “stigmatizing” (MacMillan & Reschly, 1998, p. 17). He also concluded that reliance on biased assessment instruments was closely related to overrepresentation patterns.

Although considerable efforts have been devoted to understanding African American placement in high-incidence disabilities, we know significantly less about ELL representation in these programs. This study presents preliminary evidence about the contexts of English learner overrepresentation in California’s special education programs. The goals of the study are to assess representation of English learners in various disability categories and grade levels, to examine whether English learners in various language programs and grade levels are more likely to be overrepresented and/or more isolated in distinct special education programs.

## METHODS

### Context Characteristics: The School Districts, Populations, and Programs<sup>3</sup>

We used databases from eleven urban districts in California that are currently undergoing major reforms, including in special education. The districts are heavily populated by English learners, particularly of Latino descent. The student ethnic background for the eleven districts during the 1998–1999 school year was as follows: 69 percent Latino/a, 13.6 percent African American, 10.5 percent white, 4.3 percent Asian, 1.9 percent Filipino, 0.4 percent Pacific Islander, and 0.3 percent American Indian/Alaska Native. The majority of the student population came from low-income families as defined by eligibility for free or reduced-price lunch programs—85 percent of elementary and 71 percent of secondary students in the study were eligible for these programs in the 1998–1999 school year.

*English Learner Population* Consistent with national trends, the English learner population has been growing rapidly in these districts. Enrollment data indicate that this population more than doubled between 1981 and 1997. The overwhelming majority is Latino (94% of the elementary population, 91% of the secondary population). The percentages of other groups of English learners for elementary and secondary combined were as follows: American Indian (0.1%), Asian American (3.3%), black (0.3%), white (2.5%), Filipino (0.7%), and Pacific Islander (0.1%). During the 1998–1999 school year, 42 percent of the student population was classified as English learner (53% in elementary and 27% in secondary grades). In the 1996–1997 academic cycle, 19 percent of the population was categorized as fluent English proficient, and 35 percent was designated English only (EO).<sup>4</sup> The redesignation rate from limited to fluent English has been increasing slowly in recent years, though it is still relatively small. For instance, during the four academic years from 1993 to 1997, the percent of English learners who were redesignated grew from 4 percent to 8 percent.

*Special Education Population and Programs* The special education systems in these districts provide services to a host of disability categories in a continuum of educational programs that range from special education schools to services in the general education classroom. The districts together reported that in the 1998–1999 school year 7.2 percent of the student population was receiving special education services (5.4% in K–5 and 9.34% in 6–12), and 7.6 percent of the English learner population was identified as having special educational needs (5.3% in K–5 and 14.06% in 6–12). The largest proportion of special education students was in the specific learning disability (SLD) category—about 2.2 percent in K–5 and 7.5 percent in 6–12.

*Teaching Force* The teaching force was largely white and a sizable proportion was underqualified. In the 1999–2000 academic year, the majority of teachers were white (49%) and female (68%). Latino and African American teachers represented about 40 percent of the teaching staff (25% and 15%, respectively). Almost 25 percent of the teachers had emergency credentials or waivers.

*Language Programs* In response to Proposition 227, new language programs were created; the programs used in elementary grades include straight English immersion, modified English immersion, and primary language instruction (bilingual). District policies stipulate that instruction in straight English immersion programs should be carried out “primarily in English with primary language support provided by the *paraprofessional for clarification purposes.*” The modified English immersion program provides instruction “primarily in English with primary language instructional support provided by a bi-

*lingual authorized teacher for concept development*" (California Department of Education, n.d.). Bilingual education offers core instruction in the primary language with daily (English-language development)<sup>5</sup> instruction. Students transition to English instruction. Placement in bilingual education requires that parents sign an exception waiver.

Language programs in secondary grades include English as a Second Language (ESL) and English-only classes; other language support programs are not available in these grades. English learners with limited English proficiency are placed in classes for students with different proficiency levels. English learners who have not been redesignated as fluent English proficient are placed in English-only classes.

*Restrictiveness in Special Education Services* The resource specialist program (RSP) is a more integrated special education option than the special day class (SDC) program by virtue of the time students spend in the general education classroom. In the SDC program, students spend more than 60 percent of the school day restricted from the regular education classroom. In RSP, students are restricted from the regular education classroom between 20 percent and 60 percent of their time in school. Considering that current language programs provide distinct types of language support for English learners, we examined whether placement in a given language program would affect English learners' odds of being placed in more or less segregated special education programs (SDC or RSP, respectively). We report only elementary grades because little language support is available at the secondary level.

### Data Analysis Procedures

The district databases contain student demographic, achievement, English proficiency, and program placement data. We used the districts' databases for the academic year 1998–1999. These databases are compiled for internal compliance and administrative purposes. Thus, the ideal level of accuracy and precision that would normally be expected for research purposes does not exist. Nevertheless, the databases provide a window into current practice. The special education data were downloaded from each of the eleven districts' centralized databases. Each district was assured of student anonymity.

We focused on the aggregate of districts as the unit of analysis. We conducted descriptive analyses to discern placement patterns for various student categories (e.g., by language proficiency, special education service, disability category, grade level). We targeted disability categories (MR, LAS) that have been historically affected by overrepresentation.<sup>6</sup> We could not include students with SLD because of technical problems with the database. Although

overrepresentation in the ED category has been persistently documented (particularly for African Americans), enrollment data indicated that English learners were rarely placed in this category, thus we did not include it in the study. Two special educational programs (RSP, SDC), as well as two grade levels (elementary, secondary), and three language programs (straight English immersion, modified English immersion, bilingual), were also included in the analyses.

The study objectives focused on disproportionate representation patterns. We defined disproportionate representation as the "extent to which membership in a given group affects the probability of being placed in a specific special education disability category" or service option, such as RSP (adapted from Oswald, Coutinho, Best, & Singh, 1999, p. 198). Consistent with Chinn and Hughes' (1987) definition, "We have reported as disproportionate percentages exceeding plus or minus 10 percent of the percentage expected on the basis on the school-age population (e.g., for blacks in 1978, the percentage of total school enrollment was 15.72 percent. According to the 10 percent criterion a range from 14.15 percent to 17.29 percent would be considered proportionate representation for blacks that year)" (p. 43).

In order to contextualize these analyses, we calculated odds ratios for English learners and English-proficient students in two disability categories (MR, LAS) at the elementary and secondary levels. We also calculated odds ratios for English learners in two special education services (RSP, SDC) in three language programs. The odds ratio is calculated as follows:

The basic element in the index is the "odds" of being assigned to a particular special education category. For example, a measurement of the odds of [an English learner's] being assigned to an [MR] class is the percentage of [English learners] who are classified as [MR] divided by the percentage of [English learners] who are not in special programs. . . . The odds of [an English-proficient] student's being designated [MR] is the percentage of English proficient students classified as MR divided by the percentage of English-proficient students who are not in special programs]. . . . The disproportion index is the ratio of these two odds. (adapted from Finn, 1982, p. 328)

When the odds ratio is equal to one, then English learners and English-proficient students are equally likely to be assigned to a disability category. If the odds ratio is greater than one, then English learners are more likely to be assigned to a disability category, and if the odds ratio is less than one, then English learners are less likely to be assigned to a disability category than are English-proficient students.

## RESULTS

### Unpacking English-Learner Overrepresentation: Language Proficiency, Disability Category, and Grade Level

English learners were disproportionately overrepresented in special education when their percentage exceeded “plus or minus 10% of the percentage expected on the basis on the school-age population” (Chinn & Hughes, 1987, p. 43). For example, 42 percent of students in these districts were English learners. This means the thresholds for English learner under- and overrepresentation would be 37.8 percent and 46.2 percent (i.e.,  $42 - 4.2$  and  $42 + 4.2$ ). Table 1 presents English-learner enrollment in the target districts and by grade level (elementary, secondary). These data suggest that English-learner overrepresentation at the district level is not a problem; the same pattern is observed in elementary grades. However, English learners are overrepresented in secondary grades; in fact, the percent of English learners in secondary special education programs is 11.3 percent above the overrepresentation threshold (see Table 1).

A different configuration of patterns emerges when the data are disaggregated by grade. Table 2 presents English-learner enrollment in general and special education by grade; under- and overrepresentation thresholds are also reported. These data suggest that, although English learners are not overrepresented in grades K–4, the problem emerges in grade five, and it remains clearly visible until grade twelve. The percentage of English learners placed in special education that is above the overrepresentation threshold in grades 6–12 ranges between 19 percent and 26 percent.

Moreover, the proportion of all English learners enrolled in special education increases consistently from kindergarten until grade six from 2 percent to 16 percent, and it varies from grade seven to grade twelve (between 11% and 16%, rising in the junior and senior years; see Figure 1).

A different set of patterns is observed when the data are examined by disability category. At the district level, the English-learner population is overrepresented in the mental retardation and language and speech impairment categories (see Table 3). In contrast, at the district level the English-proficient group is not overrepresented in special education programs or disability categories. Furthermore, although English-learner representation in elementary grades is not a problem, they are noticeably overrepresented in secondary grades in the MR and LAS categories (see Table 3), as well as in the aggregate special education population (see Table 1). The only instance of English-proficient student overrepresentation is observed in the aggregate special education population in elementary grades, where it is only 1.3 percentage points above the overrepresentation threshold.

TABLE 1  
Overrepresentation of English Learners in Special Education at the  
District and Grade Levels, 1998–1999

	<i>ELL Per-centage of Total Student Population</i>	<i>Percentage in Special Education</i>	<i>Under-/Overrepresentation Thresholds<sup>a</sup></i>	<i>Overrep-resented? (Yes/No)</i>	<i>Percentage Points above 10% Over-representation Threshold</i>
<i>All Districts</i>	42	45	37.8–46.2	No	NA
<i>Elementary Grades (K–5)</i>	53	50	47.7–58.3	No	NA
<i>Secondary Grades (6–12)</i>	27	41	24.3–29.7	Yes	11.3

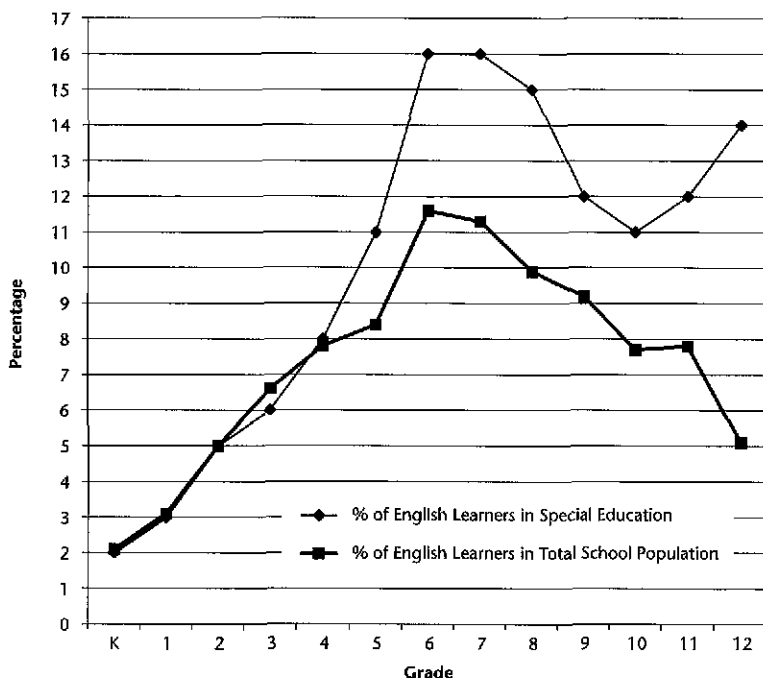
<sup>a</sup> Minus or plus 10% of the percentage expected on the basis of the general education population.

TABLE 2  
Overrepresentation of English Learners in Special Education by Grade,  
1998–1999

<i>Grade</i>	<i>ELL Per-centage of Total Student Population</i>	<i>Percentage in Special Education</i>	<i>Under-/Overrepresentation Thresholds<sup>a</sup></i>	<i>Overrep-resented? (Yes/No)</i>	<i>Percentage Points above 10% Over-representation Threshold</i>
K	62	58	55.8–68.2	No	NA
1	62	56	55.8–68.2	No	NA
2	60	55	54–66	No	NA
3	54	49	48.6–59.4	No	NA
4	47	48	42.3–51.7	No	NA
5	32	43	28.8–35.2	Yes	7.8
6	18	46	16.2–19.8	Yes	26.2
7	17	45	15.3–18.7	Yes	26.3
8	15	43	13.5–16.5	Yes	26.5
9	21	42	18.9–23.1	Yes	18.9
10	14	35	12.6–15.4	Yes	19.6
11	8	28	7.2–8.8	Yes	19.2
12	8	33	7.2–8.8	Yes	24.2

<sup>a</sup> Minus or plus 10% of the percentage expected on the basis of the general education population.

FIGURE 1  
English Learners in Special Education



Compared to their English-proficient counterparts, larger proportions of English learners are placed in each disability category. The percentage of all English learners placed in MR was 0.94, whereas it was 1.87 in LAS. In contrast, the percentage of English-proficient students in MR and LAS is 0.29 and 1.39, respectively.

With regard to placement odds, English learners are more likely to be placed in special education than are English-proficient students (see Table 4); specifically, English learners are 27 percent more likely to be placed in elementary grades and almost twice as likely to be placed in secondary grades. For the MR category, the situation in secondary classes is dramatic, as English learners are more than three times as likely to be placed in this program. English learners are less at risk for placement in elementary LAS classes than their English-proficient counterparts (see Table 4), but are 38 percent more likely to be placed in LAS secondary classes

TABLE 3  
Overrepresentation of English Learners by Disability Categories at the  
District and Grade Levels, 1998-1999

	ELL Percent- age of Total Student Population	Percentage in Mental Retardation (MR)	Percentage in Language and Speech Impairment (LAS)	Overrepre- sented? (Yes/No)	Percentage Points above 10% Overrep- resentation Threshold
<i>Districts</i>	42	55	49	Yes in MR and LAS	MR = 8.8 LAS = 2.8
<i>Elementary Grades (K-5)</i>	53	—	51	No	NA
<i>Secondary Grades (6-12)</i>	27	55	34	Yes in MR and LAS	MR = 25.3 LAS = 4.3

#### RESTRICTIVENESS IN SPECIAL EDUCATION SERVICES BY LANGUAGE PROGRAM IN ELEMENTARY GRADES

We examined whether placement in a given language program would affect English learners' odds of being placed in more (SDC) or fewer (RSP) segregated special education programs. The enrollment in RSP classes is 48 percent English learner (52% English proficient), whereas English learners represent 42 percent of the population in more separate SDC classes (i.e., 58% is English proficient). In turn, 14.5 percent of all elementary English learners are receiving services in RSP and SDC classes (8.7% in RSP and 5.8% in SDC). Of the 8.7 percent of all elementary English learners placed in RSP classes, 4.8 percent is not receiving language support, since the students are placed in English immersion programs, while 2.2 percent is placed in modified English immersion classes and only 1.7 percent in bilingual programs. Similarly, many English learners placed in SDC classes (2.2% of the 5.8%) receive little support, as they are being educated in English immersion classes, whereas 1.7 percent is placed in modified English immersion and 1.9 percent in bilingual classes.

Table 5 suggests that English learners receiving the least language support (i.e., placement in straight English immersion) have a greater chance of being placed in RSP classes. More specifically, elementary English learners in the straight English immersion program are more than twice as likely to receive RSP services than are English learners placed in the modified English immer-

TABLE 4  
 English Learner Special Education Placement Odds<sup>a</sup> (Odds Ratios)  
 by Disability and Grade Level, 1998–1999

	<i>Special Education Placement Odds<sup>b</sup> of ELLs (Compared to English-Proficient Students)</i>	
	<i>Elementary (Grades K–5)</i>	<i>Secondary (Grades 6–12)</i>
Special Education	1.27	1.91
Mental Retardation	Data not available	3.25
Language and Speech	0.94	1.38

<sup>a</sup> The comparison group for the calculation of placement odds is English-proficient students. The category "English proficient" includes children and youngsters classified as English-only students (native English speakers), "redesignated fluent-English proficient" students (former English learners who have been redesignated as English proficient); and IFEP students who have a non-English home language but test as English proficient upon enrollment in school.

<sup>b</sup> If odds ratio = 1, then ELLs and English-proficient students are equally likely to be assigned to a disability category. If odds ratio > 1, then ELLs are more likely to be assigned to a disability category than English-proficient learners, and if odds ratio < 1, then ELLs are less likely to be assigned to a disability category than English-proficient students.

sion model, and almost three times more likely than English learners placed in bilingual programs (see Table 5).

The odds for being placed in SDC classes are higher for English learners in straight English immersion programs when compared with English learners in modified English immersion programs and bilingual education programs (32% and 19%, respectively; see Table 5).

## DISCUSSION

Several interesting trends were identified through a descriptive analysis of placement data in large urban school districts in California. We highlight three patterns observed in the placement data and discuss the implications of the findings.

### Grade Level

A consistent pattern was considerable overrepresentation for English learners in special education in secondary grades. Although problems begin to emerge toward the end of elementary school, significant overrepresentation was observed in grades 6–12, particularly in the last grades of high school. It is not clear whether this pattern is associated with a lack of language support in secondary

TABLE 5  
 Disproportionate Representation of English Learners in Straight English  
 Compared to Bilingual and Modified English Classes, by Restrictiveness of Special  
 Education Placement, 1998–1999

<i>Type of Special Education Service</i>	<i>ELLs in Straight English Immersion Compared to ELLs Assigned to Modified English Immersion</i>	<i>ELLs in Straight English Immersion Compared to ELLs Assigned to Bilingual Education</i>
<i>Resource Specialist Program (RSP)<sup>a</sup></i>	2.26	2.95
<i>Special Day Class<sup>b</sup> (SDC)</i>	1.32	1.19

<sup>a</sup> *Resource Specialist Program*: Services are provided outside the general education classroom for at least 21 percent but not more than 60 percent of the school day.

<sup>b</sup> *Special Day Class*: Services are provided outside the general education classroom for more than 60 percent of the school day.

grades or whether these students entered secondary grades with a preexisting label assigned in elementary school. Let us remember that minority dropout rates are exacerbated in secondary grades as well. In addition to the need for advocacy that will bring more language support for this age group, future research ought to assess the dynamics of special education placement in secondary grades and trace the consequences of a preexisting disability label on English learners' high school careers.

#### Overrepresentation in Disability Categories

Although the district-aggregated data suggested that there is English-learner overrepresentation in the MR and LAS categories, the problem was basically in the secondary grades. It is intriguing that English-learner overrepresentation was found in the LAS category. Unfortunately, we know little about how school psychologists and communication disorders professionals in these districts apply exclusionary linguistic and cultural considerations and eligibility criteria when working with English learners. Future studies must address these trends and should disaggregate student ethnicity, language-proficiency subgroups, and generational differences within the English-learner population to obtain a more complete understanding of how and who gets labeled LAS and the consequences of such labels for student opportunities to learn. Future studies also need to focus on the SLD category, as this is the largest category in the special education field.

### Language Support and Special Education Programs

ELLs who were receiving the least support in their primary language (i.e., straight English immersion programs) had a greater chance of being placed in resource specialist programs and special day class programs than ELLs placed in language programs with greater native-language support. There are at least two important differences between straight English immersion and the other language programs. First, language support is provided by adults with notable training differences (paraprofessionals v. certified bilingual teachers), which may translate into meaningful differences in the nature and quality of language support provided to English learners in these programs. Second, the goals of language assistance vary in important ways. Straight English immersion stresses the use of students' first language (e.g., Spanish) for clarification purposes, whereas modified immersion and bilingual programs promote concept development in the primary language. This important difference in goals might translate into uneven opportunities to develop complex conceptual understanding and ultimately more sophisticated cognitive and language skills.

This finding resonates with other research that found that language support makes a difference in the educational experiences of English learners (Baca & Cervantes, 1989; Finn, 1982; Ortiz, 1997; Portes, 1999; Rumbaut, 1998). Unfortunately, California law demanded that English learners achieve native-like English proficiency in a very short time and with less than optimal primary language support. As we know, such an approach contradicts the evidence about the critical importance of language as a key tool that mediates development and learning (Cole, 1996). In fact, a basic premise of bilingual education is that students' primary language must be developed first so that they can use it to learn a second language (August & Hakuta, 1997). On the other hand, we cannot gauge the potential influence of the multiple reforms that are being implemented in these districts, such as Proposition 227 and reading instruction. Future studies must examine the impact of the confluence of many (often contradictory) reforms on English learners' educational experiences.

Placement in more or less segregated special education settings may also have important consequences for English learners, particularly in terms of access to interactions with nondisabled peers and access to programs that can enrich the social and cultural capital of English learners. Moreover, the law requires that all students with disabilities be educated in the least restrictive environment to the maximum appropriate extent. However, we should not assume that the quality of educational experiences in inclusive settings is always better than what is experienced in RSP classes—indeed, schools that have moved hastily to full inclusion programs with little preparation and support

could create rather negative experiences for both special education and English learners.

Future studies must document the potential interactions between level of program segregation, type of language support, and opportunities to learn. In the districts we studied, the question about misclassification certainly arises in the LAS category. Are school personnel equating a lack of English proficiency with language and speech impairments? What measures are being taken to discriminate between English learners who have special needs and English learners who only need second-language support? Students should be diagnosed as having LAS when adequate instruments are used to document deficiencies in speech and/or language (e.g., articulation, fluency, voice, expressive or receptive language). In short, a lack of English proficiency should not be interpreted as language impairment.

## IMPLICATIONS

### Monitor English-Language Learners

It is imperative that federal agencies and districts develop monitoring systems to track English learners and their subgroups, especially in light of the multiple policy changes and reforms being undertaken. At the federal level, for example, annual reports to Congress should include information on English learners—for the population as a whole and its subgroups. While this information has appeared in some reports in the past, the numbers and issues are now significant enough that it should be included in every report. At the district level, especially when problems of overrepresentation are suspected, monitoring efforts should try to “unpack” important factors such as linguistic, cognitive, academic performance, and socioeconomic data.

### Conduct Theoretically Grounded Research

We argue that an explicit theoretical framework plays a decisive role in defining what becomes relevant or unimportant in the analysis of minority special education representation. Our approach suggests that such analyses should use a multilevel perspective and pay close attention to contextual, historical, political, and cultural forces. In the work presented here, for example, a focus on context and language, allowed us to unpack intriguing placement patterns related to language programs for English learners. This is important, because if we had used gross measures of special education placement for the whole district we would not have seen English-learner overrepresentation. Although we cannot infer causality, we have a better sense of the association between language sup-

port and special education placement. It is an important question, but we do not know whether the patterns reported here were caused by these language policies (e.g., Proposition 227) or by the interaction of the multiple reforms being carried out in the districts.

There is also a need to create comprehensive databases and to coordinate efforts so that researchers have easier access to multiple databases that will allow them to conduct more complex studies. These have been elusive tasks because efforts to create national and state databases tend to exclude students with disabilities (Vanderwood, McGrew, & Ysseldyke, 1998). Therefore, we might run the risk of losing track of English learners if increasing numbers continue to be placed in special education. Another potential obstacle is that existing national datasets do not always allow analyses of sociocultural factors without combining more than one database (e.g., OCR and NCES), a cumbersome task requiring considerable expertise often not available on research teams.

#### Design and Implement a Research Program

We need to design a comprehensive research program that traces not only the dynamics of special education placement patterns, but also their antecedents (e.g., eligibility decision meetings, assessment practices, prereferral/referral interventions, tracking in general education). Funding for prospective studies is needed to document how dynamic sociocultural forces (e.g., opportunity to learn, instructional practices, beliefs) mediate learning in general education as well as referrals to and placement in special education (Artiles, 2000).

As one example of the interaction of these complex variables, Gutierrez and her colleagues (2000) recently reported achievement data from a school district in California that had implemented an English-only literacy program for the past three years with heavy reliance on phonics instruction. They reported that the proportion of Spanish-speaking children scoring at the 50th percentile on the SAT9 decreased from 32 percent in the first grade to 15 percent by the third grade. It is uncertain what sort of outcomes can be expected for English learners in California as Proposition 227 and other reforms are implemented and as their consequences are felt in placements in the special education system.

We believe it is best to study these complex questions with a variety of methods to map out the multiple facets of overrepresentation. Longitudinal studies with mixed quantitative and qualitative components are also needed to assess the origins, evolution, and consequences of the current wave of conservative reforms on minority representation in special education. Similarly, the inclusion of longitudinal data and both short-term and long-term outcomes will greatly enhance the breadth and sophistication of future analyses, particularly

as investigators trace the repercussions of multiple reforms for English-learner representation in special education.

To conclude, as schools become more diverse, it is important to examine how people grapple with notions of difference (Minow, 1990). What are the criteria used to compare these students that lead to judgments about difference? Who defines such norms? What are the views of difference upon which authors of systemic reforms base their positions? What are the bureaucratic and procedural consequences of these assumptions?

These are indeed critical questions educators face as a new wave of English-only initiatives gains momentum. While our data do not allow us to make causal inferences, they suggest that English learners may be adversely impacted by such legislation, as reflected by increased special education placement patterns.

We also suggest, as have others, that English-learner overrepresentation must be tackled beyond the boundaries of special education. Heller, Holtzman, and Messick (1982) conclude that disproportionate representation is a problem "if [children] are unduly exposed to the likelihood of [special education] placement by virtue of receiving poor-quality regular instruction, [and] if the quality and academic relevance of the special instruction programs block students' educational progress, including decreasing the likelihood of their return to the regular classroom" (p. 18). As we examine special education placement evidence twenty years after the publication of Heller et al.'s report, we concur on the imperative to monitor the quality of educational programs offered to English learners in general, bilingual, and special education, as well as the long-term consequences of placement decisions for these students.

## NOTES

1. The first author acknowledges the support of the COMRISE Project at the University of Virginia under grant #H029J60006, awarded by the U.S. Department of Education, Office of Special Education Programs.
2. These two initiatives denied basic educational and health services to illegal immigrants and discontinued affirmative action programs in university admissions and employment, respectively.
3. We tend to aggregate the data to describe the characteristics of the participating school districts to ensure anonymity.
4. In these districts, "English learner" is defined as "limited English proficient students acquiring English and speakers of non-mainstream language forms acquiring mainstream English." "FEP" is defined as "students identified through the formal initial assessment process as having sufficient English academic language proficiency to successfully participate in a mainstream English program." In turn, "English-only [EO] students are identified on the basis of parent responses to the Home Language Survey at the time of

- enrollment. English-only students speak various language forms, including mainstream and non-mainstream forms" (California Department of Education, n.d.).
5. English-language development (ELD) is the "state-designated term for instructional programs to develop listening, speaking, reading and writing skills in English" (California Department of Education, n.d.).
  6. The state definitions of disability categories are available from the California Department of Education.

## REFERENCES

- Artiles, A. J. (2000, July). *The inclusive education movement and minority representation in special education: Trends, paradoxes, and dilemmas*. Keynote address presented at the International Special Education Conference, Manchester, England.
- Artiles, A. J., & Trent, S. C. (2000). Representation of culturally/linguistically diverse students. In C. R. Reynolds & E. Fletcher-Jantzen (Eds.), *Encyclopedia of special education, vol. 1* (2nd ed., pp. 513–517). New York: John Wiley & Sons.
- August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.
- Baca, L. M., & Cervantes, H. T. (1989). *The bilingual special education interface* (2nd ed.). Columbus, OH: Merrill.
- California Department of Education. (n.d.). Unpublished data on file with the authors.
- Chinn, P. C., & Hughes, S. (1987). Representation of minority students in special education classes. *Remedial and Special Education, 8*, 41–46.
- Cole, M. (1996). *Cultural psychology*. Cambridge, MA: Harvard University Press.
- Diana v. State Board of Education, Civil Action No. C-7037RFP (N. D. Cal. Jan. 7, 1970 & June 18, 1973).
- Figueroa, R. A., & Artiles, A. J. (1999). Disproportionate minority placement in special education programs: Old problem, new explanations. In A. Tashakkori & S. H. Ochoa (Eds.), *Education of Hispanics in the U.S.: Politics, policies, and outcomes* (pp. 93–117). New York: AMS Press.
- Finn, J. D. (1982). Patterns in special education placement as revealed by the OCR surveys. In K. A. Heller, W. H. Holtzman, & S. Messick (Eds.), *Placing children in special education: A strategy for equity* (pp. 322–381). Washington, DC: National Academy Press.
- Gándara, P., Maxwell-Jolly, J., García, E., Asato, J., Gutierrez, K., Stritikus, T., & Curry, J. (2000). *The initial impact of Proposition 227 on the instruction of English learners*. Santa Barbara, CA: Linguistic Minority Research Institute.
- Gutierrez, K. D., Asato, J., Santos, M., & Gotanda, N. (2000). *Backlash pedagogy: Language and culture and the politics of reform*. Unpublished manuscript, University of California, Los Angeles.
- Heller, K. A., Holtzman, W. H., & Messick, S. (Eds.). (1982). *Placing children in special education: A strategy for equity*. Washington, DC: National Academy Press.
- Larry P. v. Riles. (1979). C-71-2270, FRP. Dis. Ct.
- Lerner, J. W. (1997). *Learning disabilities: Theories, diagnosis, and teaching strategies* (7th ed.). Boston: Houghton Mifflin.
- MacMillan, D. L., & Reschly, D. J. (1998). Overrepresentation of minority students: The case for greater specificity or reconsideration of the variables examined. *Journal of Special Education, 32*, 15–24.

- Minow, M. (1990). *Making all the difference: Inclusion, exclusion, and American law*. Ithaca, NY: Cornell University Press.
- Ortiz, A. A. (1997). Learning disabilities occurring concomitantly with linguistic differences. *Journal of Learning Disabilities, 30*, 321-332.
- Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. *Journal of Special Education, 32*, 194-206.
- Portes, P. R. (1999). Social and psychological factors in the academic achievement of children of immigrants: A cultural history puzzle. *American Educational Research Journal, 36*, 489-507.
- Reschly, D. J. (1997). *Disproportionate minority representation in general and special education: Patterns, issues, and alternatives*. Des Moines: Iowa Department of Education.
- Rumbaut, R. G. (1998, March). *Transformations: The post-immigrant generation in an age of diversity*. Paper presented at the annual meeting of the Eastern Sociological Society, Philadelphia.
- Vanderwood, M., McGrew, K. S., & Ysseldyke, J. E. (1998). Why we can't say much about students with disabilities during education reform. *Exceptional Children, 64*, 359-370.

## CHAPTER SEVEN

---

# Disability, Race, and High-Stakes Testing of Students<sup>1</sup>

---

JAY P. HEUBERT

### INTRODUCTION

This chapter focuses on tests that have high stakes for individual students. They are “high-stakes” tests because they are used in making decisions about which students will be promoted or retained in grade and which will receive high school diplomas.

Students with disabilities—and the minority students who are often overrepresented in programs for students with disabilities—have a lot to gain or lose from the standards movement and from high-stakes testing in particular. On the one hand, students with disabilities and minority students are often the victims of low expectations and weak instruction, and stand to benefit from efforts to provide high-quality instruction for all students (National Research Council [NRC], 1997).

On the other hand, low expectations and weak instruction increase the risk that students with disabilities will fail high-stakes tests and suffer the well-documented negative consequences associated with being retained in grade or denied standard high school diplomas. As discussed more fully below, even as their pass rates improve in some states, students with disabilities are now failing some state graduation tests at rates as high as 70 percent to 95 percent, and nonpass rates would be even higher if they accounted for students with disabilities who drop out before they have taken graduation tests. Heightened pressure to achieve high pass rates among general education students may also fuel inappropriate referrals to special education (Allington & McGill-Franzen, 1992). Moreover, minority students are often overrepresented among those improperly placed in special education (Individuals with Disabilities Education Act [IDEA], 1997), and there is evidence that states with high minority enrollments in special education are also likely to have high-stakes testing policies.<sup>2</sup>

Thus this study, which focuses generally on high-stakes testing of students with disabilities, is particularly relevant to minority students with disabilities.

This paper argues that if states and school districts use test scores in deciding whether individual students will be promoted or given high school diplomas, they should do so only after students have been taught the kinds of subject matter and skill the tests measure. This position is one with two decades of support in the law (*Debra P. v. Turlington*, 1981) and in the standards of the testing profession (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999; AERA, 2000; NRC, 1999). This paper also reports evidence suggesting that many students, and especially many students with disabilities, are not yet being taught the subject matter and skills they need to meet state standards and pass high-stakes tests.

The objective of the "standards" movement in U.S. public education is to enable all students to attain high levels of academic achievement. In principle, standards-based reform has three key elements: 1) state standards that identify what students should know and be able to do, 2) efforts to align teaching and learning with the state standards, and 3) student assessments, also aligned with the state standards, the results of which can be used to measure student progress and to promote accountability for improved teaching and learning (Elmore, 2000).

Accountability provisions can take many forms. High-stakes testing is designed to hold individual students accountable for their own test performance. *System* accountability measures are those aimed at the providers of education, such as states, school districts, and schools. Federal law, for example, now requires states and school districts (a) to include students with disabilities in large-scale testing programs, with appropriate accommodation and, if necessary, alternative assessment; and (b) to report performance data for students with disabilities, publicly and in disaggregated form (IDEA, 1997; Improving America's Schools Act [IASA], 1994). Under federal legislation enacted in 2002 and effective in 2005, most school districts will have to demonstrate on state assessments that students with disabilities, English-language learners, minority students, and low-socioeconomic-status (SES) students have made adequate yearly progress, and that overall graduation rates are rising (No Child Left Behind Act [NCLBA], 2002).<sup>3</sup> Similarly, some states subject school districts or schools to specific rewards or sanctions based on student performance, and it is now common for schools and school districts to receive favorable or adverse publicity based on student test scores (Goertz & Duffy, 2001). It remains the case, however, that far more states sanction individual students for poor test performance than impose sanctions on individual adults, be they teachers, ad-

ministrators, school board members, legislators, parents, or taxpayers (Goertz & Duffy, 2001).

The section below briefly describes the growth and current scope of graduation testing and promotion testing in the United States. The second section explores current controversies regarding the likely effects of promotion and graduation tests on minority students and on students with disabilities. (As noted throughout this volume, test-score data on minority students in special education are often limited.) The third section describes some important and broadly accepted norms of appropriate test use, which, if observed, would reduce the negative effects of high-stakes testing. The final section describes some elements of a sound testing program.

### THE EXTENT OF HIGH-STAKES TESTING IN THE UNITED STATES

At present, about twenty states require students to pass graduation tests as a condition of getting standard diplomas (Olson, 2001), up from sixteen in 1997 (NRC, 1997) and eighteen in 1998 (NRC, 1999). Of these twenty, more than two-thirds set graduation-test standards at the tenth-grade level or higher (AFT, 1999).

The number of states with exit exams is expected to reach between twenty-six and twenty-nine within the next few years (AFT, 2001; Goertz & Duffy, 2001; NRC, 2001; Shore, Madaus, & Clarke, 2000). Some states, however, facing very high diploma-denial rates, have postponed or are considering postponing the dates by which graduation-test requirements would go into effect; these include Alabama, Alaska, California, Maryland, North Carolina, and Wisconsin (Blair, 2002; Keller, 2001; Olson, 2001). New York has delayed application of its general graduation requirements to students with disabilities, and other states are considering doing so. Graduation testing is thus expanding but its growth has been gradual and somewhat uneven.

In recent years, promotion testing has grown far more rapidly than graduation testing. In response to concerns about "social promotion," a rapidly growing number of states—seventeen in 2001, compared with only six in 1999—require students to pass standardized tests as a condition of grade-to-grade promotion or soon will do so, and thirteen states have both middle school and elementary school promotion-test policies (AFT, 1999, 2001, Table 12). In addition, many school districts, particularly in urban areas, have adopted promotion-test policies even where states have not. For example, New York City has a promotion-test policy although New York State does not, and Boston has a promotion-test policy although Massachusetts does not. This means that large

numbers of the nation's minority students—and increasing numbers of all students—are subject to state or local promotion-test programs.

These high-stakes testing policies plainly apply to students of color. How do they apply to students with disabilities? As noted earlier, federal law requires states and school districts to include students with disabilities in large-scale assessments, and to report their scores publicly, in disaggregated form, as a way of determining how well schools are serving these students. This is a matter of system accountability. Federal law is silent, however, on whether states or schools districts should impose *high-stakes consequences on individual students with disabilities* who fail large-scale tests. In other words, while federal law mandates participation in large-scale tests and public reporting of disaggregated scores, it is for states to decide whether large-scale tests will result in individual high-stakes consequences and, if so, for which students.

States have addressed this question in different ways where students with disabilities are concerned. For example, some states authorize Individualized Education Program (IEP) teams to make individual decisions about whether students with disabilities who do not pass a promotion test may nonetheless advance to the next grade (Quenemoen, Lehr, Thurlow, & Thompson, 2000), or to decide whether students with disabilities who do not pass the state exit exam may nonetheless receive standard diplomas if they meet the requirements of their IEPs (Guy, Shin, Lee, & Thurlow, 1999; Thurlow & Thompson, 1999). Other states require students with disabilities (with appropriate accommodation) to pass promotion tests as a condition of advancing to the next grade (Quenemoen et al., 2000) and/or to pass graduation tests as a condition of receiving standard diplomas (Office of Special Education Programs [OSEP], 2000).

In some states, students with disabilities who fail state exit tests are eligible for alternative diplomas or certificates, such as IEP diplomas, certificates of completion, or certificates of attendance (Guy et al., 1999; Thurlow & Thompson, 1999). Unfortunately, there is little research on the value of such certificates and alternate, nonstandard diplomas in terms of a student's future opportunities for education or employment. The only alternative certificate on which there is extensive research is the General Equivalency Diploma, or GED, and evidence suggests that GED holders are more like high school dropouts in terms of future educational and employment opportunities than they are like individuals who hold standard high school diplomas (NRC, 2001). Indeed, the U.S. Department of Education's Office of Special Education Programs treats GED holders as dropouts rather than as high school graduates (OSEP, 2000, Table AD4), and under the Individuals with Disabilities Education Act (IDEA) a student with disabilities who has not received a *standard* high school diploma

is entitled to special education and related services until the age of twenty-one or twenty-two (IDEA, 1997). States and school districts should therefore think carefully before they decide to award students alternatives to standard diplomas.

### EFFECTS OF HIGH-STAKES TESTING

Many researchers and practitioners believe that standards-based reform will have the greatest impact on students—including many minority students and students with disabilities—who do not now have access to rigorous, high-quality education. There are serious disputes, however, over whether promotion and graduation testing will help such students or hurt them. As discussed below, the story is complex and the evidence incomplete. It seems fair to say, however, that the benefit will be greater, and the harm less, if students are taught the relevant subject matter and skills before they must pass high-stakes tests.

#### Graduation Tests

Even on graduation tests that measure basic skills, minority students and students with disabilities usually fail at higher rates than other students, especially in the years after such tests are first introduced. For example, in the 1970s, when minimum competency tests gained popularity, 20 percent of black students, compared with 2 percent of white students—a discrepancy of ten to one—initially failed Florida's graduation tests and were denied high school diplomas (*Debra P. v. Turlington*, 1979). And while many students with disabilities were excluded from state graduation-test programs (NRC, 1999), those who did participate failed at rates over 50 percent (McLaughlin, 2000).

For a variety of reasons, failure rates typically decline among all groups in the years after a new graduation test is introduced (Linn, 2000). This was true of "minimum competency" graduation tests that many states adopted in the 1970s and 1980s; after a few years; for example, black failure rates in Florida were far lower than 20 percent. It also appears to be true for graduation tests adopted more recently. Texas, for example, which has a graduation test set at the seventh- or eighth-grade level (Schrag, 2000), reports that pass rates of blacks and Latinos roughly doubled between 1994 and 1998, and that the gap in failure rates between whites, blacks, and Latinos narrowed considerably during that time (Viadero, 2000). More recent research, discussed below, questions whether the achievement gap between whites, blacks, and Latinos has actually narrowed in Texas (Klein, Hamilton, McCaffrey, & Stecher, 2000; Linn, 2001). In any case, 1998 data from the Texas graduation tests show continuing disparities: cumulative failure rates of 17.6 percent for black students, 17.4 per-

cent for Hispanic students, and 6.7 percent for white students (Natriello & Pallas, 2001).

Data for students with disabilities are harder to find, but they show a similar pattern: higher pass rates over time accompanied by continuing, disproportionately high failure rates. For example, New York has reported that the number of students with disabilities who passed the state's new Regents English Exam in 1998–1999 was nearly twice as high as the number who took the exam two years earlier (Keller, 2000). While this suggests dramatic improvement, the data can be interpreted in different ways. New York reports the following pass rates for students with disabilities on the Regents English Exam: 5.1 percent in 1997–1998, 6.1 percent in 1998–1999, and 8.0 percent in 1999–2000 (New York Department of Education, 2000, 2001). This represents a 2.9 *percentage point* increase and a 57 *percent* increase over two years in the proportion of students with disabilities earning Regents Diplomas. At the same time, it suggests that high percentages did *not* pass the Regents Exam during these years: 94.9 percent in 1997–1998, 93.9 percent in 1998–1999, and 92.0 percent in 2000. These “nonpass” rates are particularly high, considering that New York calculates them using only students with disabilities who completed high school (New York Department of Education, 2000, 2001).<sup>4</sup> A recent study (Koretz & Hamilton, 2001) confirms highly disproportionate failure rates among students with disabilities in New York and raises concerns about possibly excessive levels of difficulty of the Regents English Exam for some students with disabilities, which the authors believe could cause very high failure rates or undesirable responses by teachers or students, such as excessive coaching. In June 2001, New York decided to extend from 2004 to 2008 a special safety net under which students with disabilities who fail one or more of the new Regents Exams—by 2004 there will be five such exams—may nonetheless receive standard *local* high school diplomas if they pass the older, less rigorous Regents Competency Test for each subject required. In 1999–2000, 54.1 percent of students with disabilities who completed high school that year received such standard local diplomas (New York Department of Education, 2001).

In Massachusetts, the proportion of students with disabilities who passed both state graduation tests in the tenth grade has risen considerably, from 11 percent in 2000 to 29 percent in 2001, and students will have four additional opportunities to pass any test they have failed. At the same time, disproportions remained high in 2001: 71 percent of enrolled tenth-grade students with disabilities had yet to pass both graduation tests, compared with 24 percent of enrolled students without disabilities, and the rates for black students (63% not passing both tests), Hispanic students (71%), and English-language learners (70%) were two to three times higher than the nonpass rates for white students

(23%) and Asian students (32%) (Massachusetts Department of Education, 2001).<sup>5</sup> These statistics are based on total tenth-grade enrollment of students with disabilities. Thus, they do not account for pre-tenth-grade dropout or retention, even though ninth-grade retention apparently increased statewide in the years before 2001. Pass rates would be lower if the statistics took dropouts and retention into account.

Similar gaps between students with and without disabilities can be found in data from other states. In 2001, Alaska's tenth-grade students with disabilities failed different portions of the state graduation test in the following percentages: reading, 78.9 percent (compared with 34.1% for other students); writing, 95.7 percent (compared with 53.4% for other students); and math, 91.1 percent (compared with 56.0% for other students) (Alaska Department of Education, 2001). In 2001, failure rates for Alaska's eleventh-grade students with disabilities showed even higher failure rates. In both years, failure rates for Alaska Natives, blacks, and Hispanics were higher than those for white students. Unfortunately, the state does not post data indicating how many students have passed all three exams, which is what students must do to receive standard diplomas. The statistics just cited, however, do not bode well for students with disabilities. It is perhaps not surprising, therefore, that Alaska has postponed the date at which its graduation requirement goes into effect (Olson, 2001).

In California, where most special education students are minority students (OSEP, 2000),<sup>6</sup> ninth graders had the option in spring 2001 of taking two state exams that they will have to pass to receive standard diplomas in 2004. Only 10.3 percent of students with disabilities passed both tests, compared with 42.2 percent of all students (Wise et al., 2002, Table 5.1, p. 80). The rate at which English-language learners passed both exams was also quite low (11.9%), and the pass rates for black students (22.8%) and Hispanic students (22.8%) were well below those for white students (61.4%) and Asian students (64.5%) (Wise et al., 2002, Table 5.1, p. 80). Moreover, when one includes the students who chose not to take the exams in 2001, only 6.5 percent of all ninth-grade students with disabilities passed both tests, and only 8.1 percent of all ninth-grade English-language learners did so (Wise et al., 2002, p. 81). Students who failed California's exit exam as ninth graders in spring 2001 will have additional opportunities to pass the new state graduation tests.

In states with higher overall pass rates, the performance gaps between students with and without disabilities are smaller but noteworthy and disproportionate. In April 2001, for example, Alabama reported that 3 percent of all seniors had failed the reading test and 4 percent had failed the math test. Comparable figures for students with disabilities in the twelfth grade were 23 percent

and 27 percent, respectively, six to nine times as high as for all Alabama seniors.<sup>7</sup> Moreover, these statistics understate the actual diploma-denial rate for students with disabilities, both because students had to pass *both* tests to receive standard diplomas—which as many as 50 percent of twelfth-grade students with disabilities may not have done—and because it appears that most students with disabilities had dropped out before twelfth grade. Students with disabilities represented only 4.6 percent of twelfth-grade enrollment (Alabama Department of Education, 2001), even though students with disabilities represent a much higher percentage of total enrollments.

It is rare to find test-score data for students with disabilities that have been further disaggregated by race. Where other achievement data have been disaggregated, however, racial disparities within disability categories emerge. For example, David Osher et al. report in this volume that 66 percent of black students with emotional and behavioral disturbance received failing grades, compared with only 38 percent of white students who have this disability. Moreover, as Donald Oswald et al. point out in their chapter, post-high school outcomes for minority students with disabilities are substantially lower than those for white students with disabilities. These studies suggest that it would be valuable to disaggregate test-score data to show the combined effects of disability and race, both in publicly available reports and in test-score data available to researchers.

An important, largely unanswered question concerns the extent to which improved pass rates on graduation tests actually reflect improved teaching and learning. Such improvements are plainly a possible explanation, and the most desirable one. During the 1980s, however, when many states reported sharply improved pass rates on graduation tests, scores on the National Assessment of Educational Progress (NAEP)—a highly regarded nationally administered examination—showed little or no improvement in student learning. Indeed, evidence that minimum competency tests were not producing improved student performance on the NAEP is one reason why the current standards movement emphasizes higher standards, and why some states have been raising graduation-test standards.

More recent fourth- and eighth-grade NAEP scores suggest improvements in student mathematics performance during the period 1990–1996, particularly in some states (including Texas and North Carolina) that pursued certain “systemic reform policies” (Grissmer, Flanigan, Kawata, & Williamson, 2000, p. 58).<sup>8</sup> At the same time, NAEP scores consistently show much less gain in student performance than do the state-test results, and NAEP scores also suggest a widening racial achievement gap among thirteen- and seventeen-year-olds (National Center for Education Statistics [NCES], 2001, pp. 22–23). For example,

as noted above, data from the Texas graduation test, the Texas Assessment of Academic Skills (TAAS), suggest that the achievement gap between white and black students and between white and Latino students closed dramatically between 1994 and 1998. More recent research using NAEP data indicates, however, that the achievement gap between white students and other groups in Texas actually *increased* slightly during this period (Klein et al., 2000). For Robert Linn (2001), this evidence "raises serious questions about the trustworthiness of the TAAS result for making inferences about improvements in achievement in Texas or about the relative size of the gains for different segments of the student population" (p. 28). It also raises questions about the factual basis of the decision in *GI Forum v. Texas Education Agency* (2000), in which a federal judge relied heavily on evidence of a narrowing racial achievement gap on the TAAS in upholding the legality of the Texas graduation test. Moreover, as Daniel Losen and Kevin Welner discuss in this volume, the low TAAS participation rates of students with disabilities, most of whom are minority, also suggest that evidence before the court understated the racial achievement gap.

Unfortunately, NAEP does not yet include enough students with disabilities (or English-language learners) in its samples to provide meaningful *state-level* performance scores for these groups (NCES, 2001). Time will tell whether future state NAEP results for students with disabilities confirm the state-test gains that some states have reported for students with disabilities.

What factors other than improved achievement may explain increased pass rates on state tests? First, it is well known that scores on a test can increase as students become familiar with that test's format, "with or without real improvement in the broader achievement constructs that tests and assessments are intended to measure" (Linn, 2000, p. 4). Studies show that improvements on a state's tests may not be confirmed when students take other tests that supposedly measure the same knowledge and skills (Koretz & Barron, 1998; Koretz, Linn, Dunbar, & Shepard, 1991). When teachers "teach to the test," for example, student scores typically rise as students become familiar with particular item formats, whether or not they actually know more about the subjects being tested (Madaus & Clarke, 2001; Mehrens, 1998).

Second, some states may reduce high failure rates, actual or projected, by making the state graduation tests easier or by setting lower cutoff scores that students must achieve to pass. In New York, for example, failure rates on a state test dropped substantially after the state created a temporary "low-pass" category for students who were below the state's original passing score. Similarly, increased pass rates in Texas may be due in part to changes in the test that made it easier for students to pass (Schrag, 2000).

Third, if low-achieving students are not part of the test-taking population, then the pass rates of those who remain will be higher—even if the achievement of those who actually take the test has not improved. Studying an administration of New York's new Regents Exam, for example, Koretz and Hamilton (2001) found that only about 6 percent of actual test takers were students with disabilities, even though students with disabilities represented about 12 percent of the relevant student population; given significantly lower pass rates among students with disabilities, the absence of half of these students would produce increased pass rates for those who did take the test. Sometimes low-performing students with disabilities are encouraged to postpone taking the test until they are likelier to pass;<sup>9</sup> this may be legitimate, but it does boost the pass rate for those who do take the test.

Special education placement practices can also distort pass-rate information. If low-performing general education students are improperly placed in special education, some states or school districts may not count these students' scores in calculating pass rates for general education students, and the pass rates for those who remain in general education will be inflated artificially (Allington & McGill-Franzen, 1992).

Similarly, if a graduation test is administered in tenth grade and large numbers of low achievers were retained in ninth grade the year before, those retained will not be part of the test population and the pass rate for those who were promoted to tenth grade will be higher. Such retention is not uncommon. It is well documented, for example, that ninth-grade retention in Texas has increased dramatically since the mid-1980s (Murnane & Levy, 2001). In Massachusetts, improved pass rates among tenth graders in 2001 followed increased ninth-grade retention in previous years. In California, a 2001 survey of educators indicates that 55 percent of principals and 32 percent of teachers anticipate that the state's new graduation test will have "a strongly negative or negative impact on student retention rates" (Wise et al., 2002, p. 45). As noted above, minority students and English-language learners are often disproportionately represented among those retained in grade. There is also evidence that low-achieving students with disabilities have sometimes been retained in grade when the alternative would have been for them to take state tests (Thurlow & Johnson, 2000).

Last but not least, if low-performing students have dropped out of school before taking a graduation exam, the pass rates will be higher for those who remain in school. There is considerable debate about whether graduation testing causes increased dropout rates. Walt Haney (2001) offers evidence that the Texas graduation test has led to significantly increased dropout rates, especially for minority students. Other scholars (Carnoy, Loeb, & Smith, 2001), while agreeing that ninth-grade retention in Texas has increased dramatically since

the mid-1980s (Murnane & Levy, 2001), dispute claims that graduation testing is the cause. Brian Jacob (2001), using a national longitudinal database, finds no general relationship between graduation testing and dropping out but concludes that such tests do increase the probability of dropping out among the lowest achieving students. A 2001 survey of educators in California suggests that 80 percent of principals and 61 percent of teachers believe that the state's new graduation test will have "a strongly negative or negative impact on student dropout rates" (Wise et al., 2002, p. 45).

On the one hand, it appears that many low achievers start to disengage from school well before graduation tests loom (NRC, 2001). On the other hand, failing a graduation test can increase the likelihood that low achievers will leave school (Clarke, Haney, & Madaus, 2000). Also, the current climate of accountability places new pressures on schools to increase student pass rates, which in turn can lead to increased and/or understated dropout rates (Schrag, 2000); for example, an education research group concluded in May 2001 that the Texas dropout rate is more than twice that reported by the state education department (Benton, 2001). Unfortunately, this critical issue is complicated by a lack of uniformity among the states in defining and counting dropouts (NRC, 2001); Texas counts GED holders as high school graduates, for example, while the U.S. Department of Education counts such individuals as high school dropouts (OSEP, 2000, Table AD4). Under the 2002 No Child Left Behind Act, school districts will soon be required to show improved high school graduation rates based on the size of the entering class, a change that should increase uniformity in the counting of dropouts and the calculation of dropout rates.

In sum, reported graduation-test pass rates should be viewed in the context of such factors as (a) improper exemptions, exclusions, or absences of students with disabilities or English-language learners from the test-taking population, which are far higher in some states than in others (Citizens' Commission on Civil Rights, 2001; Robelin, 2001); (b) improper special education placements; (c) grade retention in the years prior to high-stakes testing; (d) dropout rates and the formulas by which they are computed; and (e) improper testing accommodations that may artificially inflate some students' scores (Allington, 2000; Sack, 2000).

### Promotion Tests

As noted above, promotion testing has been growing in the elementary and secondary grades (AFT, 2001), and especially in urban school districts. In Chicago, New York, and other cities, tens of thousands of students, the vast majority of them minority students, have been retained in grade. And while the application of such policies to students with disabilities varies, as previous dis-

cussion indicates, there are states and school districts in which students with disabilities who fail promotion tests are subject to retention in grade (Quemmoen et al., 2000).

How well do students with disabilities fare on such tests? Two multistate studies conducted by the National Center for Educational Outcomes provide evidence that students with disabilities are much likelier than nondisabled students to fail state achievement tests. One such report (Ysseldyke et al., 1998) examines tests that twelve states administered during 1995–1996 and 1996–1997 in grades six through eleven. It shows that students with disabilities typically failed these tests at rates thirty-five to forty percentage points higher than those for “all” students, and the gap would have been even higher had students with disabilities been compared with nondisabled students rather than with “all” students.<sup>10</sup> Similar gaps are evident in a study of pass rates on tests that seventeen states administered, mostly in grades eight through twelve, in two subsequent years, in 1997–1998 and 1998–1999. This report (Thurlow, Nelson, Teelucksingh, & Ysseldyke, 2000, Tables 4–9, 12) also shows large performance gaps between students with disabilities and “all” students: 23 to 47 percentage points in reading, 19 to 42 percentage points in math, and 25 to 44 percentage points in writing, all of which would have been even higher had the comparison been between students with and without disabilities. While not focusing specifically on tests used for promotion, both these reports provide strong evidence that the overall achievement gap is large at both the elementary and secondary levels. It is possible that some students with disabilities may be more highly motivated to pass promotion tests than they are on other state tests (Roderick & Engel, 2001). On the other hand, as discussed above, students with disabilities fail *graduation* tests at highly disproportionate rates. Overall, therefore, these studies suggest that students with disabilities fail promotion tests (and other state achievement tests) at substantially higher rates than nondisabled students.

If students with disabilities and minority students who fail promotion tests are retained in grade, they are at substantially increased risk of dropping out. Students retained in grade even once are much likelier to drop out later than are students not retained, and the effects are even greater for students retained more than once (Hauser, 2001; NRC, 1999; Shepard & Smith, 1989). “[T]here is no dispute that retention in grade is a very strong predictor of who will drop out” (NRC, 2001, p. 41), and some scholars (Lillard & DeCicca, 2001) have concluded that retention is the single strongest predictor of which students will drop out of school.

Promotion testing is thus likely to increase, perhaps significantly, the numbers of students with disabilities and minority students who suffer the serious

consequences of dropping out. These consequences include much lower average earnings and substantially reduced opportunities for employment and further education. Congress has already expressed serious concern about the disproportionately high dropout rates of students with disabilities (IDEA, 1997).

Given the relationships between promotion testing, retention in grade, and increased dropout rates, the National Research Council (1999) has described simple retention in grade as "an ineffective intervention" (p. 285). There is thus good reason to question the value of promotion-test policies, even as such policies proliferate.

Promotion and graduation testing may also have unintended consequences for teachers. As noted above, high-stakes testing is intended to raise teacher motivation and effectiveness, and there is evidence that with appropriate professional development, support, resources, and time, teaching effectiveness can improve significantly (Elmore, 2000). There is already evidence, however, that the negative publicity associated with poor test scores can lead experienced teachers to leave urban schools for the suburbs (Lee, 1998). Such trends exacerbate a nationwide teacher shortage that is already most acute in urban schools and that is at least as serious for special education teachers (Gonzalez & Carlson, 2001) as for teachers in general education. Unfortunately, efforts to improve low-performing schools—and to educate all children effectively—will be undermined if those schools lose strong teachers.

On the other hand, testing policies that lead to improved teaching and learning are likely to benefit minority students, English-language learners, and students with disabilities even more than they do other students. New York Education Commissioner Richard Mills defends stringent graduation-test requirements partly because he hopes they will bring an end to low-track classes, in which students—most of them black students, Hispanic students, and/or English-language learners—typically receive poor-quality, low-level instruction. This position is grounded in solid evidence that placement in typical low-track classes is educationally harmful for students (NRC, 1999; Oakes, Gamoran, & Page, 1992), and that students will learn more if they are placed in more demanding classes (NRC, 1999; Weckstein, 1999).

Advocates for minority children and low-SES children hope that high standards will provide the political and legal leverage needed to improve resources and school effectiveness so that all children receive—*beforehand*—the high-quality instruction they need to be able to meet demanding academic standards. Disability-rights groups likewise hope that high standards will provide the political and legal leverage needed to improve resources and school effectiveness so that students with disabilities get the help they need *in time to meet*

*demanding academic standards.* They count on state standards and tests to drive improvements in IEPs so that IEPs reflect more of the knowledge and skills that all students are expected to acquire (NRC, 1997). There is certainly evidence that higher expectations and improved instruction lead to improved achievement (Elmore, 2000; IDEA, 1997; Thurlow & Johnson, 2000).

#### STANDARDS OF APPROPRIATE TEST USE

Whether high-stakes testing helps or hurts depends largely on whether such tests are used to promote high-quality education for all children—the stated objective of standards-based reform—or to penalize students for not having the subject matter and skills that they have not been taught in school.

This is the principal theme that former U.S. Education Secretary Richard Riley, a strong proponent of standards-based reform, emphasized in his February 22, 2000, State of American Education address. Riley called for a “mid-course review” of the standards movement, a step he said was needed “because there is a gap between what we know we should be doing and what we are doing” (Riley, 2000, p. 6).

The sections that follow focus chiefly on two issues of appropriate test use: the principle that promotion tests and graduation tests should measure only the knowledge and skills that schools have afforded students the opportunity to acquire and the principle that high-stakes decisions should be based on multiple measures of student achievement, rather than on a single test score.

#### Teaching Students the Necessary Subject Matter and Skills before Using Test Results to Make High-Stakes Decisions about Individual Students

In former secretary Riley’s call for a “midcourse review,” he said that state standards should be “challenging but realistic. . . . [Y]ou have to help students and teachers prepare for these [high-stakes] tests—they need the preparation time and resources to succeed, and the test must be on matters that they have been taught” (Riley, 2000, p. 7).

Not coincidentally, these concerns are also reflected in norms of appropriate test use that the testing profession, the National Research Council, and American Educational Research Association (AERA) have articulated. The *Standards for Educational and Psychological Testing*, issued in December 1999 by the AERA, the APA, and the NCME (and referred to here as the *Joint Standards*), assert that promotion and graduation tests should cover only the “content and skills that students have had an opportunity to learn” (AERA et al.,

1999, Standard 13.5, p. 146). The congressionally mandated NRC study, *High Stakes: Testing for Tracking, Promotion, and Graduation*, reached a similar conclusion in 1999: "Tests should be used for high-stakes decisions . . . only after schools have implemented changes in teaching and curriculum that ensure that students have been taught the knowledge and skills on which they will be tested" (NRC, 1999). So does the AERA, which, in its July 2000 *Position Statement Concerning High-Stakes Testing in Pre-K-12 Education*, recommends the following "condition[] essential to sound implementation of high-stakes educational testing programs. . . . When content standards and associated tests are introduced as a reform to . . . improve current practice, opportunities to access appropriate materials and retraining consistent with the intended changes should be provided before . . . students are sanctioned for failing to meet the new standards" (p. 2).

Moreover, a committee of the National Research Council expressly recommended that this principle be applied individually to each student with disabilities:

If a student with disabilities is subject to an assessment used for promotion or graduation decisions, the IEP team should ensure that the curriculum and instruction received by the student through the individual education program is aligned with test content and that the student has had adequate opportunity to learn the material covered by the test. (NRC, 1999, p. 295)

Are students being taught or given "adequate opportunity to learn" the requisite subject matter and skills *before* individual high-stakes consequences such as grade retention and diploma denial take effect? Researchers, advocacy organizations, educators, and federal and state governments are trying in different ways to answer this question, for students generally and/or for students with disabilities. Some focus on indicators of student achievement, such as test scores, on the assumption that "*the best evidence that a school system is providing its students adequate opportunity to learn the required material is whether most students do, in fact, learn the material*" (Wise et al., 2002, p. 93, emphasis in original). Others (Citizens' Commission on Civil Rights, 2001; Cohen, 2001) are looking at whether states and school districts have met system accountability standards that are intended to gauge how well schools are serving different groups of students. Some (Porter & Smithson, 2000, 2001) are conducting surveys that ask teachers and administrators how much alignment they see between standards, curriculum, instruction, and tests, and developing techniques for expressing the amount of alignment. Others are examining written documents—state standards, the state curriculum, the tests that are administered,

the actual lesson plans from which teachers teach—to determine how well they are all aligned.

By these measures, there is evidence of progress. As discussed earlier, increasing proportions of students—all students, minority students, students with disabilities—appear to be passing state tests over time.<sup>11</sup> More states meet current federal system accountability requirements than did so two years ago (Citizens' Commission on Civil Rights, 2001; Cohen, 2001; Robelin, 2001).

At the same time, there are plainly many students who are not yet being taught the subject matter and skills that state standards reflect and that students need if they are to pass state tests. Several different types of evidence support this conclusion.

One kind of evidence consists of recent graduation–test score data showing failure rates of 60 percent to 90 percent for students with disabilities, minority students, and English-language learners. If “all children can learn,” as the standards movement and at least three federal statutes assert (IASA, 1994; IDEA, 1997; NCLBA, 2002), these failure rates must be due at least in part to poor-quality instruction for the groups whose failure rates are so high.<sup>12</sup>

Other kinds of evidence tend to reinforce the view that many educational systems are not yet at the point where they offer all students instruction that enables them to meet state standards. For example, many states do not yet meet federal system accountability standards that require them to include all their students in large-scale assessment programs and to report disaggregated scores for students with disabilities, English-language learners, and different racial groups (Citizens' Commission on Civil Rights, 2001; Cohen, 2001; Goertz & Duffy, 2001; Robelin, 2001; Thompson & Thurlow, 2001). States not meeting these standards lack basic information without which it is difficult even to *know* how well low-achieving groups are performing, much less to improve instruction so as to *address* any problems that the data might reveal. In other words, some important preconditions to systemic improvement, designed to identify and help address the needs of low achievers, have yet to be met in a number of states.

Studies call particular attention to the need for improved standards-based education for students with disabilities. For example, Don Dailey, Kathy Zantal-Wiener, and Virginia Roach (2000), in a three-state, OSEP-funded study of standards-based reform and students with disabilities, found that special education teachers “lacked guidance about how to align IEPs with the standards,” that they were “by and large . . . not involved in school-wide discussions about standards,” that special education teachers “tended to use the IEPs rather than the standards as a guide for instruction,” and that “most IEPs were not

aligned with the standards” (pp. 8–9). They also found that many special education and general education teachers did not know how to link pedagogy, standards, and content, “lacked the knowledge and skills to co-teach in a classroom,” and “tended to have a ‘wait and see’ attitude about exposing students with disabilities to and engaging them in standards-based instruction” (pp. 8–9). The authors did not identify which three states they studied, and it is therefore unclear whether these states administer high-stakes graduation or promotion tests. It is also unclear how generalizable these findings are. In these states and many others like them, however, it does not appear that IEP teams “ensure that the curriculum and instruction received by the student through the individual education program is aligned with test content and that the student has had adequate opportunity to learn the material covered by the test” (NRC, 1999, p. 295). This is also a concern for minority students, who are overrepresented among students with disabilities.

Similarly, while there are not many published empirical studies that explore actual alignment within states between standards, assessments, curriculum, and instruction, research indicates that there remain discrepancies between what high-stakes tests measure and what students have been taught. Preliminary results of a ten-state study by Andrew Porter and John Smithson (2000) suggest that there is little overlap between a state’s standards and what fourth- and eighth-grade teachers in the state say they teach students. The overlap teachers reported between state tests and instruction ranged from a low of 5 percent to a high of 46 percent, depending on the subject, grade level, and state (Boser, 2000; Porter & Smithson, 2000). However, these results are preliminary, the teacher samples are small, and the study is limited to the fourth and eighth grades. More recent studies by Porter, Smithson, and their colleagues offer a few more examples of the overlap between teaching and tests within particular states; while it is not clear how representative the examples are, the overlap is small in each case.<sup>13</sup>

All these statistics and findings have their limitations, and research suggests that alignment will increase as teachers increasingly focus instruction on the subjects that state tests measure (Madaus & Clarke, 2001). Taken together, however, they suggest that many teachers are not yet teaching students the full range of subject matter and skills that state tests measure, and that the gap is probably greatest for students with disabilities, minority students, and English-language learners. Where this is the case, it would be inappropriate to use results of these tests in making promotion or graduation decisions for individual students. It seems problematic, therefore, that so many states and school districts are moving forward with high-stakes graduation and/or promotion tests.

### Using Multiple Measures to Make High-Stakes Decisions about Individual Students

As noted above, increasing numbers of states and school districts automatically deny grade promotion or high school diplomas to students who fail state or local tests, regardless of how well the students have performed on other measures of achievement, such as course grades. Former secretary Riley is not alone in believing that states and school districts should “incorporate multiple ways of measuring learning” (Riley, 2000, p. 6), particularly in making high-stakes decisions about promotion and graduation.

The National Research Council (1999) emphasizes that educators should always buttress test-score information with “other relevant information about the student’s knowledge and skills, such as grades, teacher recommendations, and extenuating circumstances” (p. 279) when making high-stakes decisions about individual students. This is consistent with the testing profession’s *Joint Standards*, which state that “in elementary or secondary education, a decision or characterization that will have a major impact on a test taker should not automatically be made on the basis of a single test score. Other relevant information . . . should be taken into account if it will enhance the overall validity of the decision” (AERA et al., 1999, Standard 13.7, p. 146). Similarly, the AERA *Position Statement* (2000) provides that “[d]ecisions that affect individual students’ life chances or educational opportunities should not be made on the basis of test scores alone. Other relevant information should be taken into account to enhance the overall validity of such decisions” (p. 2).

Why is it so important to use multiple measures in making such critical decisions about individuals? One reason is that decisions based on grades may have less disproportionate racial impact than test scores; this is the conclusion of a recent study examining student grades and scores on the Massachusetts graduation exam (Brennan, Kim, Wenz-Gross, & Siperstein, 2001).

More broadly, the answer is that *any* single measure is inevitably imprecise and limited as to the information it provides. Proponents of high-stakes testing sometimes point out problems often associated with exclusive reliance on student grades in making promotion and graduation decisions: that there has been grade inflation during the last three decades, for example, and that there is variation among teachers, schools, and school districts in what particular grades mean.

The evidence on K–12 grade inflation is less clear than many people seem to assume. Daniel Koretz and Mark Berends (2001), using national databases to explore possible math grade inflation between 1982 and 1992, concluded that high school math grades increased slightly during this period but that grades ac-

tually *declined* slightly after taking into account modest improvements in math achievement during this time. Even assuming that there are valid concerns about grades, however, it does not follow that grades should be ignored altogether.

Standardized tests, like grades, are limited in what they measure. It is well known, for example, that standardized-test scores are no better than high school grades in predicting first-year college achievement, and that grades and test scores together provide a better prediction of freshman grades than either measure alone. Grade-point averages are also better indicators than standardized tests of student motivation over time, a factor strongly related to later success in school and the workplace. Moreover, as the following examples illustrate, even the best standardized tests are typically less precise than most people think:

- What are the chances that two students with identical “real achievement” will score more than ten percentile points apart on the same Stanford 9 test? For two ninth graders who are really at the 45th percentile in math, the answer is 57 percent of the time. In fourth-grade reading, the probability is 42 percent.
- How often will a student who really belongs at the 50th percentile according to national test norms actually score within 5 percentile points of that ranking on a test? The answer is only about 30 percent of the time in mathematics and 42 percent in reading. (Rogosa, 1999, cited in Viadero, 1999, p. 3)

Unfortunately, as former secretary Riley noted, “There is a gap between what we know we should be doing and what we are doing” (2000, p. 7). This is the case in the many states and school districts that make promotion or graduation decisions relying solely on student test scores. Such practices, though widespread, do not seem consistent with norms of appropriate test use.

To complicate matters, there is at present no satisfactory mechanism for ensuring that states and school districts respect even widely accepted norms of appropriate, nondiscriminatory test use. The two existing mechanisms—professional discipline through the associations that produce the *Joint Standards* or legal enforcement through the courts or administrative agencies—have complementary shortcomings. Professional associations such as the AERA, APA, and NCME have detailed standards but lack mechanisms for monitoring or enforcing compliance with those standards. For courts and federal civil rights agencies, the reverse is true; they have complaint procedures and enforcement power, but lack specific, legally enforceable standards on the appropriate use of

high-stakes tests. Recognizing the problem, the U.S. Department of Education's Office for Civil Rights (OCR, 2000) has released a carefully crafted resource guide that, while not legally binding, aims to promote the appropriate use of tests used in promotion and graduation decisions. In 2001, the new Bush administration "embargoed" this resource guide, leaving its status uncertain; it remains available, however, and is helpful on a wide range of issues associated with promotion and graduation testing, assessment of students with disabilities and English-language learners, and other civil rights issues.

### ELEMENTS OF A SOUND TESTING POLICY

Given these concerns, what are some elements of a sound high-stakes testing policy within the larger context of standards-based reform? First, states should adopt standards for what students should know and be able to do. And while such standards continually evolve, this is something virtually all the states have done (AFT, 2001). Second, policymakers and educators should strive to align each of the following with state standards: (a) state and local large-scale assessments; (b) state and local curricula; and, perhaps most important, (c) actual instruction. This objective is a challenging one, and there is evidence of major gaps. Often, graduation testing and promotion testing precede the alignment of curriculum and instruction with state standards (Elmore, 2000), and in many cases the tests are not well aligned with state standards: "There is little evidence to suggest that exit exams in current use have been validated properly against the defined curriculum and actual instruction; rather, it appears that many states may not have taken adequate steps to validate their assessment instruments, and that proper studies would reveal important weaknesses" (Stake, 1998, cited in NRC, 1999, p. 179).

The steps mentioned thus far do not include high-stakes testing. Even before alignment is complete, states and school districts can use large-scale assessments to help drive improvements in curriculum and instruction, and virtually all do. But the *Joint Standards*, the 1999 NRC study, and the July 2000 AERA *Position Statement* all assume that alignment will occur *before* such instruments become high-stakes tests for students. As noted above, all three say that tests should be used to decide whether individual students will be promoted or given high school diplomas only *after* students have been taught the kinds of subject matter and skill the tests measure.

The *Joint Standards* (1999), the NRC study (1999), and the AERA *Position Statement* (2000) describe measures a state or school district should take if it elects to use tests for high-stakes purposes. One, just noted, is not to use tests

for high-stakes purposes until schools are actually teaching students the relevant subject matter and skills. Second, test users should make sure that a high-stakes test is valid for its intended purpose. This may sound obvious, but it is not something every test user does. Chicago, for example, received national publicity for its use of the Iowa Test of Basic Skills (ITBS) in making promotion decisions, even though the district's chief accountability officer acknowledged that the ITBS is not valid as a measure of which students should be promoted or held back (NRC, 1999).

Third, a test use is inappropriate unless it leads to the best available treatment or placement for students (NRC, 1999). This means that states and school districts should refrain from using test scores (or other information) to justify educational decisions that are demonstrably harmful to students. Based on the weight of research evidence, two placements or treatments that typically harm students are retention in grade and placement in typical low-track classes (Hauser, 2001; NRC, 1999; Oakes et al., 1992). Retention and low-track placements are inimical to the goal of helping all students reach high levels of achievement. Both are inconsistent with principles of appropriate test use.

Fourth, test developers should take students with disabilities, English-language learners, minority students, and other groups into account beginning with initial test development, and should take steps to ensure that the test is equally valid for all major student populations that will take it (AERA, 2000; AERA et al., 1999; NRC, 1999).

Fifth, test users should not rely solely on test-score information in making promotion and graduation decisions (AERA, 2000; AERA et al., 1999; NRC, 1999). Instead, as colleges do, states and school districts should look at multiple measures of student achievement and readiness, and allow high achievement on one measure to balance lower performance on another.

Further, some states measure not only absolute achievement in the form of a percentage of students passing a test but also improvement over time (i.e., higher percentages of students passing a test). And some states measure whether school districts or schools are succeeding in closing the gap between high-achieving and low-achieving students. Each of these measures adds something important. An absolute standard signals that schools set high expectations for all students rather than lower expectations for some. A standard based on improvement recognizes that different students, schools, and school districts start out at different places and rewards progress. A standard based on whether schools are closing the achievement gap—between white students and minority students, between nondisabled students and students with disabilities, between native English speakers and English-language learners—encourages schools to

pay more attention to these very important goals. This is the theory behind the new federal yearly progress requirements for Title I recipients. The baseline for these improvements will be established in 2002–2003 (NCLBA, 2002).

Sixth, the debate over high-stakes testing is often framed in terms of either-or choices: whether a student who does not seem ready for the next grade should be retained or promoted, or whether a student who has not mastered the necessary knowledge and skill should receive a diploma. In each case, the choice is between unattractive alternatives. Though often unacknowledged, there is almost always a preferable third option: Any information schools can use to make a promotion or graduation decision can be used years earlier—before students reach a “gatepost”—to determine which children are performing poorly and to help get them the support they will need to be able to meet high standards. Teachers typically know, long before a promotion or graduation test, which students will need help if they are to pass. Effective early intervention is critical, as recent research shows (Grissmer et al., 2000).

Seventh, tests by themselves do not improve learning, any more than a thermometer reduces fever. At best, good tests provide information that can be used to improve instruction. It is important that this information, along with information from other sources, be available—in an understandable form—to policymakers, educators, parents, and students. And it is equally important for all concerned to know which policies and practices are likeliest to produce improved teaching and learning (Elmore, 2000; Grissmer et al., 2000). Educators and parents also need access to the resources that it takes to make the necessary changes in teaching and learning. Unfortunately, it is well known that many school districts and schools lack the resources they need to enable all children to reach high levels of achievement (National Academy of Education, 1995; NRC, 1999).

Finally, these questions all call for additional research: on what interventions work, on how treatments effective in some settings can be implemented widely, and, not least, on how high-stakes testing policies affect student learning and dropout rates, for students generally and for such important groups as students of color, English-language learners, and students with disabilities.<sup>14</sup> There is also a need for improved special education data broken out by race so researchers, policymakers, and practitioners can better understand the status and needs of minority students with disabilities.

In conclusion, the standards movement and high-stakes testing present both opportunities and risks to students with disabilities, minority students, English-language learners, and minority students with disabilities. These students are among those who stand to benefit most if all students receive high-quality instruction. Such students are also at great risk, however, especially in

states that administer high-stakes promotion and graduation tests before having made the improvements in instruction that will enable all students to meet the standards. Even failure rates well below 70–95 percent are plainly unacceptable, for these students and for society at large.

Educators and policymakers are right to be concerned about educating all students to high levels, and reaching this objective is obviously no simple matter. Promotion and graduation tests are one part of this picture, and debates over the necessity and desirability of such testing will continue even as it becomes more widespread. One thing is clear, however: If states and school districts are going to use high-stakes testing, then it is important that such testing be done properly. The basic principles of appropriate test use are relatively clear and enjoy broad support among researchers and practitioners. States and school districts that disregard these principles put their students—and themselves—at risk. The prospect of high failure rates has already produced a political backlash against some states' high-stakes testing programs. Lawsuits are also likely, if only because no reliable alternatives exist by which to ensure appropriate use of tests that affect students' life chances in such important ways. The stakes are high indeed.

## NOTES

1. This paper was written with support from the Carnegie Scholars Program of the Carnegie Corporation of New York, and from the National Center on Accessing the General Curriculum (NCAC), pursuant to cooperative agreement #H324H990004 under CFDA 84.324H between the Center for Applied Special Technologies (CAST) and the U.S. Department of Education, Office of Special Education Programs (OSEP). The opinions expressed herein are the author's; they do not necessarily reflect the position or policy of the Carnegie Corporation or of OSEP, and no endorsement should be inferred. The author holds the copyright for this paper and all rights pertaining thereto.
2. For example, of the seven states in which data for 1998 show that minority students outnumber white students in special education—Hawaii, California, Louisiana, Mississippi, New Mexico, South Carolina, and Texas (Office of Special Education Programs [OSEP], 2000, Table AA3)—all but California had exit exams in 1997–1998 (NRC, 1999), and California has since decided to administer exit tests as well. In the nine states where OSEP data for 1998 show that minority students represent 40 percent to 50 percent of special education enrollments—Alabama, Alaska, Arizona, Delaware, Florida, Georgia, Maryland, New York, and North Carolina (OSEP, 2000, Table AA3)—five (Alabama, Florida, Georgia, New York, and North Carolina) had exit exams in 1998 (NRC, 1999) and the remaining four have since decided to adopt them (American Federation of Teachers [AFT], 2001).
3. The new federal statute defines graduation rates in terms of the percentage of secondary school students who earn standard diplomas in the customary amount of time (No Child Left Behind Act, 2002).

4. Pass rates for each year would be lower if New York's calculations included all students with disabilities who enrolled that year, rather than only those who completed high school that year, and if the calculations included students with disabilities who had once been part of that year's graduating class but had dropped out or been retained in grade and thus had not completed high school that year. In either case the same numerator would be divided by a larger denominator and the resulting pass rate would be lower.
5. The data on students with disabilities were furnished upon request by Jeffrey Nellhaus, Associate Commissioner for Student Testing, Massachusetts Department of Education, by email dated November 16, 2001. I am grateful to Mr. Nellhaus. Similar data by race and ethnicity are publicly available in the report cited above (Massachusetts Department of Education, 2001); a student who has passed both tests is referred to as one "earning a competency determination."
6. See note 2 for further information about the source of data on minority representation among special education students in California.
7. Since "all seniors" includes students with disabilities, who in effect are counted as part of both groups, the actual difference between students with disabilities and *nondisabled* students is even higher than Alabama's figures suggest.
8. These include "state standards by grade, assessment tests linked to these standards, good systems for providing feedback to teachers and principals, some accountability measures, and deregulation of the teaching environment" (Grissmer et al., 2000, p. 58). The same study found that after controlling for family characteristics, results on the 1996 fourth-grade NAEP test showed black students in Texas outscoring black students in the other forty-nine states; white students in Texas outscoring white students in the other forty-nine states; and Latino students outscoring Latino students in forty-five of the other forty-nine states (Grissmer et al., 2000, p. 72).
9. This appears to be the case in New York. Personal conversation with James Kadamus, Deputy Commissioner of Education, New York, June 2000.
10. Ysseldyke et al. (1998) compare the performance of students with disabilities with that of "all" students, which means that students with disabilities are being counted in each group. Since students with disabilities have lower pass rates than nondisabled students, the 35–40 percentage point difference is smaller than what one would have found by comparing students with disabilities and students without disabilities. The gap between students with disabilities and students without disabilities, had it been calculated, would thus have been even higher than the 35–40 percentage points reported.
11. The previous section also notes, however, that pass rates on state tests are often not confirmed by scores on other tests, such as NAEP, that supposedly measure much of the same knowledge and skills as state tests, and that state-test pass rates would be lower if states accounted more fully for dropouts, retained students, and other students who are sometimes not included when pass rates are calculated.
12. That some of these data come from ninth-, tenth-, or eleventh-grade students, who still have time to acquire the requisite knowledge and skills, reduces only partially the seriousness of such high failure rates.
13. The examples offered in one study (Council of Chief State School Officers [CCSSO], 2001, pp. 24–25) show a .37 overlap in one state between instruction and the state fourth-grade math test, and a .33 overlap in one state between instruction and the state eighth-grade science test. The other recent report, while providing specific information on overlap in only one unnamed state, notes that in that state "instructional content was

not very well aligned with either the state test or the NAEP test for Grade 8 science (.17 for the state test, and .18 for the NAEP test)" (Blank, Porter, & Smithson, 2001, p. 26).

14. As the NRC study (1999) notes, "High-stakes testing programs should routinely include a well-designed evaluation component. Policymakers should monitor both the intended and unintended consequences of high stakes assessments on all students and on significant subgroups of students, including minorities, English-language learners, and students with disabilities" (p. 281).

## REFERENCES

- Alabama Department of Education. (2001, April 25). *Alabama high schools exceed expectations on new higher standards graduation exam*. Montgomery: Author.
- Alaska Department of Education. (2001). *Statewide spring 2001 HSGQE student test results*. Juneau: Author.
- Allington, R. (2000, May 10). Letters: On special education accommodations. *Education Week*, p. 48.
- Allington, R., & McGill-Franzen, A. (1992). Unintended effects of reform in New York. *Educational Policy*, 6, 397-414.
- American Educational Research Association. (2000). *AERA position statement concerning high-stakes testing in pre-K-12 education* [On-line]. Available: <http://www.aera.net.about/policy/stakes.htm>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Federation of Teachers. (1999). *Making standards matter 1999*. Washington, DC: Author.
- American Federation of Teachers. (2001). *Making standards matter 2001: A fifty-state report on efforts to implement a standards-system*. Washington, DC: Author.
- Benton, J. (2001, May 20-24). Falling through the cracks: Dropout figures vary with formula. *Dallas Morning News*, p. 1.
- Blair, J. (2002, February 13). Citing deficit, governor now proposes Wis. delay exam. *Education Week*, p. 23.
- Blank, R., Porter, A., & Smithson, J. (2001). *New tools for analyzing teaching, curriculum and standards in mathematics and science: Results from survey of enacted curriculum project final report*. Washington, DC: Council of Chief State School Officers.
- Boser, U. (2000, June 7). Teaching to the test? *Education Week*, pp. 1, 10.
- Brennan, R., Kim, J., Wenz-Gross, M., & Siperstein, G. (2001). The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS). *Harvard Educational Review*, 71, 173-216.
- Carnoy, M., Loeb, S., & Smith, T. (2001, January 13). *Do higher test scores in Texas make for better high school outcomes?* Paper prepared for the forum of The Civil Rights Project at Harvard University and Achieve, Inc., Dropouts in America: How severe is the problem? What do we know about intervention and prevention? Harvard University, Cambridge, MA.

- Citizens' Commission on Civil Rights. (2001). *Closing the deal: A preliminary report on state compliance with final assessment and accountability requirements under the Improving America's Schools Act of 1994*. Washington, DC: Author.
- Clarke, M., Haney, W., & Madaus, G. (2000). High-stakes testing and high school completion. *National Board on Educational Testing and Public Policy*, 1(3), 1-11.
- Cohen, M. (2001, January 19). *Review of state assessment systems for Title 1*. Memorandum to Chief State School Officers from the Assistant Secretary for Elementary and Secondary Education, U.S. Department of Education. Available at <http://www.ed.gov/offices/OESE>
- Council of Chief State School Officers. (2001). *Using data on enacted curriculum in mathematics and science: Sample results from a study of classroom practices and subject content* (Summary Report from Survey of Enacted Curriculum Project). Washington, DC: Author.
- Dailey, D., Zantal-Wiener, K., & Roach, V. (2000). *Reforming high school learning: The effect of the standards movement on secondary students with disabilities*. Alexandria, VA: Center for Policy Research on the Impact of General and Special Education Reform.
- Debra P. v. Turlington, 474 F. Supp. 244 (M.D. Fla. 1979); aff'd in part and rev'd in part, 644 F.2d 397 (5th Cir. 1981); rem'd, 564 F. Supp. 177 (M.D. Fla. 1983); aff'd, 730 F.2d 1405 (11th Cir. 1984).
- Elmore, R. (2000). *Building a new structure school leadership*. Washington, DC: Albert Shanker Institute.
- GI Forum v. Texas Education Agency, 87 F. Supp. 2d 667 (W.D. Tex. 2000).
- Goertz, M., & Duffy, M. (2001). *Assessment and accountability systems in the 50 states: 1999-2000*. Philadelphia: Consortium for Policy Research in Education.
- Gonzalez, P., & Carlson, E. (2001, April 30). *Preliminary results from the study of personnel needs in special education (SPeNSE)*. Paper presented at the annual meeting of CSPD, Washington, DC. Available at [www.spense.org](http://www.spense.org)
- Grissmer, D., Flanigan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP scores tell us*. Santa Monica, CA: Rand.
- Guy, B., Shin, H., Lee, S., & Thurlow, M. (1999). *State graduation requirements for students with disabilities* (Technical Report No. 24). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Haney, W. (2001, January 13). *Revisiting the myth of the Texas miracle in education: Lessons about dropout research and dropout prevention*. Paper prepared for the forum of The Civil Rights Project of Harvard University and Achieve, Inc., Dropouts in America: How severe is the problem? What do we know about intervention and prevention? Harvard University, Cambridge, MA.
- Hauser, R. (2001). Should we end social promotion? Truth and consequences. In G. Orfield & M. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in education* (pp. 151-178). New York: Century Foundation Press.
- Improving America's Schools Act of 1994, 20 U.S.C. sections 6301 et seq.
- Individuals with Disabilities Education Act, 20 U.S.C. section 1401 et seq. (1997).
- Jacob, B. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23, 99-122.
- Keller, B. (2000, April 12). More N.Y. special education students passing state tests. *Education Week*, p. 33.

- Keller, B. (2001, October 3). Calif. to study whether graduation test should be delayed. *Education Week*, p. 24.
- Klein, S., Hamilton, L., McCaffrey, D., & Stecher, B. (2000). *What do tests scores in Texas tell us?* Santa Monica, CA: Rand.
- Koretz, D., & Barron, S. (1998). *The validity of gains on the Kentucky Instructional Results Information Systems (KIRIS)*. Santa Monica, CA: Rand.
- Koretz, D., & Berends, M. (2001). *Changes in high school grading standards in mathematics, 1982-1992*. Los Angeles, CA: Rand.
- Koretz, D., & Hamilton L. (2001). *The performance of students with disabilities on New York's revised Regents Examination in English*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Koretz, D., Linn, R., Dunbar, S., & Shepard, L. (1991, April). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Lee, J. (1998, December 4). *Using high-stakes test results to give disadvantaged kids access to outstanding responsive teachers*. Paper presented at The Harvard Civil Rights Project/Teachers College conference on high-stakes testing and civil rights, New York.
- Lillard, D., & DeCicca, P. (2001). Higher standards, more dropouts? Evidence within and across time. *Economics of Education Review*, 20, 459-474.
- Linn, R. (2000). Assessments and accountability. *Educational Researcher* 29(2), 4-16.
- Linn, R. (2001). *The design and evaluation of educational assessment and accountability systems*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Madaus, G., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from one hundred years of test data. In G. Orfield & M. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high stakes testing in education* (pp. 85-106). New York: Century Fund.
- Massachusetts Department of Education. (2001). *Spring 2001 MCAS tests: State results by race/ethnicity and student status*. Boston: Author
- McLaughlin, M. (2000, June 30). *High-stakes testing and students with disabilities*. Paper presented at the National Research Council conference on the role of the law in achieving high standards for all, Washington, DC.
- Mehrens, W. A. (1998, April). *Consequences of assessment: What is the evidence?* Paper presented at the annual meeting of the American Educational Research Association, San Diego.
- Murnane, R., & Levy, F. (2001). Will standards-based reforms improve the education of children of color? *National Tax Journal*, 54, 401-416.
- National Academy of Education. (1995). *Improving education through standards-based reform* (M. McLaughlin, L. Shepard, & J. O'Day, Eds.). Washington, DC: Author.
- National Center for Education Statistics. (2001). *The condition of education 2001*. Washington, DC: U.S. Government Printing Office.
- National Research Council. (1997). *Educating one and all: Students with disabilities and standards-based reform* (L. M. McDonnell, M. J. McLaughlin, & P. Morison, Eds.). Washington, DC: National Academy Press.
- National Research Council. (1999). *High stakes: Testing for tracking, promotion, and graduation* (J. Heubert & R. Hauser, Eds.). Committee on Appropriate Test Use. Washington, DC: National Academy Press.

- National Research Council. (2001). *Understanding dropouts: Statistics, strategies, and high-stakes testing* (A. Beatty, U. Neisser, W. Trent, & J. Heubert, Eds.). Washington, DC: National Academy Press.
- Natriello, G., & Pallas A. (2001). The development and impact of high stakes testing. In G. Orfield & M. Kornhaber (Eds.), *Raising standards or raising barriers: Inequality and high-stakes testing in public education* (pp. 19–38). New York: Century Foundation Press.
- New York Department of Education, Office of Vocational and Educational Services for Students with Disabilities. (2000). *2000 pocket book of goals and results for individuals with disabilities*. Albany: New York Department of Education.
- New York Department of Education, Office of Vocational and Educational Services for Students with Disabilities. (2001). *2001 pocket book of goals and results for individuals with disabilities*. Albany: New York Department of Education.
- No Child Left Behind Act, Public Law 107-110 (January 8, 2002).
- Oakes, J., Gamoran, A., & Page, R. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In P. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570–608). New York: Macmillan.
- Office for Civil Rights. (2000). *The use of tests when making high-stakes decisions for students: A resource guide for educators and policymakers*. Washington, DC: U.S. Department of Education.
- Office of Special Education Programs. (2000). *To assure the free appropriate public education of all children with disabilities: Twenty-second annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, DC: U.S. Department of Education.
- Olson, L. (2001, January 24). States adjust high-stakes testing plans. *Education Week*, pp. 1, 18–19.
- Porter, A., & Smithson, J. (2000, April). *Alignment of state testing programs, NAEP, and reports of teacher practice in grades 4 and 8*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Porter, A., & Smithson, J. (2001). *Defining, developing, and using curriculum indicators*. Philadelphia: Consortium for Policy Research in Education.
- Quenemoen, R., Lehr, C., Thurlow, M., & Thompson, S. (2000). *Social promotion and students with disabilities: Issues and challenges in developing state policies* (NCEO Synthesis Report No. 34). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Riley, R. W. (2000, February 22). *Setting new expectations* (annual State of American Education address). Paper presented at Southern High School, Durham, NC.
- Robelin, E. (2001, November 28). States sluggish on execution of 1994 ESEA. *Education Week*, pp. 1, 26–27.
- Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23, 197–228.
- Sack, J. (2000, April 19). Researchers warn of possible pitfalls in spec. ed. testing. *Education Week*, p. 12.
- Schrag, P. (2000). Too good to be true. *American Prospect*, 4(11), 46.
- Shepard, L. A., & Smith, M. L. (Eds.). (1989). *Flunking grades: Research and policies on retention*. London: Falmer Press.

- Shore, A., Madaus, G., & Clarke, M. (2000). Guidelines for policy research on educational testing. *National Board on Educational Testing and Public Policy*, 1(4), 1-7.
- Stake, R. (1998, July). Some comments on assessment in U.S. education. *Educational Policy Analysis Archives* [Online] 6(14). Available: <http://epaa.asa.edu/epaa/v6n14.htm>
- Thompson, S., & Thurlow, M. (2001, June). *2001 state special education outcomes: A report on state activities at the beginning of a new decade*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M., & Johnson, D. (2000). High-stakes testing of students with disabilities. *Journal of Teacher Education*, 51, 305-314.
- Thurlow, M., Nelson, J., Teelucksingh, E., & Ysseldyke, J. (2000). *Where's Waldo? A third search for students with disabilities in state accountability reports* (Technical Report No. 25). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M., & Thompson, S. (1999). *Diploma options and graduation policies for students with disabilities*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Viadero, D. (1999, October 6). Stanford report questions accuracy of tests. *Education Week*, p. 3.
- Viadero, D. (2000, May 31). Testing system in Texas yet to get final grade. *Education Week*, p. 1.
- Weckstein, P. (1999). School reform and enforceable rights to an adequate education. In J. Heubert (Ed.), *Law and school reform: Six strategies for promoting educational equity* (pp. 306-389). New Haven, CT: Yale University Press.
- Wise, L., Sipes, D. E., Harris, C., George, C., Ford J., & Sun, S. (2002). *Independent evaluation of the California High School Exit Examination (CAHSEE): Analysis of the 2001 administration*. Sacramento: California Department of Education. Available at <http://www.cde.ca.gov/statetests/cahsee/2001humrroreport.html>
- Ysseldyke, J., Thurlow, M., Langenfeld, M., Nelson, J., Teelucksingh, E., & Seyfarth, A. (1998). *Educational results for students with disabilities: What do the data tell us?* (Technical Report No. 23). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

