Large State-Level Fluctuations in Mental Retardation Classifications Related to Introduction of Renormed Intelligence Test

Matthew H. Scullin
The University of Texas at El Paso

Abstract

Oppositely valenced forces may be at work to influence rates of placement of children into mental retardation programs. On one hand, educational policies regarding intellectual disability and concerns about overrepresentation of minorities in special education may contribute to lower placement rates; on the other hand, more difficult intelligence test norms may be a countervailing force, increasing placement rates. An analysis of longitudinal data on state and national level placement rates reveals that a lengthy and steep 12-year decline in students receiving mental retardation services reversed shortly after the introduction of the WISC–III in 1991. This phenomenon has relevance for death-penalty cases, because this historical pattern may affect the ability to establish whether an adult meets the developmental period onset criterion for mental retardation.

It is now well-established that IQs on the Wechsler series of intelligence tests rose steadily throughout the 20th century at the rate of about .3 IQ points per year, although this rise has been masked by the periodic renorming of intelligence tests to reset the mean to 100 (Flynn, 1984, 1987, 1998). This IQ rise has been dubbed the *Flynn effect* after James Flynn, who has extensively documented this phenomenon. There is a great deal of controversy surrounding why this gain in IQ occurred (e.g., Neisser, 1998; Neisser et al., 1996), but its effects on children and adults being tested for mental retardation are quite real (e.g., Flynn, 2000; Kanaya, Scullin, & Ceci, 2003). As will be discussed later, these effects may range from whether or not children receive special education services to whether or not an adult is eligible for the death penalty.

In this paper I have focused upon the impact of the Flynn effect upon children receiving the special education classification of mental retardation during their developmental period (ages 6 through 17), when mental retardation evaluations are most frequently conducted. A developmental period onset (prior to age 18) is one of the three defining criteria of mental retardation according to most sources (e.g., Luckasson et al., 2002). The other two criteria are subaverage intellectual functioning (typically IQ of 70–75 or below); and limitations in conceptual, social, and practical adaptive skills, such as communication and social functioning (often referred to as adaptive functioning).

Most of the researchers investigating the impact of the Flynn effect on special education classifications have used data from a limited number of school districts (e.g., Kanaya et al., 2003; Sanborn, Truscott, Phelps, & McDougal, 2003). Relatively little is known about whether the Flynn effect has had a widespread impact on children's mental retardation classifications over time. Using longitudinal data from government reports, I assessed whether mental retardation rates moved in the expected directions as the norms of the Wechsler Intelligence Scale for Children-Revised—WISC–R (Wechsler, 1974) grew older

and were eventually replaced by the renormed Wechsler Intelligence Scale for Children–3rd Revision—WISC-III (Wechsler, 1991) between 1981 and 1999.

Theoretically, over time the Flynn effect results in inflated IQs and dramatically reduces the number of people classified as having mental retardation because fewer people will score below the IQ 70–75 threshold for mental retardation (Flynn, 1984, 1987). When an IQ is calculated, it is compared to the normative performance of a standardization sample that, in theory, provides a representative sample of the population as a whole. Over time, normative performance in the population increases. When a new standardization sample is drawn years later at the time that a test is renormed, the new standardization sample will perform better on the intelligence test than did the old standardization sample. A concrete example of this phenomenon is that both the WISC-R and the WISC-III were in common use in 1991 (Kanaya et al., 2003); a 9-year-old who took the WISC-R in 1991 and received a score of 105 would obtain that score because the child was being compared to the performance of the standardization sample of children who took the test when it was normed 19 years earlier in 1972. However, normative IQ performance improved at a rate of .3 points per year between 1972 and 1989, when the WISC–III was normed. If the same child was tested on the WISC–III, the child would receive an IQ of 100 because his or her performance would be compared to the improved performance of the standardization sample who took the exam in 1989, 2 years earlier.

The Wechsler intelligence tests are the most popular instruments for assessing children being considered for special education services (Kamphaus, 1993), and changes to their norms would be expected to have an impact on children's special education placements. Kanaya et al. (2003) used triennial reevaluation data to determine the size of the impact of the Flynn effect on 743 children in the mild mental retardation (IQ 55–70) and borderline (IQ 71–85) ranges on the WISC-R and the WISC-III, approximating the American Psychiatric Association's (1994) guidelines for these classifications. Using WISC–R test and WISC–R retest data as a baseline to control for typical IQ changes upon retesting on the same test, Kanaya et al. found in regression analyses that Full-Scale IQs dropped 5.6 points on average when children were tested on the WISC-R and retested on the WISC-III. This finding was consistent with Flynn's (1998) estimate of the size of the WISC-R/WISC-III IQ difference being 5.3 points in the average IQ range and with estimates by researchers who have looked at triennial reevaluations of children receiving mental retardation services (Bolen, Aichinger, Hall, & Webster, 1995; Slate & Saarnio, 1995; Vance, Maddux, Fuller, & Awadh, 1996) and learning disability services (Sanborn et al., 2003; Truscott & Frank, 2001; Truscott & Volker, 2005). This finding suggests that the Flynn effect is not simply due to children with high IQs scoring higher, thus increasing the raw score variance of the IQ distribution, but that the Flynn effect also elevates the low end of the IQ distribution (Sanborn et al., 2003).

Kanaya et al. (2003) also established that the mental retardation classification rate of school-age children initially tested on the newly normed WISC-III was much greater than the mental retardation classification rate of children initially tested on the WISC-R in an equivalent (adjusted for the Flynn effect) IQ range shortly before the release of the WISC-III. This is because the introduction of the WISC-III would theoretically double the number of children who meet the IQ 70 or below criterion for mental retardation because about half of the children with IQs of 75 or below have an IQ of 71–75.

The impact of the introduction of a renormed intelligence test on mental retardation classifications is likely to occur over a several year period for at least three reasons. First, the proportion of children who are initially tested on a newly normed test will gradually increase for the first several years after the release of the test. For example, 1 out of 12 children who would be initially tested at some point between the ages of 6 and 17 might be tested in Year 1, another 1 out of 12 in Year 2, etc. Second, about one third of children who tested between IQ 71 and IQ 85 on the WISC-R shortly before the introduction of the WISC-III received special education services for disorders other than mental retardation (predominantly learning disability) in the Kanaya et al. (2003) sample. Many of these children became eligible for mental retardation services when they tested 70 or below on the newly normed WISC-III during their triennial reevaluations for continuing special education services. Finally, not all school districts immediately replace their intelligence tests when a new measure comes out, but

rather phase in new measures over a period of several years (Kanaya et al., 2003).

In summary, a possible implication of the findings of Kanaya et al. (2003) is that the Flynn effect has had a widespread impact on the number of children classified as having mental retardation over time. As the WISC-R norms became more obsolete, fewer children would be expected to test in the mental retardation range. The introduction of the renormed WISC-III in 1991 would be expected to gradually increase the number of children eligible for mental retardation services because the WISC-III test norms would be more in line with the population mean and cause more children to test in the mental retardation range. However, there are several factors other than the Flynn effect that could potentially have affected mental retardation classification rates during the 1981 through 1999 time frame of this study.

Researchers have noted the steep decline in the percentage of children in mental retardation programs in the 1980s, with some hypothesizing that this decrease was due to the concomitant increase in the number of children placed into learning disability programs during this same time period, including many who would have formerly been placed into mental retardation programs (e.g., Gottlieb, Alter, Gottlieb, & Wishner, 1994; MacMillan, Gresham, Siperstein, & Bocian, 1996). The percentage of children classified as having a learning disability rose from 3.90% of enrolled 6- to 17-year-old children in 1981 to 5.75% in 1999 (U.S. Department of Education, 1983, 2001). A classification of *learning disability* means that a child has academic achievement that is lower than expected for his or her age, level of schooling, or IQ. Earlier definitions of learning disability (e.g., those provided by the American Psychiatric Association, 1994, and 1997 revisions to the Individuals With Disabilities Education Act—IDEA 97, P.L. 105–17) consistently stated that these achievement problems should not be due to mental retardation, although the move away from the discrepancy model for determining learning disability classifications has resulted in a relaxing of this standard (e.g., American Psychiatric Association, 2000).

Concern about the overrepresentation of minorities in mental retardation programs. The high percentage of minority children in mental retardation programs has generated a good deal of concern and even a special mention in the IDEA 1997 legislation. In California, schools are operating under the Ninth Circuit Court of Appeals ruling *Larry P. v. Riles* (1972, 1974, 1979, 1984, 1986), and as a result they are not allowed to consider minority children's IQs when they are being considered for a mental retardation classification. Whether or not they are allowed to use intelligence tests for minority children's special education classifications, school psychologists in a number of school districts around the country have reported to me that they are under legal and educational policy pressure to not place excessive numbers of minority children into mental retardation programs.

Broadening the other health impairment category. The IDEA amendments of 1990 (P.L. 101–476) allowed children with attention deficit disorder (ADD) and attention deficit hyperactivity disorder (ADHD) to receive special education services for other health impairment. Children are eligible for other health impairment status if their limited ability to attend to their educational environment (a characteristic of both ADD and ADHD) negatively affects their educational performance. Thus, academically failing children who do poorly on intelligence tests due to their inability to focus on the task may receive other health impairment as a classification rather being classified as having mental retardation.

Extending the age range of the developmental delay category. The IDEA 1997 raised the age limit for a diagnosis of developmental delay from 5 to 9 years of age beginning with the 1997–1998 school year. A diagnosis of developmental delay means that a child has not met developmental milestones for physical, cognitive, communicative, social, or emotional development and is in need of special education services (Individuals With Disabilities Education Act, 1997). Although not all children who are classified as having developmental delay are later assigned a classification of mental retardation, students with mental retardation have often received a diagnosis of developmental delay prior to their mental retardation diagnosis.

Between-state variability in mental retardation classification rates. Nationally, the percentage of enrolled children aged 6 to 17 who were classified as having mental retardation in the 1999–2000 school year was 1.19% (including the .04% of children aged 6 to 9 reported to be receiving services for developmental delay). State policies regarding the classification of children as having mental retardation vary widely from state to state and range from a low of .33% in New Jersey to a high of 3.04% in Kentucky (U.S. Department of

Education, 2001). There were 21 states reporting that their mental retardation rate was less than 1% of their school-age population for the 1999–2000 school year, with the two leading death penalty states, California and Texas, reporting mental retardation rates of only .48% and .53%, respectively. Because more than nine times as many children are diagnosed with mental retardation in some states compared to others suggests that in many states strong efforts are made to minimize the number of children labeled with mental retardation or, alternatively, deleterious environmental factors may place many more children at risk for mental retardation. Researchers have noted that the discrepancy between states raises concerns about a lack of consistency in diagnosing mental retardation, often due to reluctance to assign the mental retardation classification or the more extensive use of the learning disability classification (reviewed by MacMillan & Reschly, 1998).

In summary, the increasing popularity of the learning disability category, concerns about excessive numbers of minorities in mental retardation programs, and inclusion of ADHD students in the other health impairment category would all be expected to reduce the number of children categorized as having mental retardation and, perhaps, mitigate the impact of the Flynn effect. The raising of the age limit for developmental delay has been more ambiguous in terms of its impact on mental retardation classifications for reasons that will be discussed below. Finally, the large differences in mental retardation classification rates between states highlights the importance of looking at fluctuations in mental retardation rates at a within-state level.

In this study, I examined the extent over time of the fluctuations in the number of children classified as having mental retardation by analyzing longitudinal mental retardation placement trends in all 50 states, the District of Columbia, and the United States as a whole. This is a broader evaluation of the impact of the Flynn effect than was provided by Kanaya et al. (2003), who examined the impact of the Flynn effect in nine school districts. I predicted that at the state and national level, rising IQs occurring as a result of the Flynn effect on the WISC-R would be expected to decrease the percentage of school-age children eligible for mental retardation services during the 1980s. By 1991, the year the WISC-III was introduced, the mental retardation classification rate would theoretically be as low as half of what it was when the WISC-R was introduced because of the 5.6 point rise in Full-Scale IQ found by Kanaya et al. (2003). This decline would then be followed by a gradual reversal of this trend after 1991, resulting in an increase in the percentage of children receiving mental retardation services.

National and state data are available on the percentage of enrolled school-age children who are receiving mental retardation services on December 1 of each school year between 1981 (i.e., the 1981–1982 school year) and 2002, so it is possible to mathematically model how the percentage of children eligible for mental retardation services changed over this time period within each state. I hypothesized that most states would exhibit a quadratic function in terms of the percentage of enrolled school-age children eligible for mental retardation services between the years of 1981 and 1999. This function would take the shape of a parabola that would reflect a declining percentage of children eligible for mental retardation services in the years after 1981. This decline would reach a low point around 1992, when the WISC-III was introduced in most school districts, followed by an increase in the percentage of children eligible for mental retardation services in the years after 1992. The 1981 school year was 7 years after the introduction of the WISC-R, and 1999 was 7 years after the introduction of the WISC-III. These years were chosen because theoretically there would be a leveling off of the increase of mental retardation classifications several years after the reintroduction of the WISC-III, followed by a decline. I worked under the assumption that the 7-year span following the introduction of the WISC-III would be sufficient to capture the inflection prior to the leveling and decline; this assumption may be visually assessed by comparing the modeled curves to the data points. Although changes in education policies, such as the broadening of the other health impairment classification and the steadily increasing popularity of the learning disability classification, would be expected to mitigate the impact of the introduction of harder IQ norms, I predicted that the quadratic pattern of a decline in the rate of mental retardation placements, followed by an inflection and upward trend, would be present in most states.

Method

Sample

Data on the percentage of enrolled students in each of the 50 states plus the District of Co-

lumbia and the United States as a whole who were being served by the Education of the Handicapped Act (EHA, P.L. 94–142) or the Individuals With Disabilities Education Act (IDEA, Part B) for mental retardation were gathered or calculated from U.S. Department of Education (1983–2004) tables beginning with the 1981 school year. Beginning with the 1997 school year, the Department of Education included developmental delay as an additional category for 6- to 17-year-olds. Initially, only a minority of states reported children in this category and these children only comprised .04% of the total number of children in this age group by the 1999 school year. Because this category largely consists of students who would formerly have been labeled with mental retardation, these students were added to the mental retardation totals. I chose to use the percentage of enrolled students rather than raw numbers of students receiving mental retardation services in order to control for the large fluctuations in the sizes of the overall student bodies that occurred between the 1981 and 1999 school years.

Procedure

I scaled the percentage of enrolled 6- to 17-year-old children receiving mental retardation services so that the quadratic equation was in the form of "percentage of enrolled students receiving mental retardation services relative to 1981" = $b_0 + b_1x + b_2x^2$, where x = school year − 1981. This sets Year 0 as 1981. The data were scaled so that the b_0 intercept was 100 for each state, the District of Columbia, and the United States as a whole in 1981. Scaling the data permitted a meaningful comparison of estimates of the changes in the percentage of children receiving mental retardation services between states that had markedly different baseline percentages of children receiving mental retardation services in 1981. For example, in 1981, California reported the smallest percentage of school-age children receiving mental retardation services at .56%; Alabama topped the list at 4.26%. Therefore, a shift of .2 percentage points of all enrolled children would have a proportionately much larger impact on the relative number of children receiving mental retardation services in California compared to Alabama. Second, this approach allowed me to assess the percentage change in mental retardation classifications since 1981 for each succeeding school year, controlling for overall enrollment trends. Finally, I estimated the year in which the quadratic function reached its high or low point (vertex) as well as the percentage of students categorized as having mental retardation at this point relative to the baseline of 100% categorized with mental retardation in 1981.

Because the intercept was scaled to a common metric, I was able to compare the vertices (low or high points) of the quadratic function with a hypothetical Flynn effect model, in which the primary determinant of children's mental retardation classification was their IQ and all extraneous factors influencing classifications (e.g., states' average classification rates) was held constant (see Figure 1). There are three assumptions to the hypothetical Flynn effect model. The first is that by 1981, 9 years after the norming of the WISC-R and 7 years after its introduction, the likely increase in mental retardation classifications due to the introduction of the WISC-R was at or near its peak. The second is that, on average, IQs rose by about .3 points per year, resulting in fewer children receiving a classification of mental retardation over time. By 1981, one would expect around 1.46% of school-age children to receive a score under 70. By 1991, 19 years after renorming and the year in which the WISC-III was introduced but was not yet in wide use, this percentage would have fallen to .87%. In other words, the 1991 rate would be

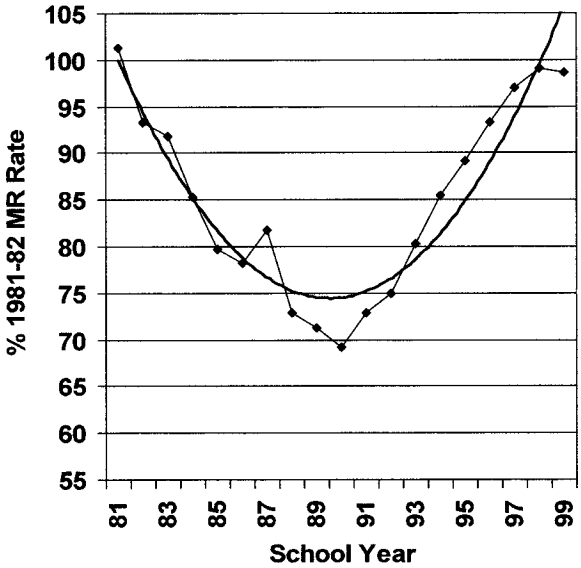

Figure 1. Graph of the change between 1981 and 1999 of the percentage of enrolled students who were in mental retardation (MR) programs in a hypothetical Flynn effect alone model. ♦ = MR rate, ___ = Poly. (MR rate).

around 60% of the 1981 rate. The third assumption is that there would be a gradual reversal of the decline in mental retardation classifications beginning around 1991 as the WISC-III spread into broader use, followed by a steady rise in mental retardation classifications for a few years. If the best-fitting parabola for the hypothetical model was symmetric around 1991, the mental retardation classification rate would reach 86% of its 1981 rate by 1999. As mentioned in the introduction, there are a number of factors that might mitigate a rebound in placements, resulting in a parabola that was steeper in its downward trajectory than in its upward trajectory in the range of the data points.

Results

Table 1 reports the quadratic curve equations that were fitted for each of the 50 states, the District of Columbia, and the United States. As can be seen from the R^2 and the coefficients, the quadratic equations fit the data for the large majority of states as well as the United States as a whole. The table shows the year in which states' mental retardation rates reached their vertex, the percentage of children receiving mental retardation services at the vertex relative to the baseline percentage of enrolled children receiving services in the 1981. The percentage of children receiving mental retardation services relative to 1981 was also calculated for 1999, the most recent data point.

There were 2 states with highly anomalous trend lines whose quadratic equations are reported in Table 1 but that were excluded from the state average analyses because they were strongly affected by state reporting practices that had nothing to do with the states' actual mental retardation classification rates. Massachusetts had a trend line that looks like a cliff, a relatively flat line followed by a sudden drop of more than 50% in reported mental retardation placements in 1992, followed by another flat line. This occurred because Massachusetts does not report the actual number of students who are receiving mental retardation or other special education services to the federal government but, instead, provides an estimate of the number of students receiving mental retardation services by calculating a fixed percentage of the total number of children receiving special education services. Beginning in 1992, this fixed percentage for mental retardation was arbitrarily reduced by more than half, resulting in an inverted parabolic function for Massachusetts. Wisconsin showed a 50% drop in its mental retardation rate in 1987, followed by an abrupt return to near earlier levels in 1995 that coincided with a change in its reporting policy to eliminate the multiple disability category and only report students by their primary disability (U.S. Department of Education, 1997). Not deleted from the state average data were Alabama, Maine, Oklahoma, Pennsylvania, and Utah, which have seen fairly steady declines in mental retardation, resulting in minimum vertices that are either nonexistent (in the case of Oklahoma) or outside of the range of the data points. South Dakota's data are much better fit by a cubic model (a rise, followed by a fall, then another rise). Excluding these states had a minimal effect on the state mean calculations and resulted in a 1999 state mean that was 2 percentage points higher (76% rather than 74%).

The vertices for the United States and the state average were very similar to the Flynn effect alone model, although the year of the low point of the model for the United States was 3 years later than the Flynn effect alone model, and the rise in the 1990s was not as large as the Flynn effect alone model. Figure 2 shows the quadratic curve for the United States superimposed over the actual data. The graph clearly shows that the steep decline in the percentage of children receiving mental retardation services during the 1980s leveled off in the early 1990s and then very slightly increased. It is important to note that the vertices of the quadratic functions are affected by factors like spread of data points around the line and lack of symmetry between the steepness of the rise and fall of the data points. Notably, in the United States data, the decline in mental retardation classifications corresponded almost exactly to what would be expected in the Flynn effect alone model. However, the rise in mental retardation classifications after the vertex was substantially lower, pulling the vertex of the function to the right of the lowest reported data point, which was in 1992.

Although the United States as a whole did not see a steep increase in children receiving mental retardation services following the introduction of the WISC-III in 1992, the overall pattern seen in Figure 1 was reflected to a greater or lesser extent in 43 of the 51 states and the District of Columbia (including 34 out of the 38 death penalty states), with many states showing a marked inflection in the quadratic function in the early to mid-

Table 1. Quadratic Curve Fitting for Change in Percentage of Enrolled Students in Mental Retardation (MR) Programs by Location From 1981–1999

State	R^2	F	b_1	b_2	Year of vertex	Vertex % of 1981 MR rate	1999 % of 1981 MR rate
Alabama[a,b]	.96	178.17	−3.61	.08	2002	61	62
Alaska	.75	23.87	−9.91	.48	1991	49	78
Arizona[a]	.89	67.26	−5.82	.23	1994	63	70
Arkansas[a]	.94	129.15	−7.31	.32	1993	58	71
California[a]	.87	54.05	5.23	.26	1991	74	90
Colorado[a]	.98	348.27	−9.60	.35	1995	34	41
Connecticut[a]	.89	66.46	−6.72	.25	1995	54	59
Delaware[a]	.83	38.77	−8.48	.41	1991	57	81
DC	.58	11.14	−7.71	.42	1990	64	96
Florida[a]	.66	15.25	−4.03	.19	1992	79	90
Georgia[a]	.65	14.91	−4.93	.22	1992	73	83
Hawaii	.98	424.07	−9.69	.73	1988	68	162
Idaho[a]	.68	17.35	−3.39	.15	1992	81	87
Illinois[a]	.94	135.28	−8.00	.31	1994	48	56
Indiana[a]	.85	44.89	−4.48	.20	1992	75	84
Iowa	.92	94.55	−3.45	.30	1987	90	136
Kansas[a]	.69	18.01	−4.83	.18	1994	68	71
Kentucky[a]	.89	62.02	−3.98	.18	1992	78	88
Louisiana[a]	.79	29.14	−6.73	.39	1990	71	105
Maine[b]	.96	176.53	−8.48	.21	2001	15	16
Maryland[a]	.95	146.67	−6.51	.28	1993	62	73
Massachusetts[c,d]	*.71*	*19.10*	*1.85*	*−.34*	*1984*	*103*	*23*
Michigan	.89	65.10	−5.77	.27	1992	69	83
Minnesota	.95	157.42	−4.62	.14	1998	61	61
Mississippi[a]	.98	351.81	−9.90	.40	1994	38	50
Missouri	.98	323.91	−5.89	.17	1998	50	50
Montana[a]	.75	24.01	−4.33	.18	1993	74	80
Nebraska[a]	.89	66.33	−5.99	.35	1990	74	106
Nevada[a]	.73	21.71	−2.62	.08	1998	78	78
New Hampshire[a]	.98	429.03	−8.79	.34	1994	43	51
New Jersey[a]	.99	870.59	−10.58	.39	1995	27	34
New Mexico[a]	.96	178.72	−5.99	.23	1994	61	67
New York[a]	.99	673.08	−7.83	.24	1997	36	36
North Carolina[a]	.87	53.18	−7.94	.39	1991	60	84
North Dakota	.96	168.93	−5.26	.17	1997	59	60
Ohio[a]	.84	41.34	−4.97	.23	1992	73	86
Oklahoma[a,b]	.87	54.13	.04	−.10	1981	100	68
Oregon[a]	.83	38.29	−5.40	.19	1996	61	63
Pennsylvania[a,b]	.93	112.36	−3.61	.05	2018	34	51
Rhode Island	.60	12.05	−4.43	.16	1995	69	72
South Carolina[a]	.95	156.11	−7.59	.35	1992	59	76
South Dakota[a,b,c]	.55	9.75	−.94	−.02	1957	111	77
Tennessee[a]	.94	125.03	−7.53	.36	1991	61	82

Table 1 continues

Table 1. Continued

State	R^2	F	b_1	b_2	Year of vertex	Vertex % of 1981 MR rate	1999 % of 1981 MR rate
Texas[a]	.96	199.88	−5.38	.17	1997	58	59
Utah[a,b]	.83	40.2	−2.24	.03	2016	60	70
Vermont	.96	185.42	−9.87	.39	1994	38	50
Virginia[a]	.81	35.09	−7.38	.41	1990	66	99
Washington[a]	.97	219.62	−4.57	.14	1997	63	63
West Virginia	.83	37.60	−5.04	.28	1990	77	100
Wisconsin[d]	*.70*	*18.26*	*−14.03*	*.76*	*1990*	*36*	*95*
Wyoming[a]	.93	104.25	−7.97	.35	1992	55	71
United States	.97	299.60	−5.85	.22	1994	61	66
State average	.87	146.04	−6.43	.28	1993	62	74
Hypothetical Flynn model			−7.57	.38	1991	63	88

Note. All $F(2, 16)$ values were significant at the $p < .05$ level. The percentage of enrolled students in each state was scaled so that the b_0 intercept (Year 0 or 1981) of the quadratic equation was equal to 100. *Year of vertex* denotes the year in which the quadratic function reaches its lowest or highest point. States with B, C, and/or D subscripts were not included in state averages. States in italics were excluded from state average analyses.
[a]Denotes a death penalty state. [b]Vertex outside range of data. [c]Inverted quadratic function. [d]Quadratic function strongly affected by changes in state reporting policies.

1990s. Figure 3 presents the data from Maryland, a state whose values for b_1 and b_2 were close to the medians and means for the average of the states. The percentage of children receiving mental retardation services dropped to 62% of its 1981 rate by 1992, then rose to 77% of its 1981 level by 1999. As can be seen in Figure 4, the state with the largest number of prisoners on death row, California, exhibited a very similar pattern, with the percentage of children receiving mental retarda-

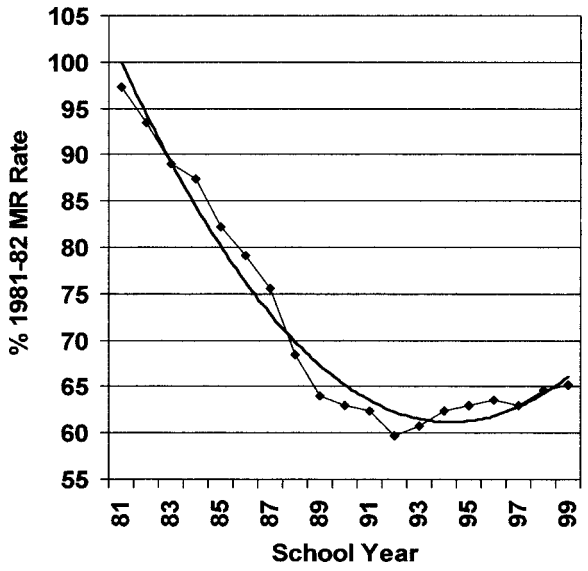

Figure 2. Graph of the change between 1981 and 1999 of the percentage of enrolled students in the United States who were in MR programs. ♦ = MR rate, ⎯ = Poly. (MR rate).

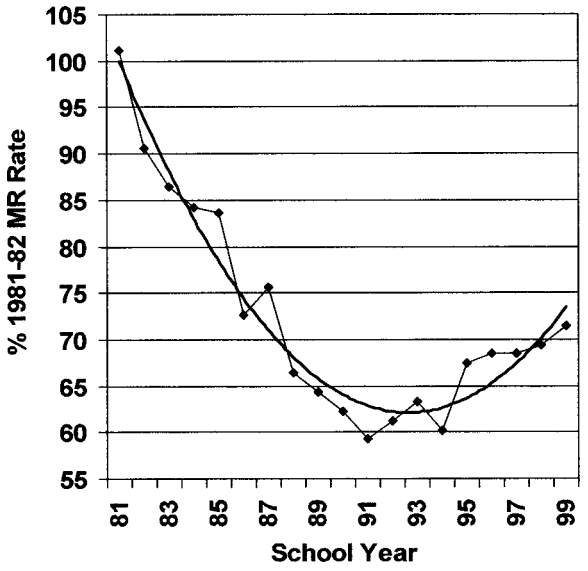

Figure 3. Graph of the change between 1981 and 1999 of the percentage of enrolled students in Maryland who were in MR programs. ♦ = MR rate, ⎯ = Poly. (MR rate).

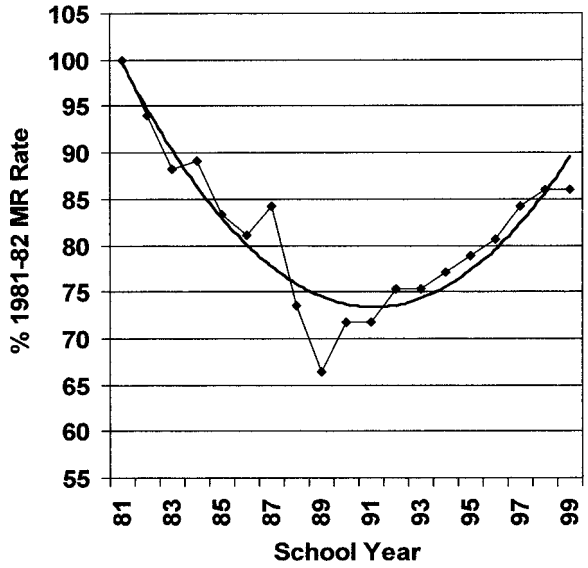

Figure 4. Graph of the change between 1981 and 1999 of the percentage of enrolled students in California who were in MR programs. ♦ = MR rate, ⎯ = Poly. (MR rate).

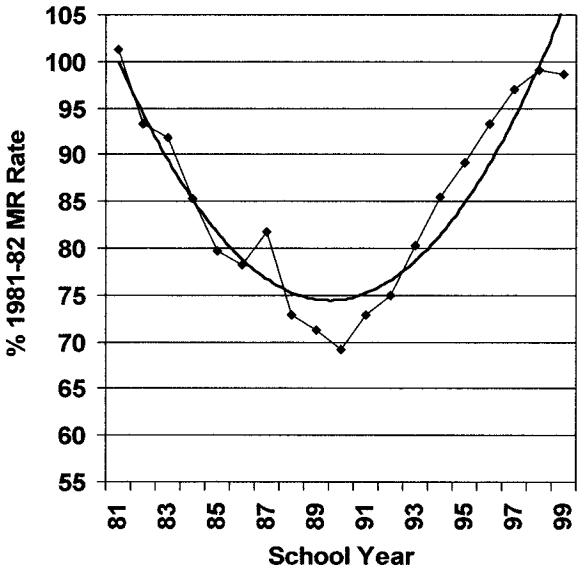

Figure 5. Graph of the change between 1981 and 1999 of the percentage of enrolled students in Nebraska who were in MR programs. ♦ = MR rate, ⎯ = Poly. (MR rate).

tion services dropping to less than 3/4 of it 1981 rate at its lowest point in the 1989 school year. Texas, the state with the second largest number of prisoners on death row, had a 1999 mental retardation rate that was only 58% of its 1981 rate, although its curve reflected more of a leveling off in the mid–1990s without a substantial rise. Florida, the state with the third largest number of prisoners on death row, had a curve with a shape very similar to California's.

In order to assess whether changes in IQ criteria at the state level might have affected the results, I surveyed over 200 school psychologists in all 50 states plus the District of Columbia about changes in eligibility criteria for mental retardation programs between 1980 and 1999 and received 78 responses from psychologists in all 50 states plus the District of Columbia. In 45 out of the 50 states and in the District of Columbia, psychologists reported no changes in their IQ criteria during this time. Psychologists in Iowa and Ohio reported that the IQ criteria in their states were lowered in the late 1990s (from 79 to 75 in Iowa and from 80 to 70 in Ohio), although this had no perceptible effect on data trends.

There were three clear instances where the Flynn effect may have interacted with state-level educational policy changes to result in trend lines that are stronger or weaker than the national average. In the mid-1990s, Nebraska raised its IQ criterion for mental retardation services from 70 to 75 in response to the American Association on Mental Retardation (AAMR) revised definition (Luckasson et al., 1992). As can be seen in Figure 5, Nebraska's increase in percentage of children receiving mental retardation services following the introduction of the WISC-III was particularly steep and brought Nebraska's 1999 mental retardation levels nearly in line with its 1981 levels. In the opposite direction, Idaho lowered its IQ criterion for mental retardation from 80 to 75 during the mid-1990s. Because this 5-point shift corresponds almost exactly to the size of the Flynn effect between the WISC-R and the WISC-III, the impact of the introduction of the WISC-III may have been mitigated. In Idaho, there was no rebound in the percentage of children receiving mental retardation services following the introduction of the WISC-III, although there was a slight increase at the end of the 1990s. During the mid-1980s, the Colorado legislature changed the IQ criterion for mental retardation from 1.75 *SD* below the mean (IQ 74) to 2 *SD* below the mean (IQ 70). This augmented the impact of the Flynn effect during the 1980s, resulting in a vertex in a mental retardation rate in 1995 that was only 34% of the rate in 1981. As expected due to the Flynn effect, the sharp decline in mental retardation rates reversed at this point.

The data analyses presented here raise some questions that cannot be answered at this point. For example, if changes in mental retardation classifications were primarily due to the Flynn effect, how long did it take before the maximum impact of the new norms was felt and the rates of classification start to drop again? There is some evidence that mental retardation classifications declined in the 3 years of data after 1999 (U.S. Department of Education, 2002; 2003; 2004). A visual examination of the trend lines shows that in the 3 years after 1999, 25 states and the United States as a whole had downward trend lines, 16 had upward trend lines, and 10 had indeterminate trend lines. Figure 5 shows the beginning of the leveling of the trend line for Nebraska.

After 1999, the developmental delay category began to have an impact on children's mental retardation classifications. The inclusion of children classified as having developmental delay in the group of children classified as having mental retardation had a minimal impact on the results reported here because it only affected 3 years, and in 1999, only .04% of 6- to 17-year-old children spread across 17 states were reported in the developmental delay category. This number tripled to .12% of this group spread across 29 states by 2002.

Three different state trends have become apparent in the reporting of children with developmental delay: Some states continued to report numbers of children receiving mental retardation services without including any children in the developmental delay category; other states placed children who previously would have been considered to have mental retardation into the developmental delay category; and, finally, other states developed programs to identify children with developmental delay, elevating the overall mental retardation plus developmental delay scores. For example, in 1997, Kentucky reported that 2.61% of school-age children were classified with mental retardation. In 2002, following the adoption of a developmental delay identification program, the percentage classified as having mental retardation was 2.49%, but the percentage of children classified with developmental delay was .94%, resulting in a marked increase in the number of children in the mental retardation plus developmental delay category.

Discussion

The replacement of the WISC-R with the renormed WISC-III in 1991 coincided with the end of a decade-long overall decline in mental retardation placements in most states and the District of Columbia. A substantial majority of states showed a decline in mental retardation rates from 1981 until the early 1990s, followed by an increase in rates. On average, among the 43 states plus the District of Columbia that showed this pattern, the decline in mental retardation placements ended in 1993 at 62% of the 1981 rate, followed by a rebound to 74% of this rate by 1999. Although the increase in popularity of the learning disability classification and further changes in educational policies may have attenuated the impact of the introduction of the WISC-III, they were not sufficient to counteract a substantial increase in the rate of mental retardation placements once the WISC-III was introduced.

This substantial longitudinal fall and rise in mental retardation rates, plus the enormous disparities at any one point in time among states in terms of the percentage of enrolled students receiving mental retardation services raises a number of questions with regard to whether mental retardation is adequately conceptualized (Flynn, 2000). These findings also pose a number of challenges in assessing the meaning of the "onset during the developmental period" criterion for mental retardation, which can have important legal and social implications beyond the developmental period.

Legal and Social Policy Implications of the Flynn Effect

A classification of mental retardation in adulthood can bring with it a number of important government benefits and legal rights. Individuals with mental retardation who are unable to maintain steady employment may obtain government assistance in the form of the Social Security Administration's (SSA) Supplemental Security Income (SSI) disability insurance as well as assistance in obtaining subsidized housing and vocational training (Parish, 2003). There is no professionally recognized requirement for a developmental period classification of mental retardation or developmental period IQs in the mental retardation range from childhood to establish mental retardation for these benefits. The guidelines for SSA disability insurance evaluations for mental retardation (Social Security, 2002, Section 12.05) were recently revised in order to correct the misperception that well-documented evidence of meeting

developmental period criteria for a classification of mental retardation is required in order to be eligible for benefits.

However, the potential underclassification of mental retardation due to dated test norms and other reasons discussed in the introduction raises an important question about the extent to which children and adults (especially urban dwellers and minorities) who should be receiving government SSI and disability insurance assistance for mental retardation are not benefiting from these services or are even aware that they might qualify for these services. Supplemental Security Income or disability insurance eligibility are typically prerequisites for some state-run programs, so not having these benefits may make it harder for individuals with low IQ to receive appropriate job training or qualify for subsidized housing (Parish, 2003). Such individuals have a great deal of difficulty finding and maintaining employment and often live in poverty (Gottfredson, 1997), and their inability to earn a subsistence living through ordinary means may contribute to their overrepresentation in the prison population. People with mild mental retardation often do not fit the public's stereotypes about how individuals with mental retardation should look or behave, so the extent of their cognitive impairments may not become apparent to caseworkers or lawyers working on their behalf unless they are formally evaluated.

There is one legal area where the establishment that a person meets the developmental period criterion has become a matter of life or death. In June of 2002, the Supreme Court held in *Atkins v. Virginia* that "the execution of people with mental retardation constituted 'cruel and unusual punishment' under the eighth amendment." In capital murder cases, the burden of proof for establishing whether a defendant meets all of the criteria for mental retardation falls upon the defendant. Thus, a defendant's intelligence testing history may play a significant role in death penalty qualification determinations. However, establishing whether a defendant meets the developmental period criterion for mental retardation is not always a simple task, because many adults who currently meet the IQ and poor adaptive functioning criteria necessary for being classified with mental retardation may have never received a formal developmental period classification. In a recent decision, the United States Circuit Court of Appeals for the Fourth District held that a lower court must take into consideration the persuasiveness of evidence that the Flynn effect had an impact on the appellant's sole surviving developmental period WISC-R score, because under Virginia law a relevant issue was whether the appellant scored two *SD*s below the mean on an intelligence test (*Walker v. True,* 2005). Understanding the impact of the Flynn effect on IQs is especially relevant for death penalty cases in which the burden of providing the evidence for mental retardation falls on the defense. Although none of the major definitions of mental retardation explicitly require an IQ in the mental retardation range or a formal classification of mental retardation during the developmental period, the presence of these factors can make the case for mental retardation more compelling.

Conclusions

In this study I found that the end of the decline in number of children receiving mental retardation services during the 1980s occurred around the time of the introduction of the WISC-III. Because of the obsolescence of intelligence test norms over time and the increasing reluctance of some school districts to assign a classification of mental retardation (especially with regard to minority children), numerous borderline individuals who were tested but never placed may in fact meet all of the criteria for mental retardation under one or more of the commonly accepted definitions.

Although I focused on intrastate variability in placements for mental retardation programs, the large variability in these placement rates raises important questions about differences in diagnostic procedures among states. These questions include the extent to which classification rates vary by race and class, the degree to which IQs are used as a key criterion for mental retardation placements, and whether there is pressure from local and state school administrators to reduce or increase the number of children eligible for special services. Although I compared state and national trend lines to a hypothetical Flynn effect only model, the similarities between the models cannot prove that the Flynn effect was the only or even the most important factor affecting placement decision rates. Furthermore, intelligence test developers have responded to the phenomenon of rising intelligence test scores by renorming their tests more frequently. Although there was a 17-year gap between the norming of the WISC-R and the WISC-III, there was only a 13-year gap between

the norming of the WISC-III and the WISC–IV. Thus, because of the use of more up-to-date norms on intelligence tests, the Flynn effect may not be as important a consideration in future mental retardation classifications.

An important research question that cannot be resolved using these findings is whether or not the Flynn effect represents a "true" rise in actual intelligence over time. If the rise in IQ represents a real increase in overall competence, then renorming gives everyone scores that more accurately reflect their performance relative to the norming sample. However, if a newly normed intelligence test causes numerous children to be put into mental retardation programs who would have done fine without placement, then this raises questions about the use of an arbitrary cut-off score as an IQ criterion. Flynn (2000) suggested that testmakers periodically assess a group of children classified with mental retardation on behavioral grounds by respected psychologists and use these data for developing test criteria for mental retardation. These assessments would need to be repeated every 7 years or so in order to determine whether the criteria had stayed constant or drifted over time.

An implicit assumption in this paper, and in the National Research Council's (Reschly, Myers, & Hartel, 2002) recommendations for revising guidelines for mental retardation SSI benefits, is that the IQ criterion for mental retardation is a norm-referenced phenomenon and that the best indicators of mental retardation in adulthood are an IQ in the mental retardation range, poor concurrent adaptive functioning, and a history consistent with having mental retardation that may or may not include developmental period intelligence tests in the mental retardation range. Although the guidelines of the Social Security Administration (2002) were recently revised to make it clearer that developmental period IQs are not required to establish eligibility for disability insurance benefits, other definitions (such as those provided by the American Psychiatric Association and AAMR) should be revised accordingly.

In summary, the fact that an individual with a low IQ was not diagnosed with mental retardation during childhood is not necessarily indicative of not having mental retardation, but may have been the result of factors specific to the year, state, or school district in which he or she attended school that reduced the rate of mental retardation diagnoses. In capital murder cases, where the criteria for meeting the IQ prong (either current or developmental period) is 2 *SD*s below the population mean, fairness would dictate that IQs be adjusted to compensate for IQ gains. A defendant's fate should not rest upon whether or not an intelligence test was administered to him or her shortly after it was normed (and be more likely to score in the mental retardation range) or long after it was normed (and be more likely to score above the mental retardation range). As Ceci, Scullin, and Kanaya (2003) recently argued, the question of who is considered to have mental retardation is complex and may affect dozens of inmates currently on death row as well as hundreds of thousands of individuals with mild mental retardation who may not be aware they are eligible for various forms of government assistance or services.

References

American Psychiatric Association. (1994). *Diagnostic and statistic manual* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistic manual* (4th ed., text rev.). Washington, DC: Author.

Atkins v. Virginia, 534 U.S. 304 (2002).

Bolen, L. M., Aichinger, K. S., Hall, C. W., & Webster, R. E. (1995). A comparison of the performance of cognitively disabled children on the WISC-R and WISC-III. *Journal of Clinical Psychology, 51,* 89–94.

Ceci, S. J., Scullin, M. H., & Kanaya, T. (2003). The difficulty of basing death penalty eligibility on IQ cut-off scores for mental retardation. *Ethics and Behavior, 13,* 11–15.

Education of All Handicapped Children Act of 1975, Pub. L. No. 94–142. 20 U.S.C.S. § 1400 *et seq.*

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains from 1932 to 1978. *Psychological Bulletin, 95,* 29–51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What intelligence tests really measure. *Psychological Bulletin, 101,* 171–191.

Flynn, J. R. (1998). WAIS-III and WISC-III: IQ gains in the United States from 1972 to 1995: How to compensate for obsolete norms. *Perceptual & Motor Skills, 86,* 1231–1239.

Flynn, J. R. (2000). The hidden history of IQ and special education: Can the problems be solved? *Psychology, Public Policy, and Law, 6,* 191–198.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79–132.

Gottlieb, J., Alter, M., Gottlieb, B. W., & Wishner, J. (1994). Special education in urban America: It's not justifiable for many. *Journal of Special Education, 27,* 453–465.

Individuals With Disabilities Education Act Amendments of 1997, Pub. L. No. 105-17, 20 U.S.C.S. § 1400 *et seq.*

Individuals With Disabilities Education Act (as amended by the Education of all Handicapped Children Act Amendments of 1990), Pub. L. No. 101-476, 20 U.S.C.S. § 1400 et seq.

Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence: A handbook for professional practice.* Needham Heights, MA: Allyn & Bacon.

Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn effect and U.S. policies: The impact of rising IQs on American society via mental retardation diagnoses. *American Psychologist, 58,* 778–790.

Larry P. v. Riles, 342 F. Supp. 1306 (N.D. Cal. 1972) (preliminary injunction) affirmed 502 F 2d 963 (9th cir. 1974); 495 F. Supp 926 (N.D. Cal 1979) (decision on merits) affirmed (9th cr. No. 80-427), January 23, 1984. Order modifying judgment, C-7-2270 RFP, September 25, 1986.

Luckasson, R., Borthwick-Duffy, S., Buntinx, W. H. E., Coulter, D. L., Craig, E. M. Reeve, A., Schalock, R., Snell, M. E., Spitalnik, D. M., Spreat, S., & Tassé, M. J. (2002). *Mental retardation: Definition, classification and systems of supports.* Washington, DC: American Association on Mental Retardation.

MacMillan, D. L., Gresham, F. M., Siperstein, G. N., & Bocian, K. M. (1996). The labyrinth of IDEA: School decisions on referred students with subaverage general intelligence. *American Journal on Mental Retardation, 100,* 533–545.

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.

Neisser, U. (Ed.). (1998). *The rising curve: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Parish, S. L. (2003). Federal income payments and mental retardation: The political and economic context. *Mental Retardation, 41,* 446–459.

Reschly, D. J., Myers, T. G., & Hartel, C. R. (Eds.). (2002). *Mental retardation: Determining eligibility for social security benefits.* Washington, DC: National Research Council, National Academy Press.

Sanborn, K. J., Truscott, S. D., Phelps, L., & McDougal, J. L. (2003). Does the Flynn effect differ by IQ level in samples of students classified as learning disabled? *Journal of Psychoeducational Assessment, 21,* 145–159.

Slate, J. R., & Saarnio, D. A. (1995). Differences between WISC-III and WISC-R IQs: A preliminary investigation. *Journal of Psychoeducational Assessment, 13,* 340–346.

Social Security Administration. (2002). *Disability evaluation under Social Security.* Washington, DC: U.S. Government Printing Office.

Truscott, S. D., & Frank, A. J. (2001). Does the Flynn effect affect IQs of students classified as learning disability? *Journal of School Psychology, 39,* 319–334.

Truscott, S. D., & Volker, M. (2005). The Flynn effect and learning disability classification: Empirical evidence of IQ score changes that could affect diagnosis. In A. Columbus (Ed.), *Advances in psychology research* (pp. 173–204). Hauppauge, NY: Nova Science.

U.S. Department of Education. (1983–1990). *Annual report to Congress on the implementation of the Education of the Handicapped Act.* Washington, DC: Author.

U.S. Department of Education. (1991–2004). *Annual report to Congress on the implementation of the Individuals With Disabilities Education Act.* Washington, DC: Author.

Vance, H., Maddux, C. D., Fuller, G. B., & Awadh, A. M. (1996). A longitudinal comparison of WISC-III and WISC-R scores of special education students. *Psychology in the Schools, 33,* 113–118.

Walker v. True, No. 04-016 (4th Cir. February 17, 2005).

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised manual.* New York: Psychological Corp.

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children-III manual.* New York: Psychological Corp.

Received 4/6/05, accepted 1/13/06.
Editor-in-charge: Frances Conners

The author thanks Stephen J. Ceci and Tomoe Kanaya for their invaluable feedback on this paper and Stanley Cohen for his statistical advice. Appreciation is also extended to Ulric Neisser, Nathan Brody, and Linda Gottfredson for their comments on an earlier version of this paper, and three anonymous reviewers for their feedback on the current version of the paper. Requests for reprints should be sent to Matthew H. Scullin, Department of Psychology, The University of Texas at El Paso, 500 W. University Ave., El Paso, TX 79968. E-mail: mhscullin@utep.edu

