Institute of Education Sciences
National Center for Education Statistics

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

WHITE PAPER

PROJECTIONS OF EDUCATION STATISTICS: PRESENTATION AND METHODOLOGY

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

WHITE PAPER

PROJECTIONS OF EDUCATION STATISTICS: PRESENTATION AND METHODOLOGY

EXECUTIVE SUMMARY

This white paper is an extended review of NCES' annually issued report containing projections, exemplified by Hussar and Bailey (2008), from methodological (§3) as well as presentation (§2) perspectives.

While the paper contains criticisms, they are meant to be constructive, and are in no sense criticisms of the authors. Especially in the discussion of presentation, every criticism is accompanied by at least one suggested alternative.[1] None of these alternatives would be difficult to implement, except to the extent that some of them may conflict with the NCES statistical standards (National Center for Education Statistics, 2004).

The discussion of projection methodology, by contrast, is critical without detailed consideration of alternatives. This reflects the magnitude of the effort that would be needed to develop new methodology, which is discussed in §3.3.

One very broadly applicable comment is important for both current and future editions. Although Hussar and Bailey (2008) is explicit about omissions, the *collective effect of these omissions is large and increasing*. Examples noted in Hussar and Bailey (2008) include home schooling (page 1) and high school completers by means other than diplomas granted by school authorities (page 11), postsecondary enrollment in non-degree-granting institutions (implicitly on page 9) and possibly others. Additional examples that may not be addressed include distance learning and US citizens enrolled in institutions outside the US.[2] It would be very valuable to list these omissions in one prominent place. More importantly, in the longer run NCES should devote effort to addressing them, since otherwise the information in successors to Hussar and Bailey (2008) will be increasingly incomplete.

[1] For expediency, virtually all of these were produced using some combination of Microsoft Excel, Microsoft Visio and Adobe Photoshop. Versions produced by professional graphics software would be even better.

[2] It is not clear, for instance, whether data in Hussar and Bailey (2008) include schools operated by the US Department of Defense.

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

WHITE PAPER

PROJECTIONS OF EDUCATION STATISTICS:
PRESENTATION AND METHODOLOGY

I. PRESENTATION OF PROJECTIONS

Hussar and Bailey (2008) is very complete, and appears to contain relatively few outright errors.[3] It is not, however, an especially user-friendly document. A relatively small number of changes would increase both the amount of information conveyed to readers *and* the clarity with which that information is transmitted. We describe these changes, none of which is esoteric, in this section. All of this section other than §2.4 addresses presentation of projections in a printed document. Some opportunities associated with interactive, web-based presentation are discussed in §2.4. 1997). Wilkinson's Grammar of Graphics (Wilkinson, 2005) does underlie many of the comments.

1.1 Graphics

This section is intended to be extremely concrete. Therefore, there are no mentions of the extremely intriguing and important but often somewhat philosophical tenets of Tufte (Tufte, 1983, 1990, 1997). Wilkinson's *Grammar of Graphics* (Wilkinson, 2005) does underlie many of the comments.

1.1.1 Principal Items

A dramatic improvement to presentation would be to *replace vertical bar charts by horizontal bar charts*.

For concreteness, consider Figure A of Hussar and Bailey (2008),[4] which is inefficient because the width of the bars there is determined by the need to put numbers above them. Consider instead Figure 1, which is a horizontal version; by comparison with Figure A of Hussar and Bailey (2008),

1. The expanded physical scale makes comparisons easier.
2. The horizontal version reveals more about small values. See also discussion of the *disparate scales issue* below.
3. The horizontal version includes both actual values and percentage changes from one time period to the next.
4. Year labels are explicit.

Note that Figure 1 contains low-key vertical lines, beneath all other graphical elements, that carry the numerical scale associated with the x -axis through the entire chart. These are much easier to follow than the tick marks in Figure A of Hussar and Bailey (2008).

---

[3] But, see discussion in §2.3.1.

[4] Throughout this section, we illustrate each problem, as well as potential solutions, using only one figure in Hussar and Bailey (2008). It is straightforward to determine which other figures share the same problem and are amenable to the same solutions.

In regard to the general comment in §1, it appears (cf. page 1 of Hussar and Bailey (2008)) that 1992 and 2005 values in Figure A include home-schooled students, but that 2017 projections do not. It is not even clear that the statement on page 1 is correct, since projected enrollments exceed current ones by reasonable numbers. Figure A is deceptive if home-schooled students are in only two of the three sets of values.

A pervasive issue in Hussar and Bailey (2008) is the *treatment of totals*. In Figure 1, the "Total" bars are the sums of the "PK-8" and "9-12" bars. At some level, this is perfectly clear, but at the same time, the principle that it be clearly indicated when some elements of a chart are derived from others is violated. Figure 2 removes this problem by means of a bi-directional ("back-to-back") bar chart. There, PK-8 enrollments are to the left of the light gray vertical line at "Enrollment = 0" and 9-12 enrollments to the right. By comparison with Figure 1, Figure 2 is superior in three senses:

- It makes explicitly clear that "Total" is the sum of PK-8 and 9-12.
- It conveys the same information in approximately one-third less space.
- It improves comparisons between PK-8 and 9-12. For instance, it is clear in Figure 2 that the rate of PK-8 growth is increasing, whereas the rate of 9-12 growth is decreasing.

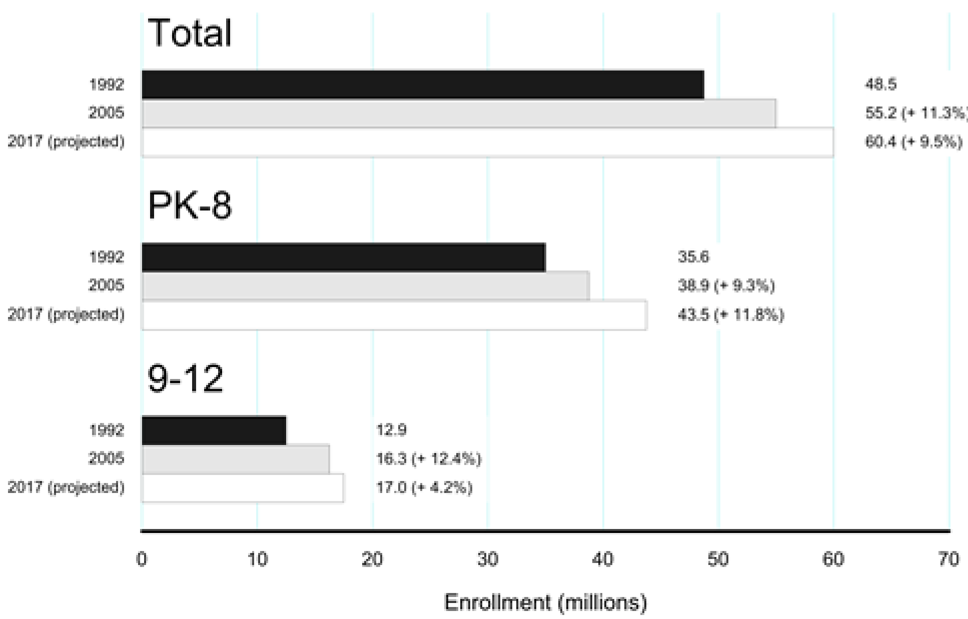


Figure 1: Alternative version of Figure A in Hussar and Bailey (2008) in the form of a horizontal bar chart.

On the other hand, the capability for direct graphical comparisons between totals is attenuated in Figure 2 as compared to Figure 1. For instance, it is apparent from the latter but not the former that the rate of growth of total enrollment is decreasing.

Graphics such as that in Figure 2 should not be employed in cases where the total of the "left" and "right" sides makes no sense, as exemplified by Figure K of Hussar and Bailey (2008).

Throughout, Hussar and Bailey (2008) deals poorly with *disparate numerical values*. The multi-part[5] Figure D is an example of the problem. No value in it exceeds 18, but the scale runs from 0 to 30. Graphical information about small values, notably in the fourth and sixth panels, is obliterated.

The underlying rationale is clear and sensible - preserving a common scale across the multiple panels of the figure. A high price is paid, however, which is avoided in the horizontal bar chart in Figure 3. While this alternative is problematic in other senses - it may contain too much information for some users,[6] it makes small values much more visible. Moreover, it permits cross- comparisons (for example, males to 18-24 year-olds) that are impossible in Figure D of Hussar and Bailey (2008).

Abandoning the “preserve common scale” principle is, of course, possible.

In addition, and independent of the disparate values problem, Figure D of Hussar and Bailey (2008) is misleading. The four panels appearing on page 9 correspond to distinct categorizations, whereas the two panels on page 10 represent two components of a single categorization. Neither is it clear why the public-private categorization is presented separately, in Figure E of Hussar and Bailey (2008), from the others.

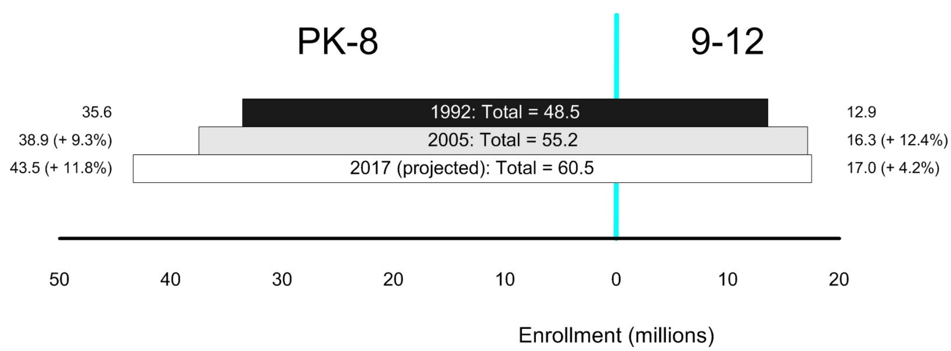


Figure 2: Alternative version of Figure A in Hussar and Bailey (2008) in the form of a bi-directional horizontal bar chart.

Hussar and Bailey (2008) is replete with *low content figures*. Figure J is a prime example: it consumes approximately 10% of a page in order to display only 3 values! The alternative in Figure 4 presents three times as much information: the coordinates of the endpoint of each line are the (Teacher, Pupil) numbers, and the slope of each line is the Pupil/Teacher ratio. The increasing rate of decline in the ratio is evident from the concavity in Figure 4, but hard to discern in Figure J of Hussar and Bailey (2008). Figure 4 has deficiencies, especially the skewed aspect ratio, but these do not interfere with its ability to communicate multiple pieces of information.

Figure 5, which is included to illustrate the range of possibilities, has strengths as well as significant weaknesses. In it, the number of teachers is encoded as the height of each rectangle and the Pupil/Teacher ratio as the width of each rectangle, so that the number of pupils is the area of the *rectangle*. This figure is very revealing about changes in the numbers of pupils and the ratio, but not - because humans are not adept at translating perceived areas to numerical values - the numbers of pupils. The weakness of this

[5] Which some would already consider to be a violation of the principles of good graphics.

[6] In addition, the labeling is poor, year is not coded into the shading of the bars and there are extraneous tick marks on the vertical scale. These result from its having been produced using Microsoft Excel, and could easily be remedied.

figure is inconsistency: one numerical value (teachers) is encoded as a length, but the other numerical value (pupils) is encoded as an area. Figure 4 does not have this problem: both numerical values are encoded as lengths, and their ratio is encoded - mathematically consistently - as a slope.

1.1.2 Additional Items

The items discussed here are important, and fixing them is both straightforward and non-controversial.

First, the *year-to-shade/color encoding* should be consistent across all figures. The predominant encoding in Hussar and Bailey (2008) is

1992 light blue: RGB values (127,179,210)

2005 mid blue: RGB values (0,85,165)

2017 (projected) white,

which translates acceptably to gray-scale hard copy.

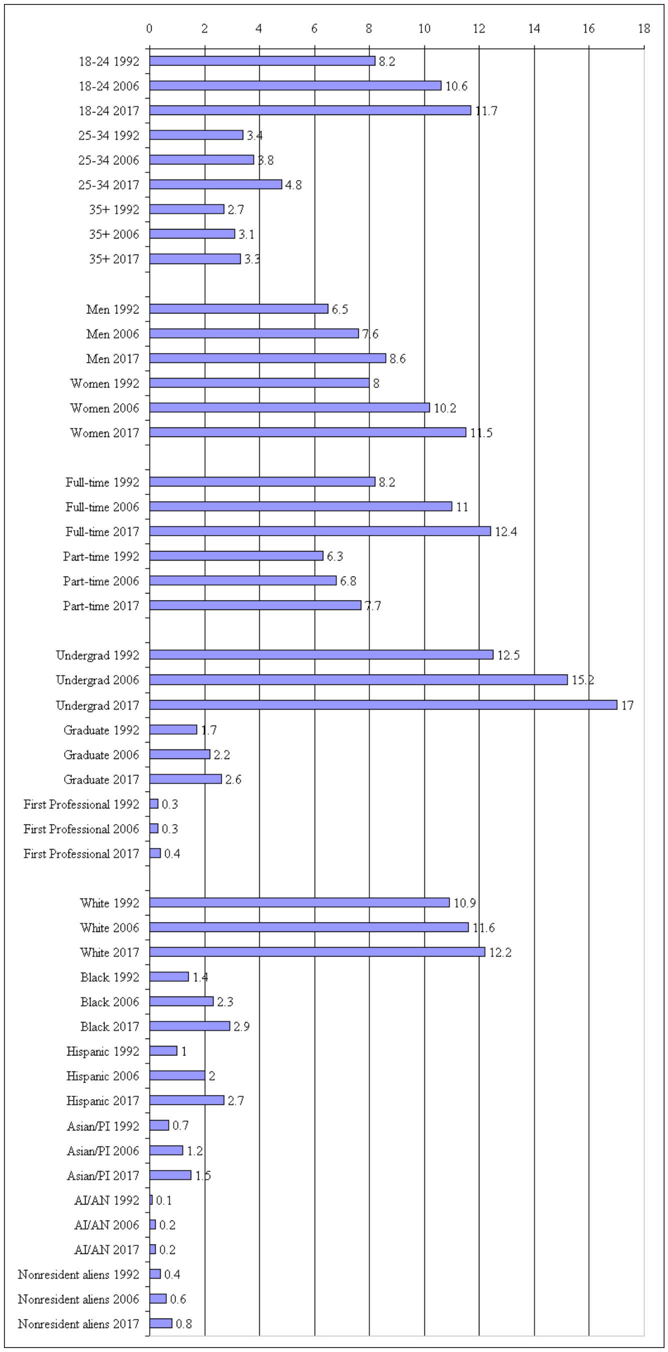


Figure 3: Alternative version of Figure D in Hussar and Bailey (2008).

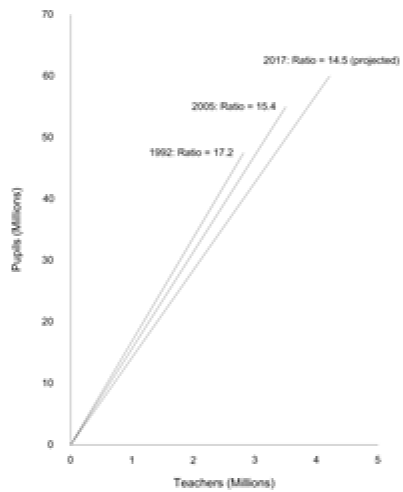


Figure 4: Alternative version of Figure J in Hussar and Bailey (2008).

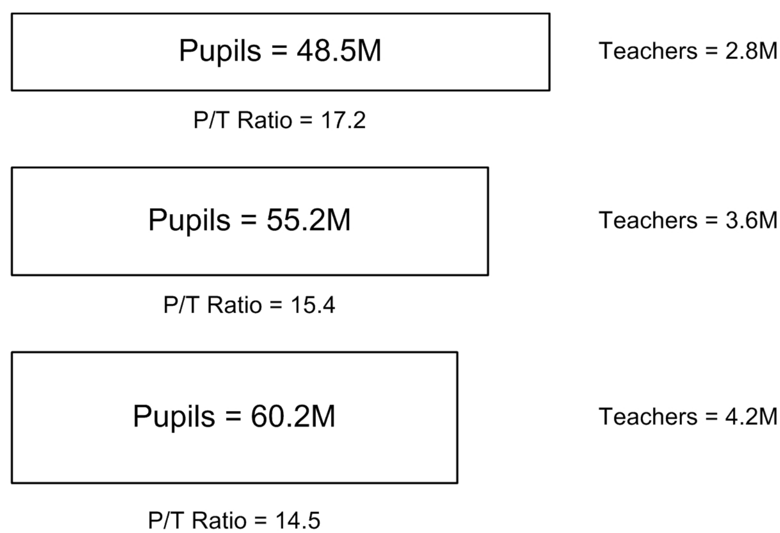


Figure 5: Another alternative version of Figure J in Hussar and Bailey (2008).

Year	Pupils (Figure A)	Students (Figure H)	Calculated Ratio	Reported Ratio (Figure J)
1992	48.5	2.8	17.321	17.2
2005/2006	55.2	3.6	15.417	14.3
2017	60.4	4.2	14.381	14.5

Table 1:

Figures C, H, L and M of Hussar and Bailey (2008), among others, violate this scheme. Yet another shading is used in some of the reference figures, for instance, Figure 11. In graphs such as Figure 1 (page 23), the blue color is used (not very effectively in color displays or hard copy, and uselessly in gray-scale hard copy) for yet another purpose - to distinguish between actual and projected values. That the same colors are used in maps, where there is graphical shading as well, does not seem to cause problems.

Second, there are *inconsistencies in the years* appearing in the graphics. For example, the years in Figure A of Hussar and Bailey (2008) are 1992, 2005 and 2017, while those in Figure F are - under the convention that NNNN refers to academic year *starting* in calendar NNNN - 1992, 2004 and 2017. Figure G of Hussar and Bailey (2008) also illustrates the problem.

Third, some things that may have clear explanations are puzzling. The Pupil/Teacher ratios in Figure J of Hussar and Bailey (2008) seem to be derived from the numbers of pupils in Figure A and teachers in Figure H.[7] However, as Table 1 shows, correct reproduction of the calculation is not possible. The values in Figure J are not ratios of those in Figures A and H, presumably because the ratios in Figure J were calculated from unrounded (or less severely rounded than in Figures A and H) numbers of pupils and teachers. It would be useful to note and explain these kinds of anomalies.

Fourth, the Reference Figures present several problems:

1. As noted above, the use of color and a virtually indiscernible increase in line thickness to distinguish actual from projected values is ineffective in gray-scale hard copy. Figure 6 illustrates one alternative for Figure 1 of Hussar and Bailey (2008). Incidentally, it appears that projected values should start with the year labeled 2006, not 2005. This is done in Figure 6 but not in Figure 1 of Hussar and Bailey (2008).
2. Some Reference Figures (examples: Figures 1, 8 and 9 of Hussar and Bailey (2008)) contain graphs that representing totals, while others (examples: Figures 3 and 4) do not. There is no evident reason for this inconsistency, especially given that there are exact parallels, for instance, between Figure 1 and Figures 8 and 9.

Figure 1. Actual and projected numbers for school-age populations, by age range: 1992 through 2017

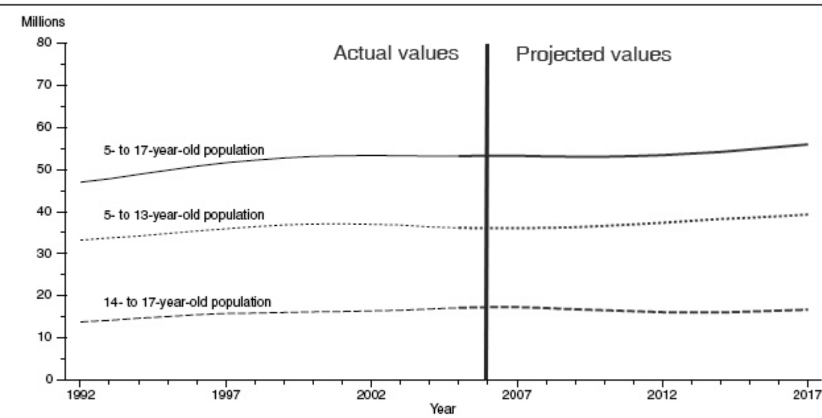


Figure 6: Alternative version of Figure 1 of Hussar and Bailey (2008) in which actual and projected values are distinguished more clearly.

3. Although it may be of interest to only a small number of readers, some discussion of the smoothing used to convert discrete data to graphs such as Figure 1 would be useful.
4. Perhaps more important, the very need for "continuous graphs" is obscure. Figure 7 contains more information than either of Figures 2 and 3 in Hussar and Bailey (2008) - and in fact more information than the two together, and at the same time it conveys at least as much visual gestalt.

[7] Something that would be helpful to readers of Hussar and Bailey (2008) to know, but is never stated.

1.2 Maps

Maps constitute a relatively minor part of Hussar and Bailey (2008), so the comments in this section are not extensive. Some of them are underlain by MacEachern (1995).

1. All maps display numerical values at the state level via generalized shading of states, but the shading scheme is linked only loosely to those values, in the sense that higher values correspond to darker/more complete shading. The yellow-brown heat scale in Pickle et al. (1997) is much more effective.[8] Nor does the shading scheme differentiate effectively be- tween increases and decreases. Even though rudimentary, Figure 10 does so much more effectively.

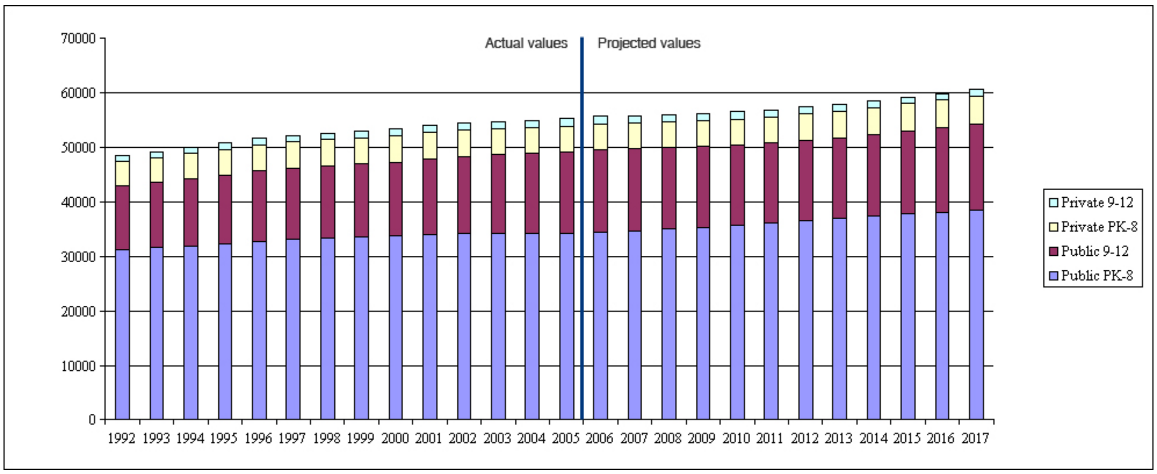


Figure 7: Alternative version of Figures 2 and 3 in Hussar and Bailey (2008) that contains more information than those two figures combined.

2. The legends are flawed. For instance, how would a decrease of 4.95% be displayed?
3. Figures 5,6 and 7 in Hussar and Bailey (2008) display projections, yet it is not stated which projections.

1.3 Tables

The principal function of the tables in Hussar and Bailey (2008) appears to be completeness in support of access to individual values. Nevertheless, improvements are possible that convey higher-level information more effectively without compromising this basic purpose.

1.3.1 Principal Items

First, the *dotted lines that appear in virtually every table are space-consuming, visually unattractive and less effective than alternatives* such as that in Figure 8, which is a version of Table B-4 of Hussar and Bailey (2008).[9] The shading for alternating years is unintrusive, yet distinguishes years perfectly. This table is physically smaller than Table B-4 of Hussar and Bailey (2008), and as well, the distance between labels and data is smaller.

[8] Although there seems to be no need for county-level information in NCES' projection reports, the maps in Pickle et al. (1997) display county-level information easily, whereas the cross–hatching in Hussar and Bailey (2008) fails badly at higher geographical resolution.

[9] Produced manually using Adobe Photoshop; Microsoft Excel would have done essentially the same thing.

Second, *some tables would be more effective as graphics*. Tables A and B of Hussar and Bailey (2008) are a clear example. Before proceeding, we note that the justification for placing states with increases in projected enrollment and those decreases in separate tables is elusive. A person trying to find a specific state of interest is forced to look at two tables rather than one. Sorting the tables by level of increase or decrease also inhibits finding specific states.

Year (July 1)	18-year-olds	18- to 24-year-olds	25- to 29-year-olds	30- to 34-year-olds	35- to 44-year-olds
Actual					
1992	3,354	26,282	20,591	22,564	40,046
1993	3,455	26,102	20,146	22,646	40,975
1994	3,428	25,821	19,809	22,648	41,877
1995	3,601	25,585	19,742	22,425	42,765
1996	3,650	25,376	19,927	21,996	43,605
1997	3,780	25,574	19,960	21,494	44,282
1998	3,984	26,155	19,863	20,999	44,802
1999	3,993	26,780	19,632	20,647	45,130
2000	4,076	27,393	19,357	20,579	45,235
2001	4,074	28,087	19,004	20,781	45,188
2002	4,033	28,601	18,997	20,878	44,869
2003	4,131	29,094	19,213	20,789	44,484
2004	4,128	29,408	19,625	20,528	44,178
2005	4,127	29,500	20,148	20,153	43,954
2006	4,190	29,610	20,800	19,764	43,748
Projected					
2007	4,272	29,809	21,313	19,713	43,379
2008	4,401	30,173	21,672	19,865	42,782
2009	4,384	30,536	21,878	20,213	42,109
2010	4,312	30,762	21,944	20,657	41,600
2011	4,250	30,894	21,981	21,205	41,318
2012	4,170	30,947	22,057	21,652	41,217
2013	4,126	30,884	22,205	22,000	41,222
2014	4,080	30,693	22,459	22,202	41,258
2015	4,007	30,297	22,783	22,271	41,270
2016	3,990	29,901	23,059	22,313	41,421
2017	4,018	29,607	23,260	22,394	41,754

Figure 8: Alternative version of Table B-4 of Hussar and Bailey (2008) in which dotted lines are replaced by shading of alternating rows.

There are at least three alternatives. One is the bar chart in Figure 9.[10] It contains the same detailed information as Tables A and B in Hussar and Bailey (2008), but has several advantages:

- Sorting by state name facilitates finding specific states.
- A visual sort by magnitude of increase or decrease is possible, at least for the largest magnitudes.
- The relative numbers of states with increases and those with decreases are clear.
- It is apparent that many projected increases exceed most projected decreases.

A map is another alternative, but if numerical values were to be included, they would need to be plotted within states, which can be problematic, especially for the New England states. Figure 10 is an extremely rudimentary,[11] manually prepared[12] map that conveys some of the power of this alternative:

- Accessibility of information for specific states is as high as in Figure 9, and higher than in Tables A and B of Hussar and Bailey (2008).
- The geographical structure of the decreases in projected enrollment "jumps out" of Figure 10.

[10] Produced by Microsoft Excel.
[11] For instance, no values are shown for Alaska, Hawaii or the District of Columbia.
[12] Using Adobe Photoshop.

WARNING: This is true only of the color version of this figure.

However, Figure 10 is weaker than Figure 9 with respect to visual comparison of the magnitudes of increases and decreases.

A final alternative is an interactively sortable version of Figure 9. See §2.4 for further discussion.

Third, some tables exhibit *poor formatting*. There are alternatives that not only are more usable but also allow detection of errors. Consider Table 10 (page 53) of Hussar and Bailey (2008). In this table, it is virtually impossible to compare middle, low and high alternative projections for a given year, given the middle-low-high order in which they appear. Table 2 is an alternative version of part of this table, in which projections are much more readily compared. This alternative form also shows that the two entries in boldface seem to be incorrect.

It is true that it would take three versions of Table 2 to replace Table 10 in Hussar and Bailey (2008). However, since the report is already large, the gain in clarity may offset the increase in length.

Fourth, *totals in tables are conventionally at the bottom and right side*, while in many tables in Hussar and Bailey (2008), totals are at the top and left. Totals appear at the right in Table 2.

The typographical conventions of placing totals in **boldface** is effective, but is employed only sporadically, for example, in Table 13. The use of indenting to distinguish totals from components, as in Table 11, is not effective. There, the grand total is at the same level of indentation as the age components.

Figure 11 is a version of Table 33 in Hussar and Bailey (2008) that incorporates several of the alterations proposed in this section:

- Dotted lines are replaced by shading alternate rows.
- All projections for a given year appear in one row.
- The most aggregated figure - representing public *and* private schools - appears at the right. (In Table 33 of Hussar and Bailey (2008), these combined values are inaccurately and mis- leadingly labeled "Total.")

This version reveals a question that seems to deserve comment in the report: why are projected pupil/teacher ratios *highest* under the high alternative projection?

1.3.2 Minor Items

Some additional comments:

1. Tables 8 and 9 of Hussar and Bailey (2008) exemplify another confusion. In the former, the regional values are sums of state values, but in the latter they are not. Nevertheless, both tables have exactly the same physical format.

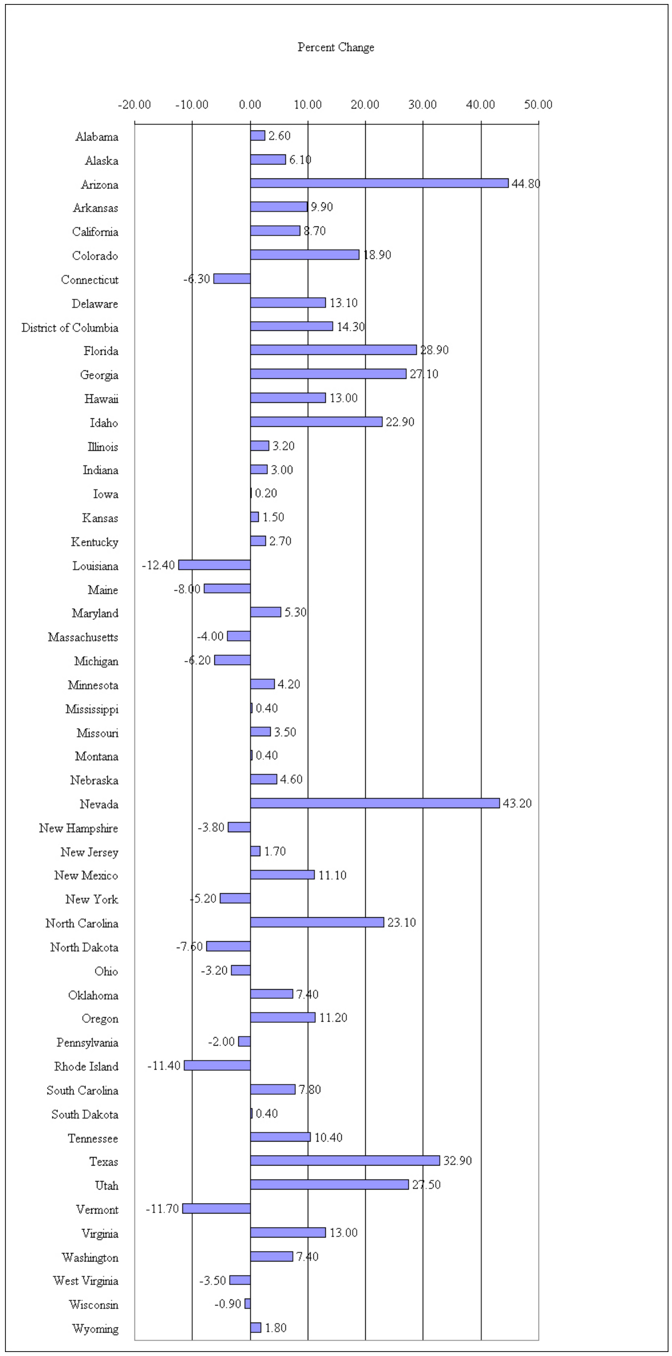


Figure 9: Graphical version of Tables A and B in Hussar and Bailey (2008).

Year	Men			Women			TOTAL		
Actual									
1992		6,524			7,963			14,486	
1993		6,427			7,877			14,305	
1994		6,372			7,907			14,279	
1995		6,343			7,919			14,262	
1996		6,353			8,015			14,368	
1997		6,396			8,106			14,502	
1998		6,369			8,138			14,507	
1999		6,491			8,301			14,791	
2000		6,722			8,591			15,312	
2001		6,961			8,967			15,928	
2002		7,202			9,410			16,612	
2003		7,260			9,651			16,911	
2004		7,387			9,885			17,272	
2005		7,456			10,032			17,487	
2006		7,575			10,184			17,759	
Projected	Low	Middle	High	Low	Middle	High	Low	Middle	High
2007	7,709	**7,704**	7,719	10,265	10,271	10,314	17,974	17,976	18,033
2008	7,829	**7,822**	7,850	10,353	10,378	10,454	18,182	18,200	18,304
2009	7,898	7,929	7,965	10,372	10,487	10,580	18,271	18,416	18,544
2010	7,957	8,022	8,071	10,397	10,590	10,714	18,354	18,613	18,785
2011	8,018	8,118	8,183	10,433	10,704	10,866	18,452	18,822	19,049
2012	8,088	8,213	8,296	10,509	10,835	11,041	18,597	19,048	19,337
2013	8,161	8,306	8,407	10,623	10,993	11,243	18,784	19,299	19,650
2014	8,227	8,387	8,499	10,742	11,146	11,426	18,969	19,533	19,924
2015	8,271	8,443	8,654	10,840	11,273	11,580	19,111	19,716	20,145
2016	8,318	8,500	8,634	10,934	11,393	11,734	19,252	19,893	20,368
2017	8,366	8,568	8,717	11,028	11,512	11,889	19,404	20,080	20,606

Table 2: Alternative version of a portion of Table 10 in Hussar and Bailey (2008). This alternative suggests that the entries in BOLDFACE may not be correct.

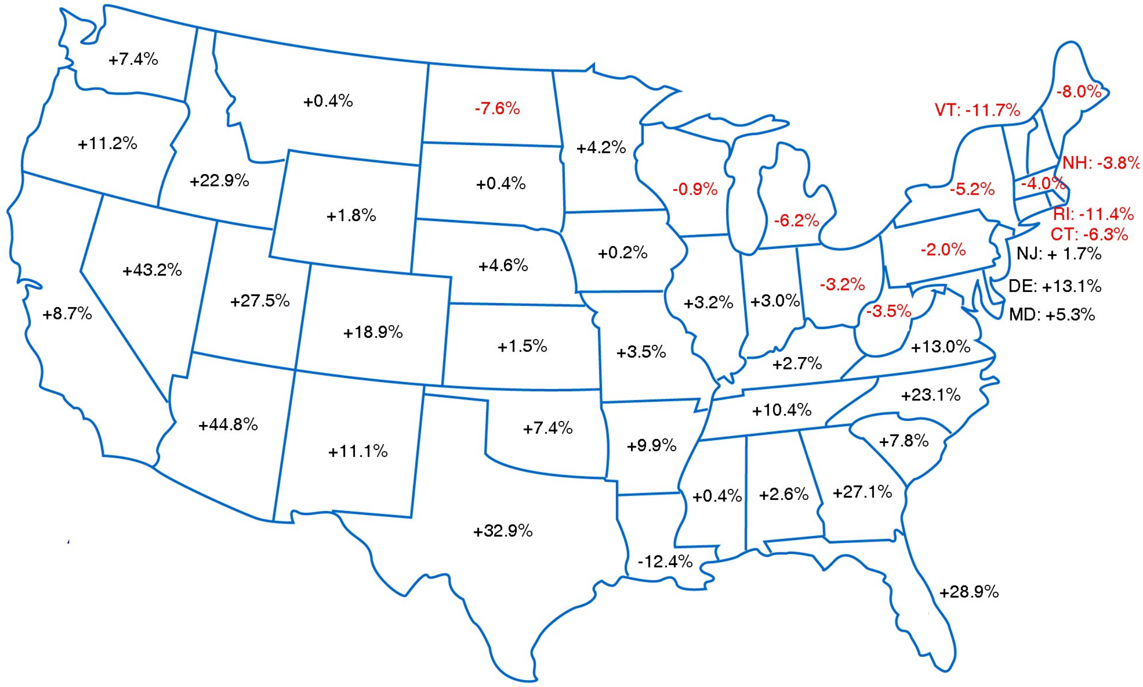


Figure 10: Prototype map version of Tables A and B in Hussar and Bailey (2008).

2. It is difficult to discern that Table 11 in Hussar and Bailey (2008) contains three distinct breakdowns of the same set of totals. Blank lines between the breakdowns are a simple but effective way of conveying this.
3. Some tables are broken across non-facing pages - that is, the first part of the table is on an odd-numbered page and the second on the following even numbered page. It is impossible to look at the entire table at once.
4. Table B-5 of Hussar and Bailey (2008) (page 130) contains some figures rounded/truncated to thousands and others rounded/truncated to millions. This imposes a gratuitous burden on readers: the populations could equally well be in thousands, as they are, for example, in Table B-4. In addition, the heading for the right-most column in Table B-5 is not correct: the value is fall enrollment as a *percentage* of the population.

1.4 Interactivity

The bulk of this paper presumes that NCES will continue to produce “hard copy” projection re- ports. These can be both distributed physically and, which is done now, as PDF files available from the NCES web site.

	Public			Private			Public and Private		
Actual									
1992		17.4			15.4			17.2	
1993		17.4			15.3			17.1	
1994		17.3			15.5			17.1	
1995		17.3			15.7			17.1	
1996		17.1			15.5			16.9	
1997		16.8			15.2			16.6	
1998		16.4			15.0			16.3	
1999		16.1			14.7			15.9	
2000		16.0			14.5			15.9	
2001		15.9			14.3			15.7	
2002		15.9			14.1			15.7	
2003		15.9			13.8			15.7	
2004		15.8			13.7			15.5	
2005		15.7			13.5			15.4	
Projected	**Low**	**Middle**	**High**	**Low**	**Middle**	**High**	**Low**	**Middle**	**High**
2006	15.5	15.6	15.5	13.5	13.5	13.5	15.3	15.4	15.3
2007	15.4	15.5	15.4	13.4	13.2	13.4	15.1	15.2	15.2
2008	15.2	15.3	15.4	13.1	13.0	13.2	15.0	15.1	15.1
2009	15.1	15.2	15.3	12.9	12.8	13.1	14.9	15.0	15.1
2010	15.0	15.1	15.2	12.7	12.6	12.9	14.8	14.9	15.0
2011	14.8	15.1	15.1	12.5	12.4	12.8	14.7	14.9	14.9
2012	14.7	15.0	15.0	12.4	12.2	12.7	14.6	14.8	14.9
2013	14.6	14.9	15.0	12.3	12.1	12.6	14.5	14.8	14.8
2014	14.5	14.9	14.9	12.1	12.0	12.4	14.4	14.7	14.7
2015	14.4	14.8	14.8	12.0	11.9	12.4	14.3	14.6	14.7
2016	14.3	14.7	14.7	11.9	11.8	12.3	14.2	14.6	14.6
2017	14.2	14.6	14.6	11.8	11.8	12.2	14.1	14.5	14.5

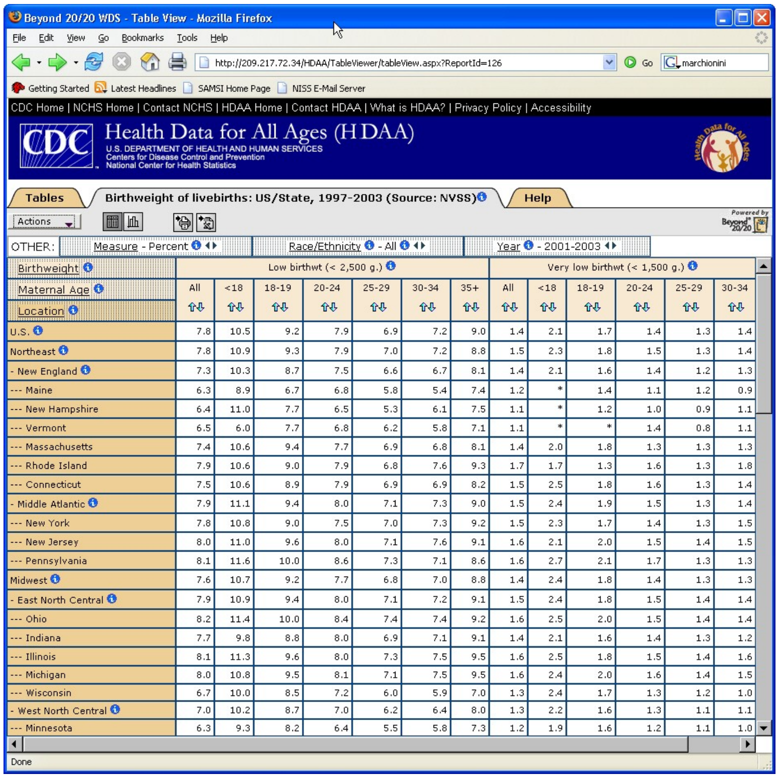
Figure 11: Alternative version of Table 33 of Hussar and Bailey (2008) incorporating multiple improvements.

At some point, however, NCES may choose to provide an interactive web version, and here we note some functionalities that are inherently useful as well as address issues raised elsewhere in this white paper. These include:

Sorting: As noted in §2.3.1, the data in Tables A and B of Hussar and Bailey (2008) and Figure 9 can be sorted by either state of magnitude of change. Both sort orders make sense, and both are informative. Capabilities for interactive sorting are well-developed and easily applied. Figure 12 shows an example containing data from the Health Data for All Ages section of the National Center for Health Statistics (NCHS)' web site.

Linked views, where the linkage allows selections to propagate from one view to the others. This *filtering* functionality is a central strength of interactive displays. To illustrate, consider Figure D of Hussar and Bailey (2008), which contains six (year, age of student, sex of student, attendance status of student, degree level, and race/ethnicity of student) distinct categorizations of the set population - students enrolled in degree-granting postsecondary institutions. In effect, the components of that figure are five two-way marginals of the underlying 6-dimensional contingency table.[13] Linked views are one means for exploring higher - dimensional structure of the data. For example, selection, using a mouse, of the 18–24 category in the first panel in Figure D would split each bar in every other panel into "18-24" and "other."

[13] year × age, year × sex, year × attendance status, year × degree level and year × race/ethnicity.



Birthweight	Low birthwt (< 2,500 g.)							Very low birthwt (< 1,500 g.)					
Maternal Age / Location	All	<18	18-19	20-24	25-29	30-34	35+	All	<18	18-19	20-24	25-29	30-34
U.S.	7.8	10.5	9.2	7.9	6.9	7.2	9.0	1.4	2.1	1.7	1.4	1.3	1.4
Northeast	7.8	10.9	9.3	7.9	7.0	7.2	8.8	1.5	2.3	1.8	1.5	1.3	1.4
- New England	7.3	10.3	8.7	7.5	6.6	6.7	8.1	1.4	2.1	1.6	1.4	1.2	1.3
--- Maine	6.3	8.9	6.7	6.8	5.8	5.4	7.4	1.2	*	1.4	1.1	1.2	0.9
--- New Hampshire	6.4	11.0	7.7	6.5	5.3	6.1	7.5	1.1	*	1.2	1.0	0.9	1.1
--- Vermont	6.5	6.0	7.7	6.8	6.2	5.8	7.1	1.1	*	*	1.4	0.8	1.1
--- Massachusetts	7.4	10.6	9.4	7.7	6.9	6.8	8.1	1.4	2.0	1.8	1.3	1.3	1.3
--- Rhode Island	7.9	10.6	9.0	7.9	6.8	7.6	9.3	1.7	1.7	1.3	1.6	1.3	1.8
--- Connecticut	7.5	10.6	8.9	7.9	6.9	6.9	8.2	1.5	2.5	1.8	1.6	1.3	1.4
- Middle Atlantic	7.9	11.1	9.4	8.0	7.1	7.3	9.0	1.5	2.4	1.9	1.5	1.3	1.4
--- New York	7.8	10.8	9.0	7.5	7.0	7.3	9.2	1.5	2.3	1.7	1.4	1.3	1.5
--- New Jersey	8.0	11.0	9.6	8.0	7.1	7.6	9.1	1.6	2.1	2.0	1.5	1.4	1.5
--- Pennsylvania	8.1	11.6	10.0	8.6	7.3	7.1	8.6	1.6	2.7	2.1	1.7	1.3	1.3
Midwest	7.6	10.7	9.2	7.7	6.8	7.0	8.8	1.4	2.4	1.8	1.4	1.3	1.3
- East North Central	7.9	10.9	9.4	8.0	7.1	7.2	9.1	1.5	2.4	1.8	1.5	1.4	1.4
--- Ohio	8.2	11.4	10.0	8.4	7.4	7.4	9.2	1.6	2.5	2.0	1.5	1.4	1.4
--- Indiana	7.7	9.8	8.8	8.0	6.9	7.1	9.1	1.4	2.1	1.6	1.4	1.3	1.2
--- Illinois	8.1	11.3	9.6	8.0	7.3	7.5	9.5	1.6	2.5	1.8	1.5	1.4	1.6
--- Michigan	8.0	10.8	9.5	8.1	7.1	7.5	9.5	1.6	2.4	2.0	1.6	1.4	1.5
--- Wisconsin	6.7	10.0	8.5	7.2	6.0	5.9	7.0	1.3	2.4	1.7	1.3	1.2	1.0
- West North Central	7.0	10.2	8.7	7.0	6.2	6.4	8.0	1.3	2.2	1.6	1.3	1.1	1.1
--- Minnesota	6.3	9.3	8.2	6.4	5.5	5.8	7.3	1.2	1.9	1.6	1.2	1.1	1.0

Figure 12: Example of an interactively sortable table, taken from the web site of the NCHS.

Mosaic plots (Friendly, 1994), which need not be interactive but are especially effective when they are, also facilitate exploration of high-dimensional structure. Figure 13 illustrates for 8-dimensional data taken from the Current Population Survey (CPS): the relationships among four categorical variables[14] come across clearly.

Micromaps (Carr and Pickle, 2010) provide similar functionality for geographically indexed data.

User-set breakpoints for maps: Multiple technologies are available that allow users interactively to manipulate category boundaries for maps such as those in Figures 5-7 of Hussar and Bailey (2008), allowing more detailed understanding of the underlying data.

[14] Race (2 categories), salary (2 categories), marital status (2 categories) and educational attainment (5 categories).

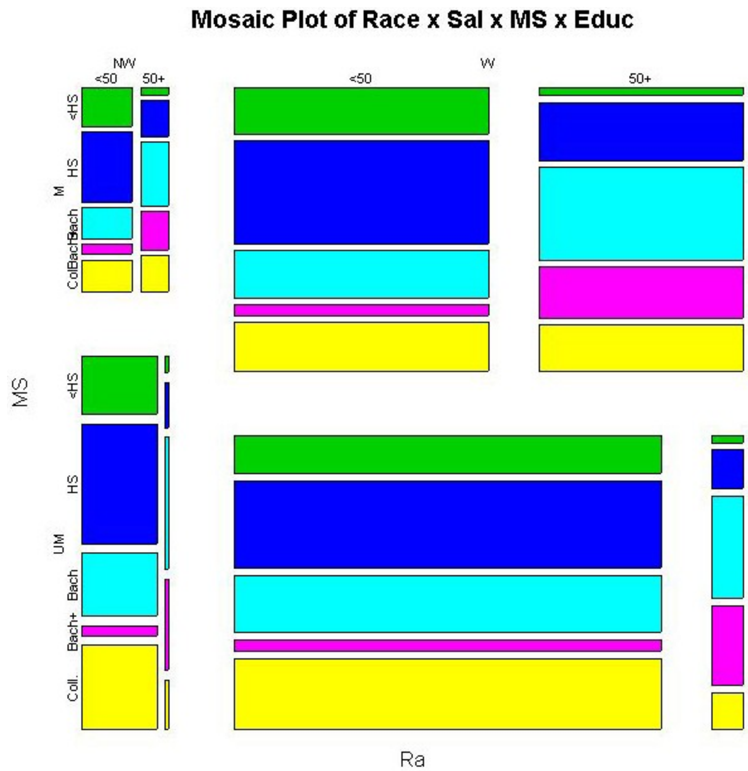


Figure 13: Example of a mosaic plot containing four-dimensional data drawn from the CPS.

II. PROJECTION METHODOLOGY

It is challenging to discuss the methodology underlying the projections in Hussar and Bailey (2008) constructively because the description of it in Appendix A is cryptic and incomplete. Some specific aspects of this description are discussed in §3.2. The methodology itself is discussed first, in §3.1.

2.1 Modeling Approach

Hussar and Bailey (2008) states explicitly and correctly on page 83 that “the equations in this appendix should be viewed as forecasting rather than structural equations.” The ensuing justification that “limitations of time and available data precluded the building of large-scale, structural models” is not persuasive. The result is a hodgepodge of models underlain loosely by one principle:

> The general methodological procedure for *Projections of Education Statistics to 2017* [that is, Hussar and Bailey (2008)] was to express the variable to be projected as a percent of a “base” variable. These percents were then projected and applied to projections of the “base” variable. For example, the number of 18-year-old college students was expressed as a percent of the 18-year-old population for each year from 1972 through 2006. This enrollment rate was then projected through the year 2017 and applied to projections of the 18-year-old population from the U.S. Census Bureau.

This principle is followed inconsistently at best.

Moreover, the failure to use structural models has multiple, negative implications, which include the following.

First, *coherence is lacking*, because the "project a base and proportions" approach is not followed consistently. If it were, the re-scaling to match totals that is mentioned, for instance, on pages 92 and 109 of Hussar and Bailey (2008) would be unnecessary.

Second, the current methodology *cannot capture "regime changes" or other shocks to the system*. In light of the current economic situation and other forces, this is a major shortcoming. It is simply not possible to believe that smooth statistical models are adequate to capture the effects of the economic downturn on state, family and personal finance, potential changes in the model of federal loans to postsecondary students, and systemic changes such as enrollment in high school via the internet. To place any credence in the financial projections in Figures L and M in future projection reports based on the current methodology seems foolish.

Third, the *current methodology does not*[15] *provide principled measures of uncertainty for projections*.[16] That the methodology is ambivalent about uncertainty is an understatement. Consider the following statements:

Page 1: "The low and high alternative projections are not statistical confidence limits."

Page 84: "These alternatives reveal the *level of uncertainty* [italics added] in making projections, was well as the sensitivity of projections to the assumptions on which they are based."

Page 85: "Therefore, alternative projections are shown for most statistical series to denote the uncertainty involved in making projections. These alternatives are not statistical confidence limits, but instead represent judgments made by the authors as to reasonable upper and lower bounds."

It is difficult for any reader, sophisticated or not, to reconcile these statements. The statement on page iii that "the first alternative set of projections (middle alternative projections) in each table is deemed to represent the most likely projections" does not seem justified, especially if "most likely" is interpreted as "mode."

The inability to quantify projection uncertainty is not academic. If uncertainties are of com- parable magnitude to the differences among low, middle and high alternative projections, then reporting alternatives is meaningless, and could be misleading. Indeed, as Table A-1 of Hussar and Bailey (2008), the demographic and economic differences among the three alternatives are subtle.[17] In the absence of principled evidence to the contrary, it is hard to believe that uncertainties associated with projections 12 years into the future do not overwhelm differences among the alternatives.

It is a major shortcoming of the current methodology that it does not offer a path to address this issue. The current methodology is not amenable to uncertainty quantification and characterization. The multiplication involved in the "project a base and proportions" approach complicates calculations by requiring characterization of dependences for which there is limited data. Bayesian methods (West and Harrison, 1999), on the other hand, inherently provide principled information about uncertainties.

It is worth noting that the current national unemployment rate of 9.8% is "off-scale" relative to all three

---

[15] And possibly cannot.

[16] All past values in Hussar and Bailey (2008) are treated as if there were no associated uncertainty, which is incorrect, but is not likely to have major consequences.

[17] Indeed, for some variables, there is no difference between alternatives.

alternatives. Therefore, *even if the current methodology were sound, there may no reason to credit any of the projections*.

Fourth, the current methodology yields little actionable insight into what is driving projections, and hence no path to inform policy or decisions. Of course, NCES may not intend or wish the projections to be used for such purposes.

Fifth, the reliance on multiplicative models, which are simply additive model for logarithms, seems tenuous. The assertion on page 84 that “Research has found that it [the multiplicative model] is a reasonable way to represent human behavior.” is subject to multiple criticisms. Multiplicative models preclude prediction techniques that entail centering. That multiplicative models are used widely is true.

Finally, neither is it justifiable to argue that the seemingly good accuracy of past projections validates the methodology. Absent action by NCES, that quality may deteriorate dramatically.

Returning to the current methodology in general, there are numerous *arbitrary and unjustified choices*. Here is one of the most striking: page 83 of Hussar and Bailey (2008) contains the statement “Projections of enrollments and public high school graduates are based on a smoothing constant of α = 0.4.” On what is this choice based? Does it make sense scientifically? How sensitive are the results to it?

There also appear to be technical flaws in some of the models. Specifically, consider the model for projection of postsecondary associated with Tables A-23 and A-24.[18] That model, for associates degrees received by men, appears to be

$$\log\left(\frac{\text{Degrees}(t)}{\text{P}(t)}\right) = 5.0 + 0.4\left[0.67\log\left(\frac{\text{FTE}(t-1)}{\text{P}(t-1)}\right) + 0.33\log\left(\frac{\text{FTE}(t-2)}{\text{P}(t-2)}\right) + 0.67\log\left(\frac{\text{PTE}(t-1)}{\text{P}(t-1)}\right) + 0.33\log\left(\frac{\text{PTE}(t-2)}{\text{P}(t-2)}\right)\right]$$

Here, P is population, FTE is full-time enrollment, and PTE is part-time enrollment. It appears that the coefficients 5.0 and 0.4 are estimated from data, whereas 0.67 and 0.33 are simply arbitrary. In any event, the model in (1) is not identifiable.

We note in passing that the substitution of verbal descriptions for equations in Hussar and Bailey (2008) is problematic. Based on the footnotes in Table A-23, equation (1) could instead have been

$$\log\left(\frac{\text{Degrees}(t)}{\text{P}(t)}\right) = a + b\left[\log\left(0.67\frac{\text{FTE}(t-1)}{\text{P}(t-1)} + 0.33\frac{\text{FTE}(t-2)}{\text{P}(t-2)}\right) + \log\left(0.67\frac{\text{PTE}(t-1)}{\text{P}(t-1)} + 0.33\frac{\text{PTE}(t-2)}{\text{P}(t-2)}\right)\right].$$

The relative inability to predict numbers of postsecondary degrees awarded, as evidenced in Table A-2 of Hussar and Bailey (2008) demonstrates clearly the weakness of the current methodology as compared to a

[18] Which are not consistent with each other in their description of the model. Moreover, the former seems to state that coefficients are estimated to only one decimal place, which seems indefensible.

structural modeling approach, whether Bayesian or not. Completion of degree requirements is an event with numerous precursors contained in institutional-level databases, so short-term prediction, which is dominated by imminence of completion of requirements, would be straightforward if based on such data.

For long-term predictions, on the other hand, variability in time-to-complete might be thought to be a "smoothing" factor, which would mean that poor predictions are largely the result of inability to predict enrollment. However, Table A-2 of Hussar and Bailey (2008) shows that reality is more complex: 10-years-into-the-future projections of postsecondary enrollment have mean absolute percentage errors (MAPEs) on the order of 10%, while errors for masters degrees awarded are more than 20%. One explanation is failure-to-complete - students who enroll in masters degree programs but do not complete requirements for their degree.

Within the current methodology, there is no avenue to resolve these kinds of questions. Of course, however, processes of postsecondary access, choice and progression are under study by NCES and many other organizations and individuals, and an alternative approach based on structural modeling could be informed by this knowledge.

2.2 Description of the Methodology in Appendix A of Hussar and Bailey (2008)

The description of the projection methodology is incomplete in several important respects. The most glaring of these is that the equation

$$\hat{X}_{t+1} = \alpha \left[X_t + (1-\alpha) X_{t-1} + (1-\alpha)^2 X_{t-2} + \dots \right]$$

on page 83, [19] which is rewritten here in a way that makes clear what is being predicted, only specifies the projected value for *next time period*. Nothing in Hussar and Bailey (2008) describes projections further into the future, which can be done in several different ways.

On page 113, Hussar and Bailey (2008) presents one multiplicative model for pupil/teacher ratios in public elementary schools as a function of teacher salaries and per-student elementary education revenue from state sources,[20] and another multiplicative model for pupil/teacher ratios in public secondary schools as a function of the fraction of the secondary school-age population enrolled in secondary school and per-student education revenue from state sources.

In both cases, multiplying the inverses of projections of these values by projected numbers of students produces projected numbers of teachers.

This approach raises multiple issues:

1. 1.Why two different models? Is the justification scientific, statistical, or something else?
2. What is the unit of analysis at which the models are estimated? State? National? Because of

[19] Which, incidentally does not make proper use of the ellipsis...

[20] As an indicator of the level of opaqueness of Hussar and Bailey (2008), on page 113, in what might be considered running text, the latter is stated to be "the level of education revenue from state sources deflated by the consumer prices chained-price index in constant 2000 dollars per public elementary student," while—two pages later—in a footnote to Table A-25, that variable is stated to be "the ratio of education revenue from state sources per capita [italics added] to public elementary school enrollment." The inclusion of logs in the definitions of the variables in the notes to Table A-25 is taken to be erroneous.

extreme state-to-state variations, fitting only at the national level appears inadequate.

3. (Hussar and Bailey, 2008) alludes in passing (page 114) to the fact that definitions of “elementary” and “secondary” in terms of grades are not uniform across public school systems. This seems important enough not to be relegated to a Technical Appendix that most readers will ignore.

4. The final equations then reveal the essentially *ad hoc* nature of the current methodology. Consider elementary school teachers. Using abbreviated notation (P = pupils, T = teachers, S = salary, E = state-derived revenue per capita but not per student), the model in Table A-25 is, after exponentiation,

$$\frac{P}{T} = 3.8 S^{.1} (\frac{E}{P})^{-.2}.$$

This means that the project number of teachers is

$$T = P\left(\frac{P}{T}\right)^{-1} = \frac{1}{3.8} S^{-.1} E^{.2} P^{.8}.$$

It does not seem possible to have confidence in this equation, especially since it predicts that the number of teachers falls as teacher salaries rise.

Some parts of the description are undecipherable. Consider the “Basic Methodology” material beginning on page 89 of Hussar and Bailey (2008):

1. The subscripting violates basic principles of notational clarity: enrollment in grade j = 1 for year t = 2009 is written as G_{12009}. Because of the importance to time to projections, it is much preferable to use $G_j(t)$, which is free of ambiguity. And why not just use 0 as a subscript for kindergarten?

2. The line “G_{1t} = enrollment in grade 1” is superfluous.

3. The expression “$EG_t = K_t + E_t + \Sigma^8_{j=1} G_{it}$” is a tautology presented as if it were the result of mathematical manipulation. The same is true of the expression for SG_t.

4. The expression “K_t = RK * (P_{5t})” misleadingly makes K_t appear to be a function of P_{5t}. Written as intended, that is,

$$K_t = \mathrm{RK}_t \times P_{5t},$$

the expression makes sense, but it is completely unclear what is being defined in terms of what. By any reasonable interpretation, the definition of RK_t is

$$\mathrm{RK}_t = \frac{K_t}{P_{5t}},$$

but this is not what (4) suggests.

5. Similarly, the equation

$$G_{jt} = R_{jt} \times G_{j-1,t-1}$$

really is a definition of R_{jt}, but is not written that way.

6. The equation

$$E_t = \mathrm{RE}_t \times \sum_{i=5}^{13} P_{it}$$

imposes an arbitrary and unnecessary assumption that only 5-13-year olds can be enrolled in elementary special and ungraded programs.

The equation $S_t = \mathrm{RS}_t \times \sum_{i=14}^{17} P_{it}$ has the same problem.

Continuing to the material on page 90, which relates to enrollment in degree-granting postsecondary institutions,

1. The use of i = 25 to represent ages 25-29, i = 26 to represent ages 30-34 and i = 27 to represent ages 35 and over in one case and 35-44 in another can only be described as torturing readers, especially given that for i = 16, . . . , 24 the subscript corresponds to true age.
2. The equation $T_{ijkt} = \sum_{i=16}^{27} E_{ijkt}$ is a tautology.
3. The equation $E_{ijkt} = R_{ijkt} \times P_{ijt}$ is actually the definition of R_{ijkt}.

2.3 Improving the Methodology

There is no question that an alternative projection methodology can be developed that addresses raised in §3.1. Because of the magnitude of the effort, this paper does not - indeed, cannot - lay out a complete path to improving such methodology. Incremental improvement does not seem feasible. A full-scale approach would rely on:

- Structural models.
- Bayesian methods (West and Harrison, 1999), in order to incorporate new information and to characterize uncertainties in a principled way.
- Modern forecasting methods, as exemplified by Alho and Spencer (2005).

The new methods would be very intensive computationally, since some uncertainty quantification would be by means of simulations.

Depending on how information in NCES projection reports is employed, the benefits of a complete revamping of the methodology may not justify the costs. A rough estimate of the effort would be a two-year project involving senior researchers and a full-time postdoctoral fellow, with deep engagement of relevant NCES staff. An accompanying effort, for example, focus groups or an expert task force, to understand how the projections are used seems vital. Careful exploration of existing and alternative sources of data would also be necessary.

III. OTHER ISSUES

The items listed here identify other improvements to NCES' projection reports.

1. There should be *consistent and complete cross-referencing*. For example, Figure 2 in Hussar and Bailey (2008) is a graphical presentation of the values in Table 2,[21], but there is no cross-reference. The projection reports should not require users to supply cross-referencing.

   Proper attention to cross-referencing would have revealed that there appears to be no table containing the values underlying Figure 1 of Hussar and Bailey (2008). Why is this?

2. The interleaved numbering of tables and figures in the body of the report is unusual. Most style guides recommend that tables and figures be numbered separately. Using letters to label figures and tables in the body of the report and numbers to label Reference Figures and Reference Tables compounds the disorganization. Labeling Reference Figures as Figure R-1, ...is as easy as it is effective.

3. Given the bulk of Hussar and Bailey (2008), there is reluctance to propose additions, yet there are multiple missed opportunities. To illustrate, consider Figures 1 and 2 of Hussar and Bailey (2008), which present actual and projected populations and PK–12 enrollments, respectively. It is true that students enrolled in PK–12 are not a subset of the 5–17 population, the report provided no way to understand the relationship between population and enrollment. For a significant segment of users, absolute declines in enrollment mean much less that declines in enrollment/population, but this information is not accessible in Hussar and Bailey (2008).

   We also note the absence in NCES' projection reports of what might be termed "measures of performance" of the education system. Examples are completion rates (or dropout rates) for secondary and postsecondary students, measures of student performance such as scores on the National Assessment of Educational Progress (NAEP), and population-based measures such as the percentage of adults with varying levels of educational attainment. Inclusion of such measures would increase readership dramatically. In any event, Hussar and Bailey (2008) does not articulate a rationale for what it does and does not contain; readers would benefit from knowing it.

4. Hussar and Bailey (2008) does not state whether there is "hidden" disaggregation underlying some projections, even if results are reported in aggregate form. To illustrate, consider Figures L and M or Table B-6 of Hussar and Bailey (2008), which contain projected expenditures. Are these values projections of national aggregates or aggregates of state-level projections? Here and elsewhere, are CPI projections at the state or national level? Another instance is discussed in §3.2.

[21] At least, it appears to be that.

APPENDICES

Appendix A: References

Alho, J. M. and Spencer, B. D. (2005). *Statistical Demography and Forecasting*. Springer-Verlag, New York.

Carr, D. B. and Pickle, L. W. (2010). *Visualizing Data Patterns with Micromaps*. Taylor & Francis, London.

Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *J. Amer. Statist. Assoc*., 89:190U˝ -200.

Hussar, W. J. and Bailey, T. M. (2008). Projections of education statistics to 2017. Technical report, National Center for Education Statistics. Available on-line at http://nces.ed.gov/programs/projections/projections2017/.

MacEachern, A. (1995). *How Maps Work*. The Guilford Press, New York.

National Center for Education Statistics (2004). Statistical Standards. Information available on- line at nces.ed.gov/statprog/stat_standards.asp.

Pickle, L. W., Mungiole, M., Jones, G. K., and White, A. A. (1997). *Altas of United States Mortality*. U. S. Department of Health and Human Services, Hyattsville, MD.

Tufte, E. (1983). *The Visual Display of Quantitative Information*. Graphics Press, Chesire, CT.

Tufte, E. (1990). *Envisioning Information*. Graphics Press, Chesire, CT.

Tufte, E. (1997). *Visual Explanations*. Graphics Press, Chesire, CT.

West, M. and Harrison, J. (1999). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York. 2nd. edition.

Wilkinson, L. (2005). *The Grammar of Graphics*. Springer-Verlag, New York. 2nd. edition.

Appendix B: Figures from Hussar and Bailey (2008)

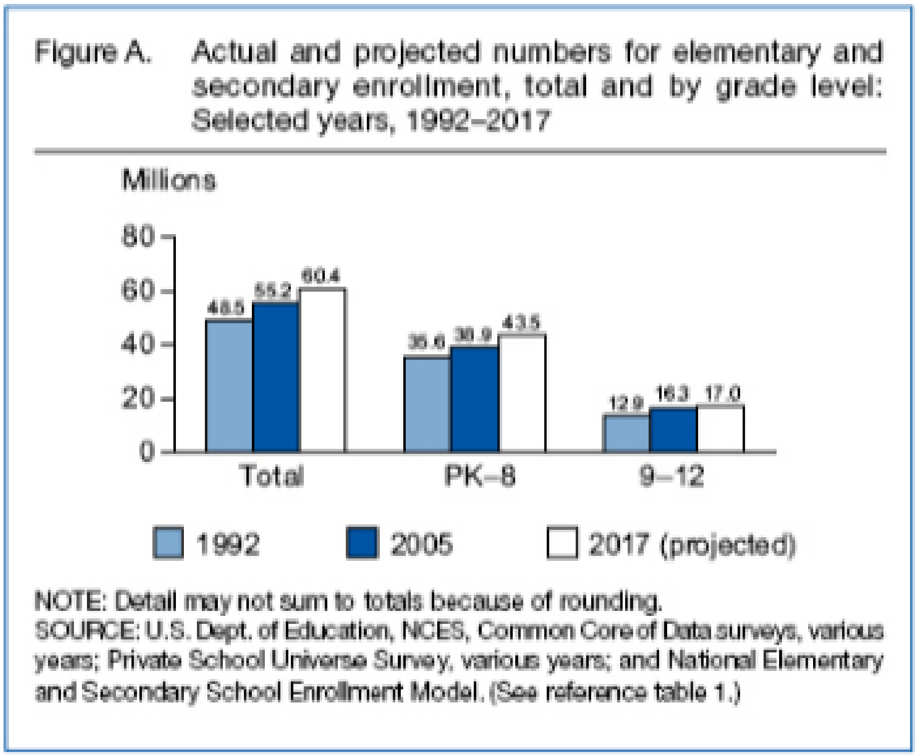


Figure 14: Figure A of Hussar and Bailey (2008).

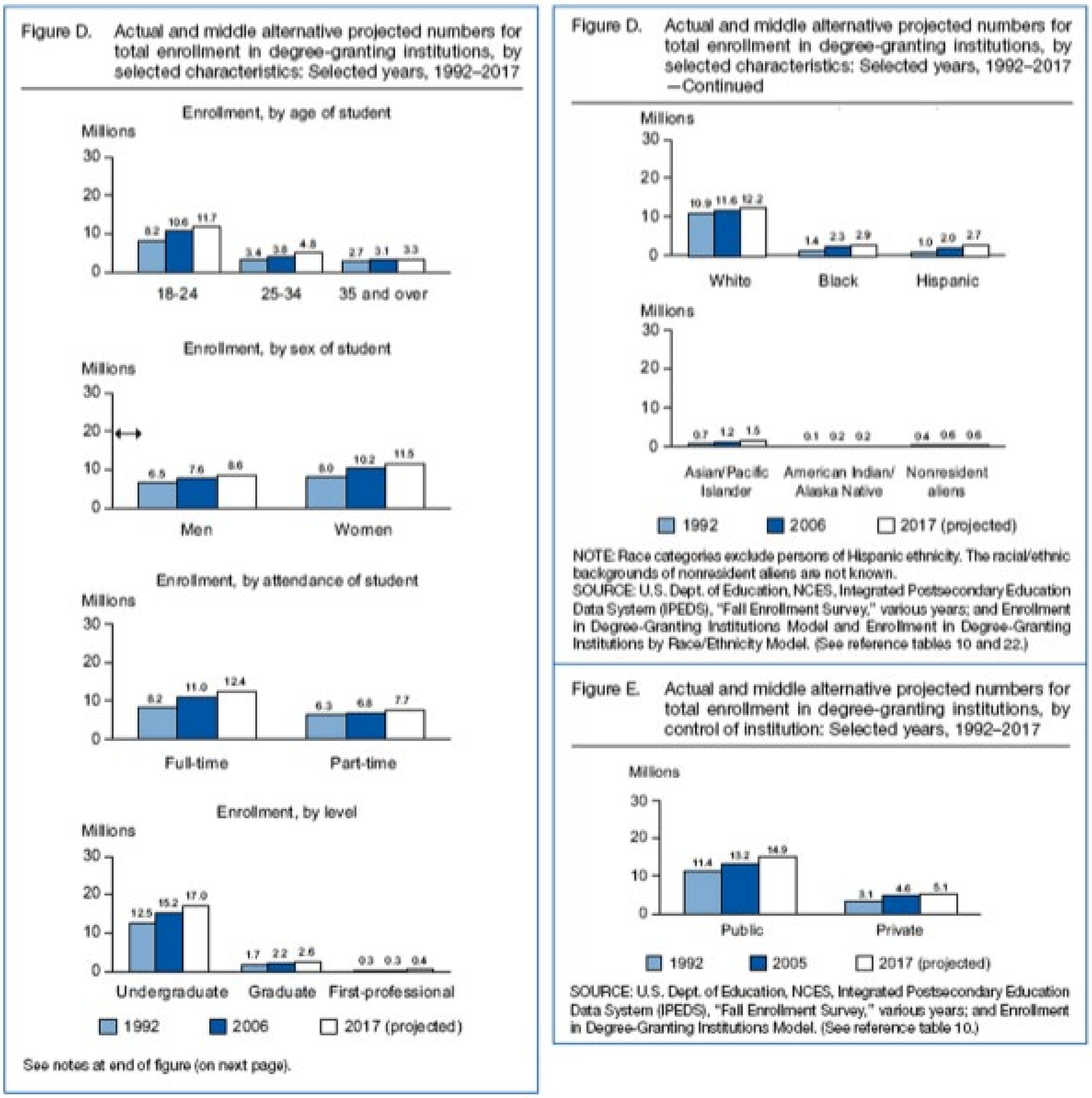


Figure 15: Figure D of Hussar and Bailey (2008).

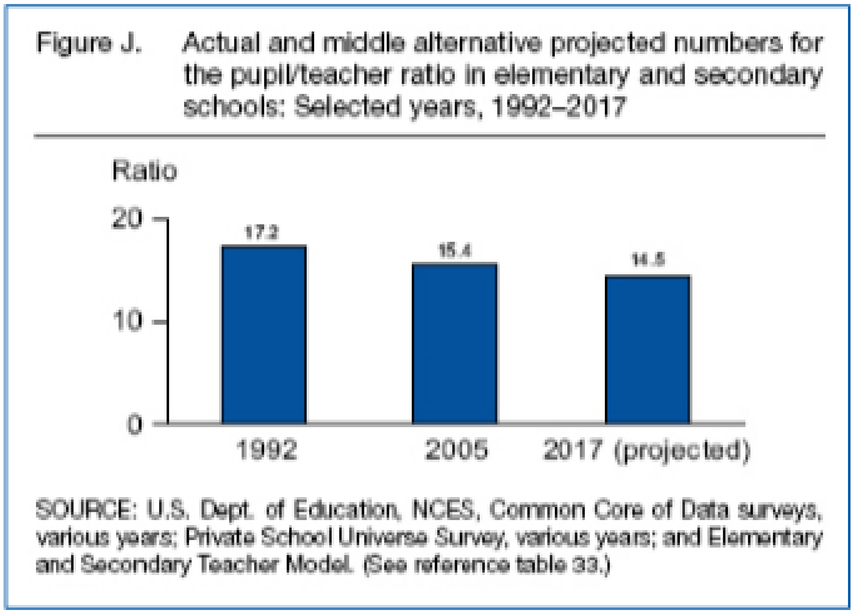


Figure 16: Figure J of Hussar and Bailey (2008).

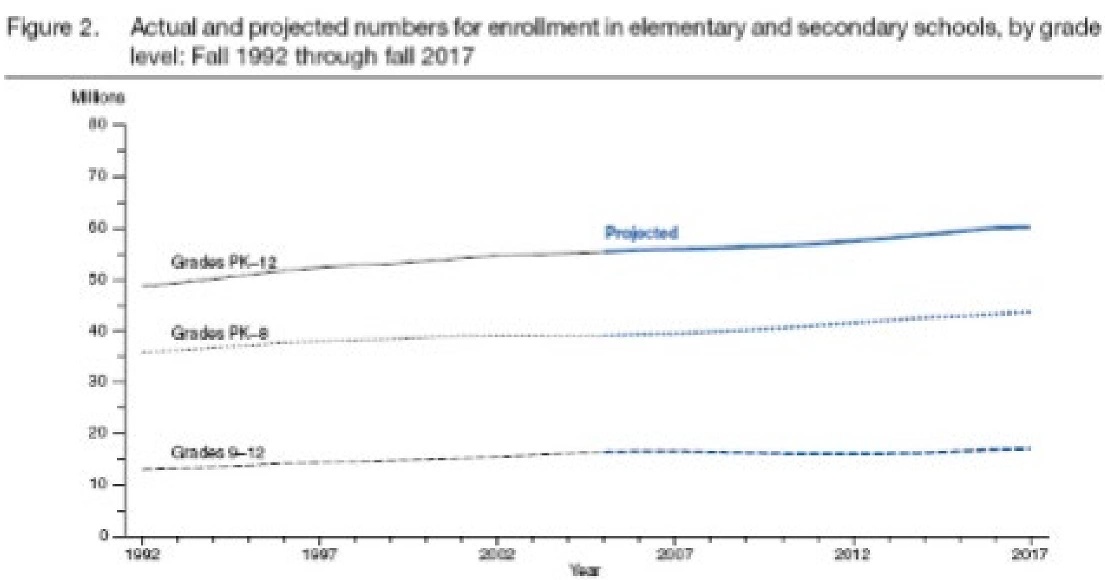


Figure 17: Figure 2 of Hussar and Bailey (2008).

Figure 3. Actual and projected numbers for enrollment in elementary and secondary schools, by control of school: Fall 1992 through fall 2017

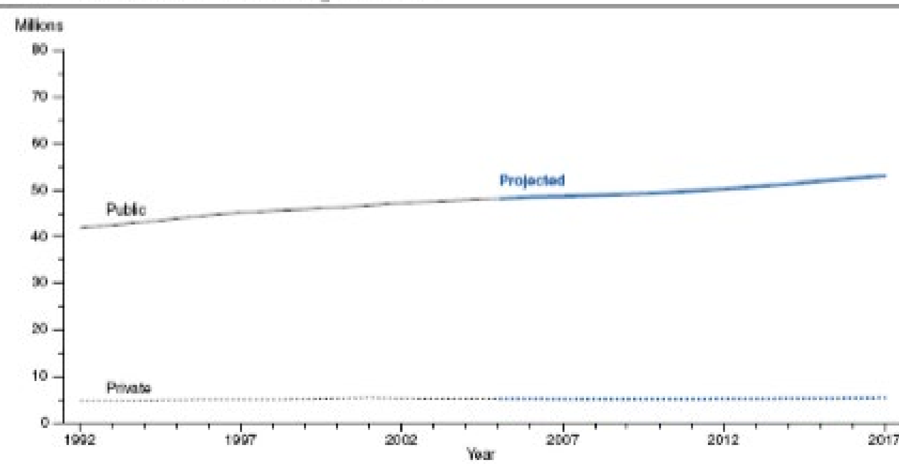


Figure 18: Figure 3 of Hussar and Bailey (2008).

Table A. Projected percentage increases in public elementary and secondary school enrollment, by state: 2005 through 2017

State	Percent change	State	Percent change
Arizona	44.8	Washington	7.4
Nevada	43.2	Oklahoma	7.4
Texas	32.9	Alaska	6.1
Florida	28.9	Maryland	5.3
Utah	27.5	Nebraska	4.6
Georgia	27.1	Minnesota	4.2
North Carolina	23.1	Missouri	3.5
Idaho	22.9	Illinois	3.2
Colorado	18.9	Indiana	3.0
District of Columbia	14.3	Kentucky	2.7
Delaware	13.1	Alabama	2.6
Virginia	13.0	Wyoming	1.8
Hawaii	13.0	New Jersey	1.7
Oregon	11.2	Kansas	1.5
New Mexico	11.1	Wisconsin	0.9
Tennessee	10.4	Mississippi	0.4
Arkansas	9.9	Montana	0.4
California	8.7	South Dakota	0.4
South Carolina	7.8	Iowa	0.2

SOURCE: U.S. Dept. of Education, NCES, Common Core of Data surveys and State Public Elementary and Secondary Enrollment Model. (See reference table 5.)

Table B. Projected percentage decreases in public elementary and secondary school enrollment, by state: 2005 through 2017

State	Percent change	State	Percent change
Louisiana	-12.4	New York	-5.2
Vermont	-11.7	Massachusetts	-4.0
Rhode Island	-11.4	New Hampshire	-3.8
Maine	-8.0	West Virginia	-3.5
North Dakota	-7.6	Ohio	-3.2
Connecticut	-6.3	Pennsylvania	-2.0
Michigan	-6.2		

SOURCE: U.S. Dept. of Education, NCES, Common Core of Data surveys and State Public Elementary and Secondary Enrollment Model. (See reference table 5.)

Figure 19: Tables A and B of Hussar and Bailey (2008).

Figure 5. Projected percentage change in grades PK–12 enrollment in public schools, by state: Fall 2005 through fall 2017

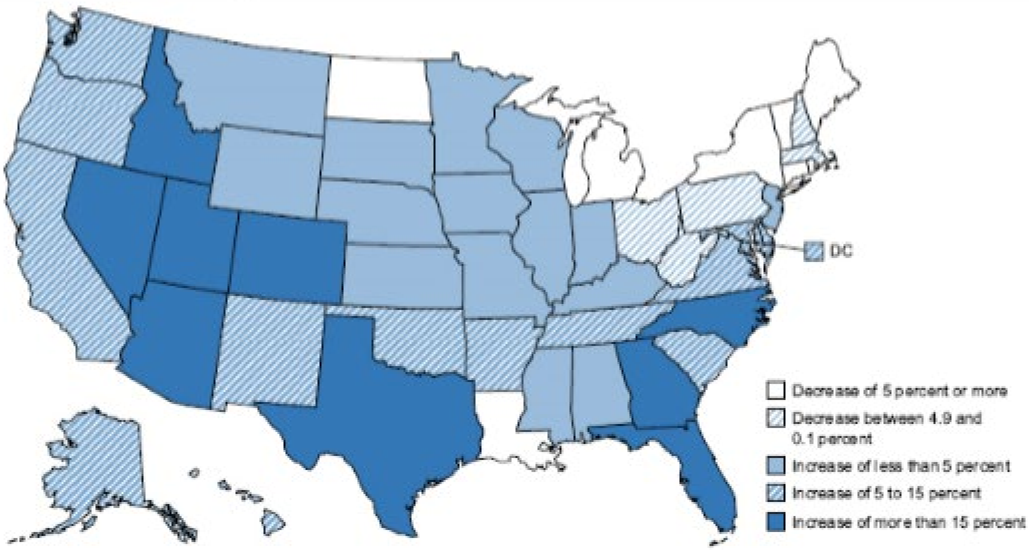


Figure 20: Figure 5 of Hussar and Bailey (2008).

Table 10. Actual and alternative projected numbers for total enrollment in all degree-granting postsecondary institutions, by sex, attendance status, and control of institution: Fall 1992 through fall 2017

[In thousands]

Year	Total	Sex		Attendance status		Control	
		Men	Women	Full-time	Part-time	Public	Private
Actual							
1992	14,486	6,524	7,963	8,161	6,325	11,385	3,102
1993	14,305	6,427	7,877	8,128	6,177	11,189	3,116
1994	14,279	6,372	7,907	8,138	6,141	11,134	3,145
1995	14,262	6,343	7,919	8,129	6,133	11,092	3,169
1996	14,368	6,353	8,015	8,303	6,065	11,121	3,247
1997	14,502	6,396	8,106	8,438	6,064	11,196	3,306
1998	14,507	6,369	8,138	8,563	5,944	11,138	3,369
1999	14,791	6,491	8,301	8,786	6,005	11,309	3,482
2000	15,312	6,722	8,591	9,010	6,303	11,753	3,560
2001	15,928	6,961	8,967	9,448	6,480	12,233	3,695
2002	16,612	7,202	9,410	9,946	6,665	12,752	3,860
2003	16,911	7,260	9,651	10,326	6,585	12,859	4,053
2004	17,272	7,387	9,885	10,610	6,662	12,980	4,292
2005	17,487	7,456	10,032	10,797	6,690	13,022	4,466
2006	17,759	7,575	10,184	10,957	6,802	13,180	4,579
Middle alternative projections							
2007	17,976	7,704	10,271	11,104	6,872	13,373	4,603
2008	18,200	7,822	10,378	11,263	6,937	13,562	4,638
2009	18,416	7,929	10,487	11,413	7,003	13,748	4,668
2010	18,613	8,022	10,590	11,546	7,067	13,890	4,722
2011	18,822	8,118	10,704	11,678	7,144	14,041	4,781
2012	19,048	8,213	10,835	11,817	7,231	14,201	4,847
2013	19,299	8,306	10,993	11,973	7,326	14,380	4,919
2014	19,533	8,387	11,146	12,114	7,419	14,547	4,986
2015	19,716	8,443	11,273	12,223	7,493	14,677	5,039
2016	19,893	8,500	11,393	12,326	7,567	14,804	5,089
2017	20,080	8,568	11,512	12,430	7,650	14,942	5,138
Low alternative projections							
2007	17,974	7,709	10,265	11,113	6,860	13,370	4,604
2008	18,182	7,829	10,353	11,277	6,905	13,545	4,637
2009	18,271	7,898	10,372	11,320	6,951	13,641	4,630
2010	18,354	7,957	10,397	11,359	6,995	13,704	4,650
2011	18,452	8,018	10,433	11,398	7,054	13,776	4,676
2012	18,597	8,088	10,509	11,469	7,128	13,880	4,717
2013	18,784	8,161	10,623	11,572	7,212	14,013	4,771
2014	18,969	8,227	10,742	11,674	7,295	14,147	4,822
2015	19,111	8,271	10,840	11,750	7,361	14,248	4,863
2016	19,252	8,318	10,934	11,825	7,427	14,349	4,903
2017	19,404	8,376	11,028	11,901	7,503	14,462	4,941
High alternative projections							
2007	18,033	7,719	10,314	11,148	6,885	13,414	4,619
2008	18,304	7,850	10,454	11,344	6,960	13,636	4,668
2009	18,544	7,965	10,580	11,513	7,031	13,839	4,705
2010	18,785	8,071	10,714	11,681	7,104	14,013	4,772
2011	19,049	8,183	10,866	11,857	7,192	14,202	4,847
2012	19,337	8,296	11,041	12,046	7,291	14,406	4,931
2013	19,650	8,407	11,243	12,252	7,398	14,628	5,022
2014	19,924	8,499	11,426	12,426	7,498	14,824	5,100
2015	20,145	8,564	11,580	12,565	7,580	14,980	5,164
2016	20,368	8,634	11,734	12,705	7,663	15,140	5,228
2017	20,606	8,717	11,889	12,850	7,756	15,314	5,293

Figure 21: Table 10 of Hussar and Bailey (2008)

Year (July 1)	18-year-olds	18- to 24-year-olds	25- to 29-year-olds	30- to 34-year-olds	35- to 44-year-olds
Actual					
1992	3,354	26,282	20,591	22,564	40,046
1993	3,455	26,102	20,146	22,646	40,975
1994	3,428	25,821	19,809	22,648	41,877
1995	3,601	25,585	19,742	22,425	42,765
1996	3,650	25,376	19,927	21,996	43,605
1997	3,780	25,574	19,960	21,494	44,282
1998	3,984	26,155	19,863	20,999	44,802
1999	3,993	26,780	19,632	20,647	45,130
2000	4,076	27,393	19,367	20,579	45,235
2001	4,074	28,087	19,004	20,781	45,188
2002	4,033	28,601	18,997	20,878	44,869
2003	4,131	29,094	19,213	20,789	44,484
2004	4,128	29,408	19,625	20,528	44,178
2005	4,127	29,500	20,148	20,153	43,954
2006	4,190	29,610	20,800	19,764	43,748
Projected					
2007	4,272	29,809	21,313	19,713	43,379
2008	4,401	30,173	21,672	19,865	42,782
2009	4,384	30,536	21,878	20,213	42,109
2010	4,312	30,762	21,944	20,657	41,600
2011	4,250	30,894	21,981	21,205	41,318
2012	4,170	30,947	22,057	21,652	41,217
2013	4,126	30,884	22,205	22,000	41,222
2014	4,080	30,693	22,459	22,202	41,258
2015	4,007	30,297	22,783	22,271	41,270
2016	3,990	29,901	23,059	22,313	41,421
2017	4,018	29,607	23,260	22,394	41,754

Figure 22: Table B-4 of Hussar and Bailey (2008).

Table 33. Actual and alternative projected numbers for the pupil/teacher ratios in elementary and secondary schools, by control of school: Fall 1992 through fall 2017

Year	Total	Public	Private
Actual			
1992[1]	17.2	17.4	15.4
1993	17.1	17.4	15.3
1994[1]	17.1	17.3	15.5
1995	17.1	17.3	15.7
1996[1]	16.9	17.1	15.5
1997	16.6	16.8	15.2
1998[1]	16.3	16.4	15.0
1999	15.9	16.1	14.7
2000[1]	15.9	16.0	14.5
2001	15.7	15.9	14.3
2002[1]	15.7	15.9	14.1
2003	15.7	15.9	13.8
2004[1]	15.5	15.8	13.7
2005	15.4	15.7	13.5
Middle alternative projections			
2006	15.4	15.6	13.5
2007	15.2	15.5	13.2
2008	15.1	15.3	13.0
2009	15.0	15.2	12.8
2010	14.9	15.1	12.6
2011	14.9	15.1	12.4
2012	14.8	15.0	12.2
2013	14.8	14.9	12.1
2014	14.7	14.9	12.0
2015	14.6	14.8	11.9
2016	14.6	14.7	11.8
2017	14.5	14.6	11.8
Low alternative projections			
2006	15.3	15.5	13.5
2007	15.1	15.4	13.4
2008	15.0	15.2	13.1
2009	14.9	15.1	12.9
2010	14.8	15.0	12.7
2011	14.7	14.8	12.5
2012	14.6	14.7	12.4
2013	14.5	14.6	12.3
2014	14.4	14.5	12.1
2015	14.3	14.4	12.0
2016	14.2	14.3	11.9
2017	14.1	14.2	11.8
High alternative projections			
2006	15.3	15.5	13.5
2007	15.2	15.4	13.4
2008	15.1	15.4	13.2
2009	15.1	15.3	13.1
2010	15.0	15.2	12.9
2011	14.9	15.1	12.8
2012	14.9	15.0	12.7
2013	14.8	15.0	12.6
2014	14.7	14.9	12.4
2015	14.7	14.8	12.4
2016	14.6	14.7	12.3
2017	14.5	14.6	12.2

Figure 23: Table 33 of Hussar and Bailey (2008).

Appendix C: Author

Alan F. Karr, PhD
Director, National Institute of Statistical Sciences