Re-analysis of NAEP Math and Reading Scores in States with and without High-stakes Tests: Response to Rosenshine

Audrey Amrein-Beardsley
David C. Berliner
Arizona State University


Abstract
Here we address the criticism of our NAEP analyses by Rosenshine (2003). On the basis of his thoughtful critique we redid some of the analyses on which he focused. Our findings contradict his. This is no fault of his, the reasons for which are explained in this paper. Our findings do support our position that high-stakes tests do not do much to improve academic achievement. The extent to which states with high-stakes tests outperform states without high-stakes tests is, at best, indeterminable. Using 1994-1998 NAEP reading and 1996-2000 NAEP math data and accounting for NAEP exemption rates for the same years, we found that states with high-stakes tests are not outperforming states without high-stakes tests in reading in the 4th grade or math in the 8th grade at a statistically significant level. States with high-stakes tests are, however, outperforming states without high-stakes tests in math in the 4th grade at a
statistically significant level. Our findings also support our earlier stance that states with high-stakes tests are exempting more students from participating in the NAEP than are states without high-stakes tests. This is more prevalent the more recent the NAEP test administration. This is illustrated in the tables below.

Introduction

In our research, we were concerned that scores on high-stakes state tests could easily be manipulated through narrowing of the curriculum, drilling on items similar to the test, increasing exclusion rates of students, increasing in dropouts and push-outs, and the like. To judge whether that concern was valid, we looked at audit tests—tests that might have some overlap with a state’s own test but where school personnel were under much less intense pressure to achieve higher scores. We chose a series of audit tests to examine—SAT, ACT and AP tests, as well as all administrations of the NAEP reading and mathematics tests. We also studied whether some unanticipated side effects were present when high-stakes tests were introduced, such as increased GED taking, increased reporting of cheating, problems of teacher morale, problems with student motivation to learn, and so forth.

Substantive criticism of our work, thus far, has been limited to the NAEP data we reported. To our knowledge, the other conclusions we reached have not yet been subject to the same kinds of thoughtful criticism. So for now, given the methods that we used for analyses, our findings in those other areas stand. We concluded that there was no systematic pattern of gains on SATs, ACTs or AP exams. That is, we found no evidence of transfer from the state tests to these other tests, tests that can be considered as audit measures. In addition, we found increased drop-out rates and decreased high school graduation rates, increased rates by which students participated in the GED program, and a host of troubling negative affects associated with high-stakes testing.

Here we address the criticism of our NAEP analyses by Rosenshine (2003). On the basis of his thoughtful criticism, we redid some of the analyses on which he focused and now have a different view of the findings. What we found contradicts what we found in both of our earlier papers (Amrein & Berliner, 2002a; Amrein & Berliner 2002b) but the data analyzed below are for different years of data from those used in the earlier papers. Following the form of the analyses done by Rosenshine the data analyzed below are only for the years 1994-1998 for the NAEP reading test, and 1996-2000 for the NAEP mathematics tests. In addition, in our earlier work we used the national trend line as the contrast or control group for our analyses. In this analysis, we use the composite score for states without high-stakes tests as the control. In addition, our findings contradict the findings reported by Rosenshine. This is no fault of his. Rosenshine used our designation of clear and unclear states with and without high-stakes tests from the second of our two papers.[2] We communicated many times and approved the states he used in his analysis. Given more consideration, however, we noticed the distinctions we made between clear and unclear states was based on our overall findings which were based on all of the available NAEP data. In other words, Rosenshine analyzed the latest two NAEP administrations in reading and math using the
distinctions we made between clear and unclear states when we used all of the available NAEP data, approximately 10 years of NAEP data per subject. To complicate things more, because we used the national trend line as our control group, our clear/unclear distinctions were also made factoring in the national average. Rosenshine did not do this which makes for differences in the findings. He used the states without high-stakes tests as the control. This makes for a better analysis and we have followed his lead here. In short, Rosenshine should not be faulted for his findings nor should he be considered wrong in what he did. He did a fine reanalysis of our NAEP examination given the information he had at that point, and here we are redoing his.

**NAEP Reading Grade 4 1994-1998**

Taking Table 1 from Amrein & Berliner (2002b) and the states in which high stakes tests were implemented before 1994 and between 1994 and 1998, we re-ran our analyses, as Rosenshine did, using all states with high-stakes tests and, as the control group, all states without high-stakes tests for which NAEP data were available. What we found in regards to reading grade 4 achievement from 1994-1998 is as follows:

**Table 1**

**Fourth grade 1994-1998 NAEP reading scores (raw data).**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona</td>
<td>206</td>
<td>207</td>
<td>Alabama</td>
<td>208</td>
<td>211</td>
</tr>
<tr>
<td>Arkansas</td>
<td>209</td>
<td>209</td>
<td>Kentucky</td>
<td>212</td>
<td>218</td>
</tr>
<tr>
<td>California</td>
<td>197</td>
<td>202</td>
<td>Louisiana</td>
<td>197</td>
<td>204</td>
</tr>
<tr>
<td>Colorado</td>
<td>213</td>
<td>222</td>
<td>Maryland</td>
<td>210</td>
<td>215</td>
</tr>
<tr>
<td>Connecticut</td>
<td>222</td>
<td>232</td>
<td>Michigan</td>
<td>n/a</td>
<td>217</td>
</tr>
<tr>
<td>Delaware</td>
<td>206</td>
<td>212</td>
<td>Mississippi</td>
<td>202</td>
<td>204</td>
</tr>
<tr>
<td>Florida</td>
<td>205</td>
<td>207</td>
<td>Missouri</td>
<td>217</td>
<td>216</td>
</tr>
<tr>
<td>Georgia</td>
<td>207</td>
<td>210</td>
<td>New Mexico</td>
<td>205</td>
<td>206</td>
</tr>
<tr>
<td>Hawaii</td>
<td>201</td>
<td>200</td>
<td>North Carolina</td>
<td>214</td>
<td>217</td>
</tr>
<tr>
<td>Iowa</td>
<td>223</td>
<td>223</td>
<td>Oklahoma</td>
<td>n/a</td>
<td>220</td>
</tr>
<tr>
<td>State</td>
<td>Change in Score</td>
<td>Change in Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td>n/a</td>
<td>222</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td>228</td>
<td>225</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>223</td>
<td>225</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td>218</td>
<td>222</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montana</td>
<td>222</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td>n/a</td>
<td>208</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>223</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>212</td>
<td>216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>n/a</td>
<td>214</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>220</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utah</td>
<td>217</td>
<td>215</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>213</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td>213</td>
<td>217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>224</td>
<td>224</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td>221</td>
<td>219</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVERALL</td>
<td>214.7</td>
<td>216.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at a p < .05 level

Table 1 illustrates that the states with high-stakes tests outperformed those states without high-stakes tests on the NAEP grade 4 reading tests over the period 1994-1998. However, as shown in our earlier research (Amrein & Berliner, 2002a; Amrein & Berliner, 2002b), the rates by which students are excluded from the NAEP must be taken into consideration to determine whether gains and losses are clear (interpretable) or unclear (not interpretable).

Clear gains can be determined if a state’s scores increase while the rates by which students are exempted from the NAEP stay the same or decrease. In other words, when the pool of students sampled to participate in the NAEP is less selective then the likelihood that their scores would increase artificially is nullified. Under these conditions such gains are clear. Clear losses can be determined if a state’s scores decrease at the same time the rates by which students are exempted from the NAEP increase. In this case, the pool of students sampled was more selective and yet the scores still went down. Under these conditions it is reasonable to interpret those findings as a clear loss.

Unclear gains are the case when a state’s scores increase while the rates by which
students are exempted from the NAEP increase. In other words, the pool of students sampled to participate in the NAEP is more selective and therefore likely to have biased the resulting gains. If lower-scoring students are pulled from the NAEP sample, scores on the NAEP will increase. This makes for unclear results. Unclear losses are the case when a state’s scores decrease at the same time the rates by which students are exempted from the NAEP sample decrease. In this case, the pool of students sampled was less selective so it is difficult to determine whether the addition of more lower-scoring students or an actual decrease in achievement caused the resulting losses.

We believe that it is absolutely necessary to make these kinds of judgments about each state because states with high-stakes tests are those states that increasingly are exempting more students from participating in the NAEP. “In states with high-stakes tests, between 0%–49% of the gains in NAEP scores can be explained by increases in rates of exclusion.” (Amrein & Berliner, 2002a)

Looking simply at those states for which clear gains or losses are applicable, an analysis of the data yields the results given in Table 2. In this table states shaded in green are those for which clear results were evident, states shaded in red are those for which unclear results were illustrated, and states shaded in yellow are those for which there were not enough data to analyze gains or losses appropriately.

As can be seen, only two states included in the states with high-stakes column can be counted as states with “clear” effects. The composite data are not significant but the table illustrates the extent to which states with high-stakes tests are not gaining in score simply because of their high-stakes testing policies.

**Table 2**

**Fourth grade 1994-1998 NAEP reading scores with states coded as clear or unclear in their gains and losses.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona</td>
<td>206</td>
<td>207</td>
<td>Alabama</td>
<td>208</td>
<td>211</td>
</tr>
<tr>
<td>Arkansas</td>
<td>209</td>
<td>209</td>
<td>Kentucky</td>
<td>212</td>
<td>218</td>
</tr>
<tr>
<td>California</td>
<td>197</td>
<td>202</td>
<td>Louisiana</td>
<td>197</td>
<td>204</td>
</tr>
<tr>
<td>Colorado</td>
<td>213</td>
<td>222</td>
<td>Maryland</td>
<td>210</td>
<td>215</td>
</tr>
<tr>
<td>Connecticut</td>
<td>222</td>
<td>232</td>
<td>Michigan</td>
<td>n/a</td>
<td>217</td>
</tr>
<tr>
<td>Delaware</td>
<td>206</td>
<td>212</td>
<td>Mississippi</td>
<td>202</td>
<td>204</td>
</tr>
<tr>
<td>Florida</td>
<td>205</td>
<td>207</td>
<td>Missouri</td>
<td>217</td>
<td>216</td>
</tr>
<tr>
<td>Georgia</td>
<td>207</td>
<td>210</td>
<td>New Mexico</td>
<td>205</td>
<td>206</td>
</tr>
<tr>
<td>State</td>
<td>2014 Score</td>
<td>2017 Score</td>
<td>Change in Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawaii</td>
<td>200</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>223</td>
<td>223</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td>n/a</td>
<td>222</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td>228</td>
<td>225</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>223</td>
<td>225</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td>218</td>
<td>222</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montana</td>
<td>222</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td>n/a</td>
<td>208</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>223</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>212</td>
<td>216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>n/a</td>
<td>214</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>220</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utah</td>
<td>217</td>
<td>215</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>213</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td>213</td>
<td>217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>224</td>
<td>224</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td>221</td>
<td>219</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVERALL</td>
<td>215.4</td>
<td>217.0</td>
<td>+1.6*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at a p < .05 level

The composite data are important in that they nullify what one might conclude looking simply at Table 1. States with high-stakes tests are not outperforming states without high-stakes tests in reading grade 4 performance. Rather, as illustrated in Table 2, states without high-stakes tests gained in reading grade 4 performance at a statistically significant level. Given only two states are included as clear states with high-stakes tests, states with high-stakes tests made insignificant gains and the differences between the two mean gains are not statistically significant.

Most importantly, what can be drawn from Table 2 is that states with high-stakes tests are exempting more students from participating in the reading grade 4 NAEP. Ninety percent of the states with “unclear” gains are states with increases in the rates by which students were exempted from the test. This supports the notion that states with high-stakes tests are not gaining in NAEP scores simply because of
their high-stakes testing policies.

**NAEP Math Grade 4 1996-2000**

Taking Table 1 from Amrein & Berliner (2002b) and the states in which high stakes tests were implemented before 1996 and between 1996 and 2000, we re-ran our analyses using all states with high-stakes tests and, as the control group, all states without high-stakes tests for which NAEP data were available. What we found in regards to math grade 4 achievement from 1996-2000 is as follows:

**Table 3**

**Fourth grade 1996-2000 NAEP mathematics scores (raw data)**

<table>
<thead>
<tr>
<th>States without high-stakes tests:</th>
<th>NAEP 1996</th>
<th>NAEP 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>224</td>
<td>n/a</td>
</tr>
<tr>
<td>Arizona</td>
<td>218</td>
<td>219</td>
</tr>
<tr>
<td>Arkansas</td>
<td>216</td>
<td>217</td>
</tr>
<tr>
<td>Colorado</td>
<td>226</td>
<td>n/a</td>
</tr>
<tr>
<td>Connecticut</td>
<td>232</td>
<td>234</td>
</tr>
<tr>
<td>Georgia</td>
<td>216</td>
<td>220</td>
</tr>
<tr>
<td>Hawaii</td>
<td>215</td>
<td>216</td>
</tr>
<tr>
<td>Idaho</td>
<td>n/a</td>
<td>227</td>
</tr>
<tr>
<td>Illinois</td>
<td>n/a</td>
<td>225</td>
</tr>
<tr>
<td>Iowa</td>
<td>229</td>
<td>233</td>
</tr>
<tr>
<td>Kansas</td>
<td>n/a</td>
<td>232</td>
</tr>
<tr>
<td>Maine</td>
<td>233</td>
<td>231</td>
</tr>
<tr>
<td>Minnesota</td>
<td>232</td>
<td>235</td>
</tr>
<tr>
<td>Montana</td>
<td>228</td>
<td>230</td>
</tr>
<tr>
<td>Nebraska</td>
<td>228</td>
<td>226</td>
</tr>
<tr>
<td>North Dakota</td>
<td>231</td>
<td>231</td>
</tr>
<tr>
<td>Oregon</td>
<td>224</td>
<td>227</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>States with high-stakes tests:</th>
<th>NAEP 1996</th>
<th>NAEP 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>212</td>
<td>218</td>
</tr>
<tr>
<td>California</td>
<td>209</td>
<td>214</td>
</tr>
<tr>
<td>Delaware</td>
<td>215</td>
<td>n/a</td>
</tr>
<tr>
<td>Florida</td>
<td>216</td>
<td>n/a</td>
</tr>
<tr>
<td>Indiana</td>
<td>229</td>
<td>234</td>
</tr>
<tr>
<td>Kentucky</td>
<td>220</td>
<td>221</td>
</tr>
<tr>
<td>Louisiana</td>
<td>209</td>
<td>218</td>
</tr>
<tr>
<td>Maryland</td>
<td>220</td>
<td>222</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>229</td>
<td>235</td>
</tr>
<tr>
<td>Michigan</td>
<td>226</td>
<td>231</td>
</tr>
<tr>
<td>Mississippi</td>
<td>208</td>
<td>211</td>
</tr>
<tr>
<td>Missouri</td>
<td>225</td>
<td>229</td>
</tr>
<tr>
<td>Nevada</td>
<td>217</td>
<td>220</td>
</tr>
<tr>
<td>New Jersey</td>
<td>227</td>
<td>n/a</td>
</tr>
<tr>
<td>New Mexico</td>
<td>214</td>
<td>214</td>
</tr>
<tr>
<td>New York</td>
<td>223</td>
<td>227</td>
</tr>
<tr>
<td>North Carolina</td>
<td>224</td>
<td>232</td>
</tr>
</tbody>
</table>
Table 3 illustrates that the states with high-stakes tests outperformed those states without high-stakes tests on the math NAEP grade 4 tests over the time period 1996-2000. However, as argued earlier, the rates by which students are excluded from the NAEP must be taken into consideration to determine whether gains and losses are clear or unclear. Using the same rules as outlined above to determine clear and unclear gains and losses, we looked at only those states for which clear gains or losses are relevant. (For 4th grade reading, Rosenshine included the following states: Arizona, Arkansas, California, Connecticut, Hawaii, Iowa, Maine, Montana, New Hampshire, Rhode Island, Utah, Washington, Wisconsin, and Wyoming. There are notable differences in the states he included and the states we included that likely came from the fact that we drew our states directly out of Table 1 of the original document.) An analysis of the data yields the following:

### Table 4

**Fourth grade 1996-2000 NAEP mathematics scores with states coded as clear or unclear in their gains and losses.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>224</td>
<td>n/a</td>
<td>Alabama</td>
<td>212</td>
<td>218</td>
</tr>
<tr>
<td>Arizona</td>
<td>218</td>
<td>219</td>
<td>California</td>
<td>209</td>
<td>214</td>
</tr>
<tr>
<td>Arkansas</td>
<td>216</td>
<td>217</td>
<td>Delaware</td>
<td>215</td>
<td>n/a</td>
</tr>
<tr>
<td>Colorado</td>
<td>226</td>
<td>n/a</td>
<td>Florida</td>
<td>216</td>
<td>n/a</td>
</tr>
<tr>
<td>Connecticut</td>
<td>232</td>
<td>234</td>
<td>Indiana</td>
<td>229</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td>Georgia 216</td>
<td>220</td>
<td>Kentucky 220</td>
<td>221</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------</td>
<td>-----</td>
<td>-------------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>Hawaii</td>
<td>215</td>
<td>216</td>
<td>Louisiana</td>
<td>209</td>
<td>218</td>
</tr>
<tr>
<td>Idaho</td>
<td>n/a</td>
<td>227</td>
<td>Maryland</td>
<td>220</td>
<td>222</td>
</tr>
<tr>
<td>Illinois</td>
<td>n/a</td>
<td>225</td>
<td>Massachusetts</td>
<td>229</td>
<td>235</td>
</tr>
<tr>
<td>Iowa</td>
<td>229</td>
<td>233</td>
<td>Michigan</td>
<td>226</td>
<td>231</td>
</tr>
<tr>
<td>Kansas</td>
<td>n/a</td>
<td>232</td>
<td>Mississippi</td>
<td>208</td>
<td>211</td>
</tr>
<tr>
<td>Maine</td>
<td>233</td>
<td>231</td>
<td>Missouri</td>
<td>225</td>
<td>229</td>
</tr>
<tr>
<td>Minnesota</td>
<td>232</td>
<td>235</td>
<td>Nevada</td>
<td>217</td>
<td>220</td>
</tr>
<tr>
<td>Montana</td>
<td>228</td>
<td>230</td>
<td>New Jersey</td>
<td>227</td>
<td>n/a</td>
</tr>
<tr>
<td>Nebraska</td>
<td>228</td>
<td>226</td>
<td>New Mexico</td>
<td>214</td>
<td>214</td>
</tr>
<tr>
<td>North Dakota</td>
<td>231</td>
<td>231</td>
<td>New York</td>
<td>223</td>
<td>227</td>
</tr>
<tr>
<td>Oregon</td>
<td>224</td>
<td>227</td>
<td>North Carolina</td>
<td>224</td>
<td>232</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>221</td>
<td>225</td>
<td>Ohio</td>
<td>n/a</td>
<td>231</td>
</tr>
<tr>
<td>Tennessee</td>
<td>219</td>
<td>220</td>
<td>Oklahoma</td>
<td>n/a</td>
<td>225</td>
</tr>
<tr>
<td>Utah</td>
<td>226</td>
<td>227</td>
<td>Pennsylvania</td>
<td>226</td>
<td>n/a</td>
</tr>
<tr>
<td>Vermont</td>
<td>225</td>
<td>232</td>
<td>South Carolina</td>
<td>213</td>
<td>220</td>
</tr>
<tr>
<td>Washington</td>
<td>225</td>
<td>n/a</td>
<td>Texas</td>
<td>229</td>
<td>233</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>231</td>
<td>n/a</td>
<td>Virginia</td>
<td>222</td>
<td>230</td>
</tr>
<tr>
<td>Wyoming</td>
<td>223</td>
<td>229</td>
<td>West Virginia</td>
<td>224</td>
<td>225</td>
</tr>
<tr>
<td>OVERALL</td>
<td>224.5</td>
<td>225.6</td>
<td>Change in Score</td>
<td>+1.1</td>
<td>OVERALL 210.4</td>
</tr>
</tbody>
</table>

*Significant at a p < .05 level

Compared to the reading data above, we now find the opposite when we look at the math grade 4 NAEP composite data. When states with clear effects are pulled out and analyzed, it is apparent that states with high-stakes tests are outperforming states without high-stakes tests at a statistically significant level. The scores posted by the clear states with high-stakes tests are significantly different than the scores posted by the clear states without high-stakes tests.

Again, however, what can also be drawn from Table 4 is that states with high-stakes tests are exempting more students from participating in the math grade 4 NAEP. Two times as many states with high-stakes tests exempted students and realized gains in grade 4 math achievement from 1996-2000 than did states without high-stakes tests. This, again, supports the notion that states with high-stakes tests
are not all gaining in NAEP scores simply because of their high-stakes testing policies.

**NAEP Math Grade 8 1996-2000**

Taking Table 1 from Amrein & Berliner (2002b) and the states in which high stakes tests were implemented before 1996 and between 1996 and 2000, we re-ran our analyses using all states with high-stakes tests and, as the control group, all states without high-stakes tests for which NAEP data were available. What we found in regards to math grade 8 achievement from 1996-2000 is as follows:

**Table 5**

**Eighth grade 1996-2000 NAEP mathematics scores (raw data)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>278</td>
<td>n/a</td>
<td>Alabama</td>
<td>256</td>
<td>262</td>
</tr>
<tr>
<td>Arizona</td>
<td>268</td>
<td>271</td>
<td>California</td>
<td>263</td>
<td>262</td>
</tr>
<tr>
<td>Arkansas</td>
<td>261</td>
<td>261</td>
<td>Delaware</td>
<td>267</td>
<td>n/a</td>
</tr>
<tr>
<td>Colorado</td>
<td>276</td>
<td>n/a</td>
<td>Florida</td>
<td>264</td>
<td>n/a</td>
</tr>
<tr>
<td>Connecticut</td>
<td>280</td>
<td>282</td>
<td>Indiana</td>
<td>275</td>
<td>283</td>
</tr>
<tr>
<td>Georgia</td>
<td>262</td>
<td>266</td>
<td>Kentucky</td>
<td>267</td>
<td>272</td>
</tr>
<tr>
<td>Hawaii</td>
<td>262</td>
<td>263</td>
<td>Louisiana</td>
<td>252</td>
<td>259</td>
</tr>
<tr>
<td>Idaho</td>
<td>n/a</td>
<td>278</td>
<td>Maryland</td>
<td>270</td>
<td>276</td>
</tr>
<tr>
<td>Illinois</td>
<td>n/a</td>
<td>277</td>
<td>Massachusetts</td>
<td>277</td>
<td>283</td>
</tr>
<tr>
<td>Iowa</td>
<td>284</td>
<td>n/a</td>
<td>Michigan</td>
<td>276</td>
<td>278</td>
</tr>
<tr>
<td>Kansas</td>
<td>n/a</td>
<td>284</td>
<td>Mississippi</td>
<td>250</td>
<td>254</td>
</tr>
<tr>
<td>Maine</td>
<td>284</td>
<td>284</td>
<td>Missouri</td>
<td>274</td>
<td>274</td>
</tr>
<tr>
<td>Minnesota</td>
<td>284</td>
<td>288</td>
<td>Nevada</td>
<td>n/a</td>
<td>268</td>
</tr>
<tr>
<td>Montana</td>
<td>283</td>
<td>287</td>
<td>New Mexico</td>
<td>262</td>
<td>260</td>
</tr>
<tr>
<td>Nebraska</td>
<td>283</td>
<td>281</td>
<td>New York</td>
<td>270</td>
<td>276</td>
</tr>
<tr>
<td>North Dakota</td>
<td>284</td>
<td>283</td>
<td>North Carolina</td>
<td>268</td>
<td>280</td>
</tr>
<tr>
<td>Oregon</td>
<td>277</td>
<td>281</td>
<td>Ohio</td>
<td>n/a</td>
<td>283</td>
</tr>
<tr>
<td>States without high-stakes tests</td>
<td>NAEP 1996</td>
<td>NAEP 2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaska</td>
<td>278</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arizona</td>
<td>268</td>
<td>271</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td>261</td>
<td>261</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>276</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>280</td>
<td>282</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>262</td>
<td>266</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawaii</td>
<td>262</td>
<td>263</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idaho</td>
<td>n/a</td>
<td>278</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>States with high-stakes tests</td>
<td>NAEP 1996</td>
<td>NAEP 2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alabama</td>
<td>256</td>
<td>262</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>263</td>
<td>262</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td>267</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>264</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>275</td>
<td>283</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td>267</td>
<td>272</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td>252</td>
<td>259</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td>270</td>
<td>276</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 illustrates the states with high-stakes tests outperformed those states without high-stakes tests on the math NAEP grade 8 1996-2000. Again, we argue that the rates by which students are excluded from the NAEP must be taken into consideration to determine whether gains and losses are clear or unclear.

Using the same rules as outlined above to determine clear and unclear gains and losses, we looked at only those states for which clear gains or losses are apparent (Note 3). An analysis of the data yields the following:

**Table 6**

**Eighth grade 1996-2000 NAEP mathematics scores with states coded as clear or unclear in their gains and losses.**

<table>
<thead>
<tr>
<th>States without high-stakes tests</th>
<th>NAEP 1996</th>
<th>NAEP 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>278</td>
<td>n/a</td>
</tr>
<tr>
<td>Arizona</td>
<td>268</td>
<td>271</td>
</tr>
<tr>
<td>Arkansas</td>
<td>261</td>
<td>261</td>
</tr>
<tr>
<td>Colorado</td>
<td>276</td>
<td>n/a</td>
</tr>
<tr>
<td>Connecticut</td>
<td>280</td>
<td>282</td>
</tr>
<tr>
<td>Georgia</td>
<td>262</td>
<td>266</td>
</tr>
<tr>
<td>Hawaii</td>
<td>262</td>
<td>263</td>
</tr>
<tr>
<td>Idaho</td>
<td>n/a</td>
<td>278</td>
</tr>
<tr>
<td>States with high-stakes tests</td>
<td>NAEP 1996</td>
<td>NAEP 2000</td>
</tr>
<tr>
<td>Alabama</td>
<td>256</td>
<td>262</td>
</tr>
<tr>
<td>California</td>
<td>263</td>
<td>262</td>
</tr>
<tr>
<td>Delaware</td>
<td>267</td>
<td>n/a</td>
</tr>
<tr>
<td>Florida</td>
<td>264</td>
<td>n/a</td>
</tr>
<tr>
<td>Indiana</td>
<td>275</td>
<td>283</td>
</tr>
<tr>
<td>Kentucky</td>
<td>267</td>
<td>272</td>
</tr>
<tr>
<td>Louisiana</td>
<td>252</td>
<td>259</td>
</tr>
<tr>
<td>Maryland</td>
<td>270</td>
<td>276</td>
</tr>
</tbody>
</table>

*Significant at a p < .05 level
<table>
<thead>
<tr>
<th>State</th>
<th>Change in Score</th>
<th>Score 1</th>
<th>Score 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illinois</td>
<td></td>
<td>277</td>
<td>277</td>
</tr>
<tr>
<td>Iowa</td>
<td></td>
<td>284</td>
<td>n/a</td>
</tr>
<tr>
<td>Kansas</td>
<td></td>
<td>n/a</td>
<td>284</td>
</tr>
<tr>
<td>Maine</td>
<td></td>
<td>284</td>
<td>284</td>
</tr>
<tr>
<td>Minnesota</td>
<td></td>
<td>284</td>
<td>288</td>
</tr>
<tr>
<td>Montana</td>
<td></td>
<td>283</td>
<td>287</td>
</tr>
<tr>
<td>Nebraska</td>
<td></td>
<td>283</td>
<td>281</td>
</tr>
<tr>
<td>North Dakota</td>
<td></td>
<td>284</td>
<td>283</td>
</tr>
<tr>
<td>Oregon</td>
<td></td>
<td>277</td>
<td>281</td>
</tr>
<tr>
<td>Rhode Island</td>
<td></td>
<td>268</td>
<td>273</td>
</tr>
<tr>
<td>Tennessee</td>
<td></td>
<td>263</td>
<td>263</td>
</tr>
<tr>
<td>Utah</td>
<td></td>
<td>276</td>
<td>275</td>
</tr>
<tr>
<td>Vermont</td>
<td></td>
<td>279</td>
<td>283</td>
</tr>
<tr>
<td>Washington</td>
<td></td>
<td>276</td>
<td>n/a</td>
</tr>
<tr>
<td>Wisconsin</td>
<td></td>
<td>283</td>
<td>n/a</td>
</tr>
<tr>
<td>Wyoming</td>
<td></td>
<td>275</td>
<td>277</td>
</tr>
<tr>
<td>OVERALL</td>
<td></td>
<td>271.1</td>
<td>271.9</td>
</tr>
<tr>
<td>Massachusetts</td>
<td></td>
<td>277</td>
<td>283</td>
</tr>
<tr>
<td>Michigan</td>
<td></td>
<td>276</td>
<td>278</td>
</tr>
<tr>
<td>Mississippi</td>
<td></td>
<td>250</td>
<td>254</td>
</tr>
<tr>
<td>Missouri</td>
<td></td>
<td>274</td>
<td>274</td>
</tr>
<tr>
<td>Nevada</td>
<td></td>
<td>n/a</td>
<td>268</td>
</tr>
<tr>
<td>New Mexico</td>
<td></td>
<td>262</td>
<td>260</td>
</tr>
<tr>
<td>New York</td>
<td></td>
<td>270</td>
<td>276</td>
</tr>
<tr>
<td>North Carolina</td>
<td></td>
<td>268</td>
<td>280</td>
</tr>
<tr>
<td>Ohio</td>
<td></td>
<td>n/a</td>
<td>283</td>
</tr>
<tr>
<td>Oklahoma</td>
<td></td>
<td>n/a</td>
<td>272</td>
</tr>
<tr>
<td>South Carolina</td>
<td></td>
<td>260</td>
<td>266</td>
</tr>
<tr>
<td>Texas</td>
<td></td>
<td>270</td>
<td>275</td>
</tr>
<tr>
<td>Virginia</td>
<td></td>
<td>270</td>
<td>277</td>
</tr>
<tr>
<td>West Virginia</td>
<td></td>
<td>265</td>
<td>271</td>
</tr>
</tbody>
</table>

After the states with clear effects are pulled out and analyzed, it seems that states with high-stakes tests are outperforming states without high-stakes tests. They are not, however, outperforming states without high-stakes tests at a statistically significant level. In addition, the scores posted by the clear states with high-stakes tests are not significantly different than the scores posted by the clear states without high-stakes tests. States with high-stakes tests are not outperforming states without high-stakes tests in math grade 8 performance.

Again, what can also be drawn from Table 6 is that states with high-stakes tests are exempting more students from participating in the math grade 8 NAEP. Thirty-three percent of the states without high-stakes tests exempted more students and realized gains in math grade 8 NAEP scores. Fifty percent of the states with high-stakes tests exempted more students and realized gains in math grade 8 NAEP scores. This, again, supports our assertion that states with high-stakes tests are not gaining in NAEP scores simply because of their high-stakes testing policies.

**Conclusion**

In short, states with high-stakes tests seem to have outperformed states without
high-stakes tests on the grade 4 math NAEP at a statistically significant level. However, gains between states with and without high stakes tests were not statistically different on the grade 4 reading or the grade 8 math NAEP. States with high-stakes tests are not outperforming states without high-stakes tests on both of these measures.

In addition, the rates by which personnel in states with high-stakes tests are exempting students are increasing at a faster rate than they are in states without high-stakes tests. There may be an underlying characteristic other than high-stakes tests that is causing this phenomenon, but this would take further analyses. What we do know, however, is that for the most part the gains posted by states with high-stakes tests on two of the three NAEP tests are more related to the rates by which students are exempted from the tests than they are related to high-stakes tests themselves.

We thank Professor Rosenshine for suggesting these alternative analytic techniques to us. In the end, for now, we remain unconvincing that the NAEP tests are showing much in the way of transfer effects. Given all the data we reported in our previous reports we remain unconvincing that the high-stakes tests used by states are showing systematic positive affects on audit tests used to assess transfer.

References


About the Authors

Audrey Amrein-Beardsley
Email: audrey.beardsley@cox.net

Audrey Amrein-Beardsley is a part-time Research and Evaluation Associate at a private foundation in Scottsdale, Arizona and a part-time researcher at Arizona State University. She received her PhD from Arizona State University in 2002 in Education Policy with an emphasis in Research Methodology. Her scholarly interests include the study of the intended and unintended consequences of high-stakes testing policies. To date, she has focused on the effects of high-stakes testing policies at a macro level given the frequency with which these policies are
being implemented across the nation.

David C. Berliner  
Regents’ Professor of Education  
College of Education  
Arizona State University  
Tempe, AZ 85287-2411  
Email: berliner@asu.edu

David C. Berliner is Regents’ Professor of Education at the College of Education of Arizona State University, in Tempe, AZ. He received his Ph.D. in 1968 from Stanford University in educational psychology, and has worked also at the University of Massachusetts, WestEd, and the University of Arizona. He has served as president of the American Educational Research Association (AERA), president of the Division of Educational Psychology of the American Psychological Association, and as a fellow of the Center for Advanced Study in the Behavioral Sciences and a member of the National Academy of Education. Berliner’s publications include The Manufactured Crisis, Addison-Wesley, 1995 (with B.J. Biddle) and The Handbook of Educational Psychology, Macmillan, 1996 (Edited with R.C. Calfee). Special awards include the Research into Practice Award of AERA, the National Association of Secondary School Principals Distinguished Service Award, and the Medal of Honor from the University of Helsinki. His scholarly interests include research on teaching and education policy analysis.

The World Wide Web address for the Education Policy Analysis Archives is epaa.asu.edu

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple  
University of Wisconsin

Greg Camilli  
Rutgers University

Sherman Dorn  
University of South Florida

David C. Berliner  
Arizona State University

Linda Darling-Hammond  
Stanford University

Mark E. Fetler  
California Commission on Teacher Credentialing
Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseis.ucla.edu

EPAA is published by the Education Policy Studies Laboratory, Arizona State University