Assessments for instructional feedback and accountability are powerful twin engines for educational improvement. As several exhaustive literature reviews have documented, assessments that provide feedback to teachers about their students’ performance are one of the highest octane technologies known to fuel student performance growth (Black & Wiliam, 1998; Crooks, 1988; Natriello, 1987). While formative assessments can be powerful dynamos for instructional improvement, assessments for accountability purposes are the driving force behind state and Federal education policy today. On the policy highway, accountability testing is king of the road. The irony of this situation is that neither assessment engine can run as efficiently without the other, yet they do not comfortably coexist. A feedback system without external indicators of progress lacks both incentives and direction. Alternatively, a system that measures progress without providing constructive feedback does little to build capacity for improvement. Unfortunately, as policymakers have been generally unsuccessful integrating multiple purposes into the same assessment system, the two operate on separate tracks (Supovitz, 1997). The lack of counter-balancing policy attention to formative assessment systems for instructional improvement is what makes our obsession with educational accountability so incongruous.

The 104th yearbook of the National Society for the Study of Education, edited by Joan L. Herman and Edward H. Haertel, is an invaluable owner’s manual for those interested in understanding the roots, technical aspects, and effects of the accountability engine. Herman and Haertel have put together a star-studded contributor team made up largely of a national network of colleagues affiliated with the Center for Research on Evaluation, Standards, and Student Testing (CRESST),
headquartered at the University of California at Los Angeles. In the yearbook, the contributors kick the tires, get under the hood, and put various accountability models through their paces. This yearbook is a veritable consumer report of assessment and accountability systems.

The 380 page yearbook is divided into four sections. The first section focuses on several of the core ambiguities that beleaguer attempts to build a tight assessment system. Haertel and Herman open the section by taking readers on a spin through American assessment history. Beginning with the well documented birth of educational testing as it diverged from its roots in intelligence testing, and proceeding through eras of measurement driven instruction, minimum competency testing, and experiments with alternative assessments, the authors take us up through the standards movement and the theory behind President Bush’s watershed No Child Left Behind Legislation and its focus on annual high stakes testing. Along the way, the authors point out some of the critical distinctions between group accountability and individual learning and the cyclical disappointment that perenni ally seems to follow unfulfilled reform ambitions that are centered on assessments. In the second part of the section, Lorraine McDonnell of UC Santa Barbara provides a clear-eyed analysis of the political elements of assessment policy. McDonnell unravels the logic of high stakes tests and points out the political incentives that drive politicians to continually reach for the assessment lever. Finally, Judith Ramaley of the University of Maine points out some of the crucial uncertainties about memory, learning, transfer, and the purposes of education that underlie many of the dilemmas of testing. She then sketches an ambitious picture of what a more robust assessment system might look like to reflect a broader notion of an educated citizenry.

The second section of the yearbook is devoted to the down and dirty technical issues of assessment design and analysis that all education grease monkeys will love. The overall theme of the section is to point out the weak joints and thin welds of current assessment technologies that make them poorly equipped to bear the weight that educational policymakers would like to place upon them. Robert Linn of the University of Colorado at Boulder, the Richard Petty of assessment researchers, revs up the section with an analysis of accountability system design issues. Harvard’s Dan Koretz empirically illustrates validity concerns with a description of a district in which scores dropped, rose and dropped again with the institution of roughly equivalent tests, suggesting that student performance was more sensitive to the test itself rather than a reasonable reflection of student learning. UCLA researchers Kilchan Choi, Pete Goldsmidt and Kyo Yamashiro present a smart analysis of the benefits of value added systems over simpler methods such as proficiency cut scores or unadjusted gains. They provide a brief description of four different value added approaches, pointing out the merits and constraints of each. David Rogosa, deriding the poor quality of the statistical work around educational research on assessment, attempts to explode some of the common misunderstandings about accountability systems. Rogosa debunks misconceptions about such issues as the error associated with school-level scores, claims of volatility in year-to-year improvements, the effects of school size on performance, and the influence of demographic characteristics on school performance.

The third section of the yearbook moves from issues of design to the consequences of accountability systems for different subgroups of students. Jamal Abedi of CREST at UCLA discusses issues and consequences for students whose first language is not English, also called English Language Learners. Diana Pullin of Springer