

Boston University School of Law

Scholarly Commons at Boston University School of Law

Faculty Scholarship

2011

Testing as Commodification

Katharine B. Silbaugh

Boston University School of Law

Follow this and additional works at: https://scholarship.law.bu.edu/faculty_scholarship



Part of the [Education Law Commons](#), [Intellectual Property Law Commons](#), and the [Law and Economics Commons](#)

Recommended Citation

Katharine B. Silbaugh, *Testing as Commodification*, in 35 *Washington University Journal of Law & Policy* 309 (2011).

Available at: https://scholarship.law.bu.edu/faculty_scholarship/1668

This Article is brought to you for free and open access by Scholarly Commons at Boston University School of Law. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Scholarly Commons at Boston University School of Law. For more information, please contact lawlessa@bu.edu.



Testing as Commodification

Katharine Silbaugh*

TABLE OF CONTENTS

I. INTRODUCTION: COMMODIFICATION IN EDUCATION REFORM.....	310
II. COMMODIFICATION AS A SINGLE METRIC OF VALUE	313
III. THE EDUCATION REFORM MOVEMENT	317
A. <i>Equity</i>	318
B. <i>Metrics and Commensurability as Translation Between Educators and Policymakers or Parents</i>	320
IV. HOW TESTS CHANGE SCHOOLS: WHAT IS “TEACHING TO THE TEST”?.....	323
A. <i>Narrowing the Curriculum</i>	324
B. <i>“Testing Gains, Not Learning Gains”</i>	326
C. <i>Corrupted Schools</i>	327
V. THE HISTORY OF PLURALISM IN EDUCATION: DOES QUALITY EDUCATION MEAN UNIFORM EDUCATION?.....	330
VI. COMMON POWER IN THE TWO METRICS: MONEY AND TEST SCORES.....	332
A. <i>Numerosity</i>	333
B. <i>Creation of Workers</i>	334
VII. CONCLUSION.....	336

“[T]he tests-and-standards movement . . . ha[s] been loaded with a coarse utilitarian toxicity and a demeaning anti-human view of childhood right from the start.”

—Jonathan Kozol¹

* Professor of Law and Law Alumni Scholar, Boston University School of Law. I wish to thank Hallie Marin for helpful research assistance, and the participants in the Love or Money conference at Washington University School of Law for helpful discussions.

1. Jonathan Kozol, *Standardized Testing: The Do-or-Die Agenda*, PRINCIPAL, Mar./Apr. 2006, at 18, 22.

I. INTRODUCTION: COMMODIFICATION IN EDUCATION REFORM

This Essay argues that criticism of the testing movement by those with expertise in education, such as the renowned anti-poverty activist Jonathan Kozol quoted above, reference concepts and concerns that map remarkably well onto philosophical discussions of commodification as well as behavioral economics discussions of intrinsic motivations. This Essay explores both the similarities among these discourses and the possible insights that may be brought to the education debate from the commodification literature in philosophy and economics.

The philosophy scholar Elizabeth S. Anderson argues that to value something differently than as a commodity is to recognize a “special intrinsic worth” to that item.² If it is appropriate to apply “use” as the proper mode of valuation, then market norms are acceptable. But if a different mode of valuation is appropriate, such as “love, admiration, honor, and appreciation,”³ then we should not treat that item as a commodity.⁴ “Use” as a value is simply utilitarian, while the other modes require a more nuanced version of value. Martha C. Nussbaum has written of the “incommensurable plurality of values”⁵—one item might be measurable in a valuation scale translatable to dollars (market value), while a different item is valued in terms that cannot enjoy a common metric with the first item (e.g., respect, love).⁶ The two are incommensurable: because they are valued on different scales, their value cannot be compared to one another. The commodification concern can be stated a number of ways, but it ordinarily contrasts the language and norms of the market with language, norms, and understandings that seem incompatible with the market.⁷ While the commodification concern has proved vulnerable

2. Elizabeth S. Anderson, *Is Women's Labor a Commodity?*, 19 PHIL. & PUB. AFF. 71, 73 n.2 (1990).

3. *Id.* at 72.

4. ELIZABETH ANDERSON, VALUE IN ETHICS AND ECONOMICS 8–11 (1993).

5. MARTHA C. NUSSBAUM, LOVE'S KNOWLEDGE: ESSAYS ON PHILOSOPHY AND LITERATURE 112 (1990).

6. *Id.* at 106–24.

7. The commodification literature is extensive. In addition to Anderson and Nussbaum, see, for example, MARGARET JANE RADIN, CONTESTED COMMODITIES (1996); HILARY PUTNAM, REASON, TRUTH AND HISTORY (1981); MICHAEL WALZER, SPHERES OF JUSTICE: A

to an array of legitimate criticism,⁸ it has proved resilient in describing a difficulty utilitarianism has accounting for plural values under some circumstances.

This Essay looks at the movement within public education toward common standards and assessments represented most prominently by the No Child Left Behind Act of 2001 (NCLB).⁹ This movement is characterized by a drive to create common metrics in the form of test scores for evaluating the quality of educational programming within each state. This movement invites measurement based on this metric and invites comparisons among districts, schools, teachers, and students along this same scale. Writing on the standards movement is split between those explicating the virtues of a common metric by which to make comparisons, measure progress, and correct stagnation and those anxious that test scores have swallowed other notions of the good in public education. The latter concern is well summarized in the words of educational equity author Jonathan Kozol, “the tests-and-standards movement . . . ha[s] been loaded with a coarse utilitarian toxicity and a demeaning anti-human view of childhood right from the start.”¹⁰

In this regard, the standards-and-testing debate mimics many familiar concerns from the commodification debate within philosophy and law. But the debate over the testing movement represents an interesting variation because tests scores play the role that prices do in the commodification literature. The commodification debate juxtaposes market valuation with all other forms of valuation, while the standards and accountability debate juxtaposes the non-market metric of test scores with more plural conceptions of educational purpose, quality, and outcomes.

DEFENSE OF PLURALISM AND EQUALITY (1983); JAMES B. WHITE, HERACLES' BOW: ESSAYS ON THE RHETORIC AND POETICS OF THE LAW (1985); VIVIANA A. ZELIZER, THE PURCHASE OF INTIMACY (2005) [hereinafter ZELIZER, THE PURCHASE OF INTIMACY]; VIVIANA A. ZELIZER, PRICING THE PRICELESS CHILD: THE CHANGING SOCIAL VALUE OF CHILDREN (1985); RETHINKING COMMODIFICATION: CASES AND READINGS IN LAW AND CULTURE (Martha M. Ertman & Joan C. Williams eds., 2005) [hereinafter RETHINKING COMMODIFICATION].

8. For an early example and discussion of the critical literature, see Katharine Silbaugh, *Commodification and Women's Household Labor*, 9 YALE J.L. & FEMINISM 81 (1997).

9. No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 stat. 1425 (codified in scattered sections of 20 U.S.C.).

10. Kozol, *supra* note 1, at 22.

The behavioral economics literature on motivations speaks to this cross-discipline project as well.¹¹ In particular, the “crowding out” effect may be a reason for concern in a testing-based educational culture. Crowding-out occurs when extrinsic motivation, particularly receipt of a payment, diminishes intrinsic motivations.¹² Crowding-out has been demonstrated even where the extrinsic motivation is non-monetary. This suggests potential applicability of crowding-out theory to student motivation to learn and school personnel motivation to teach where testing values dominate. For example, extrinsic monitoring of task completion diminishes intrinsic motivation to complete tasks, even though that extrinsic monitoring does not entail monetary rewards.¹³ This behavioral economics literature may give some empirical basis for a phenomenon long observed and discussed in the commodification literature—that a good can be changed when measured according to a metric that does not capture incommensurable modes of valuation.¹⁴

Several studies have found that paying students for test score improvement lowers test scores below the baseline once the payment is withdrawn.¹⁵ Although these studies do not demonstrate that testing alone crowds-out intrinsic motivations in education because they evaluate payment for testing, they do indicate that there are vulnerable intrinsic motivations available in the school environment under some circumstances. Where there are intrinsic motivations, external monitoring of learning through standardized testing may risk crowding out intrinsic motivations to learn. This perverse effect

11. See generally RICHARD H. THALER & CASS R. SUNSTEIN, *NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS* (2008).

12. Margit Osterloh, Jetta Frost & Bruno S. Frey, *The Dynamics of Motivation in New Organizational Forms*, 9 INT’L J. ECON. BUS. 61, 67–68 (2002); BRUNO S. FREY, *NOT JUST FOR THE MONEY: AN ECONOMIC THEORY OF PERSONAL MOTIVATION* 14–15 (1997).

13. David Dickinson & Marie-Claire Villeval, *Does Monitoring Decrease Work Effort? The Complementarity Between Agency and Crowding-Out Theories*, 63 GAMES & ECON. BEHAV. 56, 57 (2008).

14. Part II discusses the approach in the commodification literature.

15. Eric P. Bettinger, *Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores*, CESIFO (Mar. 27, 2008), http://www.cesifo-group.de/portal/page/portal/ifoContent/N/neucesifo/CONFERENCES/SC_CONF_2008/ei08/Papers/ei08_Bettinger.pdf; Ellen Garbarino & Robert Slonim, *Pay-to-Perform Educational Policy: An Experiment Examining a Crowding Out Effect*, UC SANTA BARBARA ECON. (Mar. 2005), <http://www.econ.ucsb.edu/conferences/charness06/docs/slonim.pdf>.

achieved through the transformation of motivations and values is described, using different language, in both the commodification literature within philosophy and the behavioral economics literature.

This Essay explores the standards and assessment movement in K-12 public education, and compares it to the commodification literature around market values and norms and the crowding-out literature. The comparison allows us to ask whether it is market values alone that are a problem, or instead common metrics of any sort. It lets us consider the insights from the commodification literature when evaluating the potential harms of the testing movement. Incommensurability—the inability to rank values—is a good threatened by both pricing and test scores. By positing “testing as commodification,” this Essay seeks to raise questions for further investigation about the relationship between broad and pluralistic educational goals and values, the pressures toward measurement and a single metric that standardized testing establishes, crowding out, and the commodification literature on the corruption of value.

II. COMMODIFICATION AS A SINGLE METRIC OF VALUE

In his Tanner Lectures on Human Values at Oxford University, Michael Sandel asks us to distinguish two aspects of commodification critiques: those concerned about coercion and those concerned about corruption.¹⁶ When an argument is made to prohibit market exchanges that may exploit a financially needy person into selling something personally difficult to part with, such as sex or surrogacy, the argument is first about coercion. But the pro-market response is that she chooses to make a trade to improve her situation by her own measure, and so the idea of exploitation or coercion must either be better theorized with conditions for meaningful consent, or in the alternative, abandoned. In a perfect market, a trade cannot be coerced.

16. Michael J. Sandel, Prof. of Gov't, Harvard Univ., *What Money Can't Buy: The Moral Limits of Markets*, Lecture at Brasenose College, Oxford (May 11–12, 1998), in *THE TANNER LECTURES ON HUMAN VALUES* 89, 94 (1998), available at <http://www.tannerlectures.utah.edu/lectures/documents/Sandel100.pdf>.

The concern about corruption, however, is not as easily answered with a hypothetically perfected market. Sandel explains this objection as follows: “[C]ertain moral and civic goods are diminished or corrupted if bought and sold for money.”¹⁷ This is because the proper “mode of valuation,” in Elizabeth Anderson’s terminology, is passed by in favor of the market mode of valuation. The good is corrupted, its value improperly assessed and acted upon, because the wrong *mode* of valuation is used (not the wrong ranking on a numeric metric). This theoretical contention is enhanced by the empirical one demonstrated in the economics literature on motivations: pay as “extrinsic motivation” can “crowd-out,” meaning diminish, intrinsic motivation, rather than complementing it. The behavioral economics studies demonstrate that motivations are corrupted, meaning altered and degraded, by the introduction of monetary motivations.

Anderson’s contention that modes of valuation differ immediately invites controversy over the proper mode of valuation for any given matter—controversy that’s difficult to resolve to everyone’s satisfaction in hard cases. When Anderson applied her idea of improper modes of valuation to paid surrogacy, for example, some questioned whether her idea of the proper alternative mode of valuation—emotional bonds between mother and child—are unduly maternalist and foreclose other potential valences of reproductive labor.¹⁸ Using language such as “intrinsic worth” to describe reproduction invites an essentialism concern that many feminists find objectionable. Some have also expressed concern about an elitism in denying wages for the purpose of preserving meaning.¹⁹ In part for this reason, Margaret Radin argued powerfully for the importance of *plural meanings*,²⁰ and Viviana Zelizer for avoiding either-or characterizations.²¹ The notion of worth and value need not be intrinsic to be robust and plural, however, and Anderson’s

17. *Id.*

18. Martha M. Ertman, *What’s Wrong with a Parenthood Market? A New and Improved Theory of Commodification*, in *RETHINKING COMMODIFICATION*, *supra* note 7, at 303; Katharine Silbaugh, *Commodification and Women’s Household Labor*, 9 *YALE J.L. & FEMINISM* 81, 104–07 (1997).

19. *See, e.g.*, Silbaugh, *supra* note 8, at 104–07.

20. RADIN, *supra* note 7, at 103, 107 (1996).

21. ZELIZER, *THE PURCHASE OF INTIMACY*, *supra* note 7.

philosophical characterizations are echoed in some economic literature documenting incentive effects contrary to what a rational actor model might suppose.²²

From the rich literature on commodification we might draw an important lesson about the corrupting influence of markets. We should not overstate the harms of positing that money can be a fruitful way to express aspects of certain social practices that have resisted commodification.²³ To the contrary, expressions of market attributes can at times have a liberating effect on regressive and inequalitarian social practices.²⁴ But there is a danger that market valuations may be so powerful that they will extinguish or significantly diminish alternative modes of valuation and metrics that could co-exist with market valuations. Radin expressed this concern early on using the term “domino theory”²⁵ and still considers it an important caution.²⁶ Uniformity of conception is the danger, and the literature on commodification conceives of uniformity in market terms. Market modes of valuation have trouble co-existing with other forms of valuation. A question worth investigating is why some modes of valuation would so overpower others—why market valuation is *particularly* unifying, why payment would at times diminish motivation instead of enhancing it.

It is this question—how one mode of valuation threatens to so colonize others as to either extinguish or thoroughly transform them—that can be mapped onto the standards and testing movements that have gripped the education world for several decades. In undertaking to compare the two, we may see the unifying power of the market in a different light. In the education context we can see that it is *not* the norms and values of the market that have the power to extinguish other values, but it is test scores. Education and markets

22. See, e.g., RICHARD TITMUSS, *THE GIFT RELATIONSHIP: FROM HUMAN BLOOD TO SOCIAL POLICY* (1971).

23. See generally Ertman, *supra* note 18, at 304; Martha B. Coven, *The Freedom to Spend: The Case for Cash-Based Public Assistance*, 86 MINN. L. REV. 847 (2002); Carol Sanger, *Developing Markets in Baby-Making: In the Matter of Baby M*, 30 HARV. J.L. & GENDER 67 (2007); Nancy C. Staudt, *Taxing Housework*, 84 GEO. L.J. 1571, 1575 (1996).

24. Ertman, *supra* note 18, at 305; Silbaugh, *supra* note 18, at 83.

25. RADIN, *supra* note 7, at 95–101.

26. Margaret Jane Radin & Madhavi Sunder, *The Subject and Object of Commodification*, in *RETHINKING COMMODIFICATION*, *supra* note 7, at 8, 17.

have the existence of a unitary metric in common—price for markets, test scores for education. Both unitary metrics use numbers, and so both clearly invite ranking and comparison that is difficult to achieve with non-monetary values such as respect, self-knowledge, or courage.

While there are education scholars who argue against testing altogether,²⁷ most critics of the standards movement believe in the importance of assessment to improve their teaching practice by looking for evidence of progress and learning. But as the standards and testing movement has continued to increase in force, many in the education world complain about its homogenizing effect.²⁸ Like the concern about the domino effect, educators worry that increasing the significance of a single metric in the form of standardized test results is extinguishing other values in education that are important to the purposes of public education but that are not susceptible to that single metric.²⁹

The thesis of this Essay is that commodification anxiety does not depend on markets but rather on the unifying force of single metrics. What Sandel calls the “corruption risk” in commodification discourse³⁰ is robust and visible in anxieties about education reform.³¹ Similarly, the “crowding-out” effect demonstrated in the behavioral economics literature may be frustrating the goals of reform as students’ intrinsic motivation declines in the face of broad, uniform assessment. In this sense education reform is corrupting value in public education.

27. Alfie Kohn, *Raising the Scores, Ruining the Schools*, FAIRTEST (Dec. 19, 2007), <http://www.fairtest.org/raising-scores-ruining-schools>.

28. Elliot W. Eisner, *Multiple Intelligences: Its Tensions and Possibilities*, 106 TCHRS. C. REC. 31, 33–34 (2004).

29. Alfie Kohn, *Debunking the Case for National Standards*, EDUC. WK., Jan. 14, 2010, at 28, 28; Qiuyun Lin, *Beyond Standardization: Testing and Assessment in Standards-Based Reform*, ACTION TCHR. EDUC., Winter 2002, at 43, 44–45.

30. Sandel, *supra* note 16.

31. Tsilly Dagan has recently made a similar point in a draft entitled “Commodification Without Money,” in which she argues that government regulations from taxation to healthcare regulations employ common metrics that are reductive and flattening. Tsilly Dagan, *Commodification Without Money* (Bar-Ilan Univ. Law Sch., Working Paper No. 03-10, 2010), available at http://papers.ssrn.com/Sol3/papers.cfm?abstract_id=1537586.

III. THE EDUCATION REFORM MOVEMENT

In 1983, the National Commission on Excellence in Education published the report entitled *A Nation at Risk*.³² The report enhanced fears that the public education system in the United States was failing relative to other developed countries.³³ From the time of its publication forward, standards-based reform and test-based accountability have been at the center of education reform debates and in recent years at the center of education reform practices.

Standards-based reforms aim to redouble the academic mission in schools by creating content standards, aligning those with educational outcome measures through standardized testing, requiring teacher qualifications aligned with standards, and holding schools and teachers accountable for outcomes. Demanding content, testing to measure outcomes, and accountability for student achievement have become central to education policy.

Of these reforms, the most visible are the outcome tests themselves. Reform pressures and incentives throughout the 1990s, culminating in the passage of NCLB, have led to the development of robust testing regimes in all fifty states. While each state sets its own content standards and develops its own assessments and cut scores to measure progress, the National Assessment of Educational Progress (NAEP) is used to ensure that states do not set their standards too low.³⁴ Children are tested at least once each year in grades 3-8, and once between 10-12.³⁵ The high school tests are ordinarily “high-stakes,” meaning students must pass them in order to receive a high school diploma. NCLB focuses on testing in math and language arts, with science assessments added only recently.³⁶

Every state sets a definition of “proficient” at these subject areas for each grade and creates a test to assess each student’s

32. NAT’L COMM’N ON EXCELLENCE IN EDUC., *A NATION AT RISK: THE IMPERATIVE FOR EDUCATIONAL REFORM* (1983), available at http://datacenter.spps.org/sites/2259653e-ffb3-45ba-8fd6-04a024ecf7a4/uploads/SOTW_A_Nation_at_Risk_1983.pdf [hereinafter *A NATION AT RISK*].

33. *Id.*

34. *NAEP Overview*, NAT’L CENTER FOR EDUC. STAT., <http://nces.ed.gov/nationsreportcard/about/> (last visited Jan. 17, 2011).

35. No Child Left Behind Act of 2001, 20 U.S.C. § 6311(b)(3)(C) (2006).

36. *Id.* § 6311(b)(1)(C).

proficiency.³⁷ By the year 2014, NCLB requires every school in every state to bring every child to proficiency, as defined by that state.³⁸ Each year between the law's passage and the target year 2014, each school must make "adequate yearly progress" (AYP) toward that proficiency goal.³⁹ A school is making adequate yearly progress if the percentage of students whose test scores are above that line increases at a rate that could mathematically achieve 100 percent by 2014, according to annual benchmarks the state has set for itself.⁴⁰

The law includes a series of increasingly serious sanctions for failure to make AYP, including the firing of staff, reorganization of the governance structure of a school (e.g., turning it into a charter), and closing the school altogether.⁴¹ Because the targets are ambitious, an increasing percentage of schools are labeled "failing" each year as we approach 2014.⁴² In 2009, approximately a third of the nation's schools were labeled "failing," and in states with high standards such as Massachusetts, more than half of the state's schools are now failing as measured by the Act—despite that state's comparative reputation for excellence in public education.⁴³ It would be hard to overstate the significance of these developments to the culture of public education.

A. Equity

Proponents of the law focus in particular on its equity component, aimed at closing the achievement gap between white students and students of color, as well as other disadvantaged students such as English language learners and low-income students.⁴⁴ To make adequate yearly progress, it is not enough for a school to meet the

37. *Id.* § 6311(b)(1)(D).

38. *Id.* § 6311(b)(2)(F).

39. *Id.* § 6311(b)(2)(B).

40. *Id.* § 6311(b)(2)(A)–(C).

41. *Id.* § 6316.

42. Lynn Olson, *As AYP Bar Rises, More Schools Fail*, EDUC. WK., Sept. 20, 2006, at 1.

43. CTR. ON EDUC. POL'Y, HOW MANY SCHOOLS AND DISTRICTS HAVE NOT MADE ADEQUATE YEARLY PROGRESS? FOUR-YEAR TRENDS 9 (2010), available at <http://www.cep-dc.org/index.cfm?DocumentSubSubTopicID=8>.

44. NO CHILD LEFT BEHIND AND THE REDUCTION OF THE ACHIEVEMENT GAP: SOCIOLOGICAL PERSPECTIVES ON FEDERAL EDUCATIONAL POLICY (Alan R. Sadovnik et al. eds., 2007).

increasing percentage targets for its student body as a whole. The law requires schools to report test outcomes by a number of subgroups, including race, special education, limited English proficiency, and low-income classifications.⁴⁵ The school must bring each sub-group to the annual benchmark percentage.⁴⁶

This equity component means that a school cannot make adequate yearly progress by increasing the percentage of typically-developing middle income white students who reach proficiency at a sufficient rate to compensate for stagnating percentages of English language learners, low income students, special needs students, or racial minorities. This provision of the law requires schools to focus on closing the traditional achievement gaps among these groups. NCLB is thus widely understood to contain laudable equity goals intended to focus schools on improving the performance of student sub-groups that have been “left behind” in the U.S. education system. Many reformers have observed that the United States education system is already adequate in many suburban districts, and the most appalling schools are almost exclusively attended by students with little social capital, particularly in urban school systems.⁴⁷ But even in suburban schools, the achievement gap by race persists, and NCLB prevents such schools from self-congratulatory assessments about the achievement of its white students if its African American students, for example, are not also increasing their movement toward 100 percent proficiency.

The irony of this equity component of NCLB, however, is clear to those who study high-poverty districts. It is upon those students in particular—the ones historically so badly underserved by public education—that the most negative corrupting effects of the standards movement fall. While suburban schools are likely to retain music, art, critical thinking, research papers and other project-based work because they can do so and still achieve AYP, poorer districts have been reduced to “drill-and-kill” test prep to the exclusion of other

45. 20 U.S.C. § 6311(b)(2)(C)(v)(II).

46. *Id.* § 6493.

47. See JONATHAN KOZOL, *THE SHAME OF THE NATION: THE RESTORATION OF APARTHEID SCHOOLING IN AMERICA* (2005).

aspects of teaching that are not susceptible to clear measurement.⁴⁸ The flattening effect on curriculum and content is not evenly distributed but falls most heavily on those groups historically disadvantaged in our education system. Because the standards movement does not aim to close the achievement gap by evening resources among wealthier and poorer districts, the decisions made in poorer districts must be different than those made in wealthier districts, and the poorer districts are therefore more sensitive to the negative effects of teaching to the test.

B. Metrics and Commensurability as Translation Between Educators and Policymakers or Parents

The achievement goals of NCLB were set in the political arena, not by schools or educators. This was the point: those who are politically accountable felt that schools needed external motivation, in the form of accountability, to deliver great education. As critics of the law have pointed out, the achievement goals were not set in relation to an assessment of what schools have the capacity to achieve⁴⁹ but in relation to election cycles and other exogenous pressures.

Most followers of the education reform movement would have to acknowledge a strand of distrust between policymakers and educators.⁵⁰ Crudely characterized, many policymakers believe schools need external pressure—carrots and sticks—to do better. School personnel, crudely characterized, believe that policymakers are mandating performance from schools that includes curing social ills not within a school's power to control—like mandating an end to poverty itself. Further, many school personnel feel that policymakers are forcing them to eliminate valuable educational materials that are

48. See *id.*; Monty Neill, *A Child Is Not a Test Score: Assessment as a Civil Rights Issue*, ROOT & BRANCH, Fall 2009, at 28, available at <http://www.fairtest.org/files/root%20%20branch%20fall%20-%20MN%20testing%20only.pdf>.

49. Richard F. Elmore, *The Problem of Capacity in the (Re)Design of Educational Accountability Systems*, in *NCLB AT THE CROSSROADS: REEXAMINING THE FEDERAL EFFORT TO CLOSE THE ACHIEVEMENT GAP* 230, 230 (Michael A. Rebell & Jessica R. Wolff eds., 2009).

50. See Steven Brill, *The Rubber Room*, NEW YORKER, Aug. 31, 2009, at 30.

not on the test.⁵¹ Surely many policymakers engaged in education reform view teachers and administrators as sincere and motivated. But other policymakers whose theory of motivation is tied to markets, particularly in the political arena, believe that schools have a motivation issue because they are not subject to market demands.⁵² Ideas about motivation are an important part of the reform movement.

The cultural differences between schools and policy makers are substantial. The story of motivation told in the market arena may be confounded in the schools culture. For example, in response to a survey asking whether they would want merit pay for improved test scores, fewer than 17.23 percent of teachers somewhat or strongly favored merit pay, while 60 percent strongly opposed.⁵³ This result suggests an enormous number of potential “winners” in such a system (those who would benefit from higher pay) do not want a valuation system that reflects their achievement on that metric. By contrast, 72 percent of teachers surveyed favored a pay bonus for teaching in “a high-priority situation (e.g., in an inner-city school),” a concept often given the unfortunate name “combat pay.”⁵⁴ If we decide to tell a simple economic story of motivation about these numbers (that working conditions are being traded for wages), we miss a potential insight into the intrinsic motivations of educators and the alternative (non-market) values in the schools culture.

If politicians and some policymakers like testing because it fits a market-based story of motivation, parents and some policymakers may lend their political support to testing for a different reason. The common metric of testing appears to be designed in part to let non-educators see a ranking or assessment that is otherwise inscrutable to them because they are not qualified as educators to evaluate school

51. See *supra* notes 28–29 and accompanying text.

52. See Steven Brill, *The Teachers' Unions' Last Stand*, N.Y. TIMES, May 17, 2010, § 6 (Magazine), at MM32, available at http://www.nytimes.com/2010/05/23/magazine/23Race-t.html?pagewanted=Id_rl.

53. Dan Goldhaber, Michael DeArmond & Scott DeBurgomaster, *Teacher Attitudes About Compensation Reform: Implications for Reform Implementation* 9–10 (Ctr. on Reinventing Pub. Educ., Working Paper No. 20, 2007), available at http://www.crpe.org/cs/crpe/download/csr_files/wp_sfrp20_goldhaber_aug07.pdf.

54. *Id.*

quality. It is a *translation* device between school and non-school cultures. Just as an economist might admit that rational acquisitiveness does not represent the truth of social life⁵⁵ but rather is a theoretical system that by simplification allows prediction,⁵⁶ standards proponents know that education is not summed up by test results but that the simplification of education through test results permits important prediction and discussion by non-educators. Commensurability—the possibility of comparison—is the virtue sought by those without enough expertise to make qualitative judgments along non-standard metrics. So policymakers are seeking simplicity to bridge the gap between themselves and educators. If they seek transformation, it is transformation up the scale of the metric. The insights from the commodification and behavioral economics literatures might suggest that a metric of measurement and monitoring will have unintended transformative effects other than movement up that scale.

Some policymakers see the problem in schools as one of *compliance* with expectations and are strategic in pressuring schools with the threat of sanctions.⁵⁷ School personalities tend to be less strategic and approach mandates sincerely so long as there is the capacity to achieve them.⁵⁸ But when there is not capacity, either because the institution and its personnel lack necessary skills and information or because the goal is impossible, the response of schools can be counter-productive.⁵⁹ Critics of NCLB in particular and the standards movement more generally point to these negative influences that standards have within schools.⁶⁰

The part of the reform movement that is our concern is in part a battle over the theory of motivation in schools. This is particularly visible in the political arena where schools are debated through the lens of attitudes toward government services and toward unions. But

55. See generally DAN ARIELY, PREDICTABLY IRRATIONAL: THE HIDDEN FORCES THAT SHAPE OUR DECISIONS (2008).

56. See generally JAMES BOYD WHITE, JUSTICE AS TRANSLATION: AN ESSAY IN CULTURAL AND LEGAL CRITICISM 53–55 (1990).

57. Elmore, *supra* note 49, at 230–31.

58. *Id.*

59. *Id.*

60. See *supra* notes 27–29 and accompanying text.

the second strand in the testing movement, the one that translates something about educational quality to non-education experts such as parents, provides political support for the education reform movement and may explain its success. It is understandable that a single metric of comparison is attractive to people like parents, with great reason to care about educational quality but without the expertise to know what high quality means. The two reasons for a common metric—translation and motivation—converge in the testing regime.

IV. HOW TESTS CHANGE SCHOOLS: WHAT IS “TEACHING TO THE TEST”?

The range of criticisms about school responses to the standards movement may be summed up in the frequent complaint that schools are now “teaching to the test.”⁶¹ This is an interesting complaint, because the creation of standards and assessments (tests) to check achievement of those standards is in a sense a direct call to teach to the test. Schools are expected to focus their efforts on teaching students the content and skills mandated by the standards movement, with the knowledge that success in teaching those skills is determined by test performance. Teaching the materials for the test is the goal. Therefore, to evaluate this criticism of testing, we need a fuller picture of its purported harms.

Certainly schools divert some energy to teaching test-taking skills that may not be content-oriented, such as how to eliminate options when evaluating a multiple choice question. Almost every public school now does some of this, and this is relatively easy to label wasted time.⁶² But when schools are not teaching test-taking skills, the accountability strand of the testing movement seeks to have teachers teach to the test, at least for the substance. In theory, as long as the test is “good,” meaning an accurate measure of whether

61. See, e.g., *How Standardized Testing Damages Education*, FAIRTEST (Aug. 20, 2007), <http://www.fairtest.org/facts/howharm.htm>; *Teaching to the Test*, BUS.-MANAGED DEMOCRACY, <http://herinst.org/BusinessManagedDemocracy/education/curricula/teachtest.html> (last visited Jan. 21, 2011).

62. *Teaching to the Test*, *supra* note 61.

students learned what we think they needed to learn, testing advocates can argue that there is little harm in teaching to the test.

A. Narrowing the Curriculum

Yet there is solid evidence that the curriculum in schools across the country is narrowing to align with whatever content is being tested and that many schools, children, and parents are unhappy about this development.⁶³ Reports of schools cutting recess, art, and music are routine, and even cuts to social studies are too common.⁶⁴ One can immediately see the conflict in values between those who sought a narrowing of focus to the core reading and math skills and those who see the goals of education more broadly.

Reports of narrowed curricula are widespread. From books with provocative titles like *What Happened to Recess and Why Are Our Children Struggling in Kindergarten?*⁶⁵ to government statistics showing an increase in time spent on language arts and math and a decrease in time spent on science and social studies,⁶⁶ the consensus is that schools across the country have adapted their curricula to focus on subjects that are tested by reducing the time spent on subjects that are not a part of the testing program, such as social studies, and ones that are not susceptible to standardized testing at all, such as music, art, and physical education.⁶⁷ Newspapers report on

63. See Wayne Au, *High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis*, 36 EDUC. RESEARCHER 258, 259 (2007), available at http://www.aera.net/uploadedFiles/Publications/Journals/Educational_Researcher/3605/07EDR07_258-267.pdf (study finding curriculum narrowing as a result of high-stakes testing); Brian M. Stecher, *Consequences of Large-Scale, High-Stakes Testing on School and Classroom Practice*, in MAKING SENSE OF TEST-BASED ACCOUNTABILITY IN EDUCATION 79, 91–94 (Laura S. Hamilton, Brian M. Stecher & Stephn P. Klein eds., 2002), available at http://www.rand.org/pubs/monograph_reports/MR1554/MR1554.ch4.pdf (describing phenomenon of “Negative Curriculum Reallocation”).

64. Alfie Kohn, Editorial, *Emphasis on Testing Leads to Sacrifices in Other Areas*, USA TODAY (Aug. 21, 2001), <http://www.usatoday.com/news/opinion/2001-08-22-ncguest1.htm>.

65. SUSAN OHANIAN, WHAT HAPPENED TO RECESS AND WHY ARE OUR CHILDREN STRUGGLING IN KINDERGARTEN? (2002) (an anti-testing manifesto).

66. Beth A. Morton & Ben Dalton, *Changes in Instructional Hours in Four Subjects by Public School Teachers of Grades 1 Through 4*, NAT’L CTR. FOR EDUC. STATS., 4 (May 2002), <http://nces.ed.gov/pubs2007/2007305.pdf>.

67. Claus von Zastrow with Helen Jane, *Academic Atrophy: The Condition of the Liberal Arts in America’s Public Schools*, COUNCIL FOR BASIC EDUC., 7 (Mar. 2004), <http://www.menc>.

the trend⁶⁸ as educators wring their hands over the trade-offs they are pressed to make.

Here we find an analogy to the commodification literature. One set of values are measurable, are measured, and can be made commensurable, and another set of values are not or cannot be placed on a metric. Rather than simply describing the world, this phenomenon places pressure to re-design the world so that we place our energies behind only what is measured. Here the need to make items commensurable leads to a worse result than a simple failure to describe the *character* of the good in question (education) by placing that good on a common metric. It actually transforms the character of the item. The description is self-fulfilling: education becomes the thing we have tools to measure about education.

In the context of education, it is difficult to argue that this commensurability has *not* corrupted the character of the item, because what schools do has changed since we began forcing schools to measure and compare along a common metric. We have diminished social skills development, character and self-discovery, research skills, physical education, creative endeavors from art and music to theater and even creative writing, science instruction, and the understanding of diverse social organizations conveyed in social studies.

Is this a loss? Good test performance does not appear to generalize outside of the tests themselves to these areas⁶⁹—teaching good test

org/documents/legislative/AcademicAtrophy.pdf (finding increased instruction in reading, math, and science and decreased instruction in the arts, especially in high minority districts); Jennifer McMurrer, *Choices, Changes, and Challenges: Curriculum and Instruction in the NCLB Era*, CTR. ON EDUC. POL'Y, 1 (Dec. 2007), <http://www.cep-dc.org/index.cfm?fuseaction=document.showDocumentByID&nodeID=1&DocumentID=212> (reporting an increase in reading and math instruction and a decrease in other areas). While physical fitness is periodically subject to assessment under national or statewide programs, these are in the nature of health screenings, because fitness itself is not the direct learning objective of physical education but rather the acquisition of skills and knowledge toward the development of active and healthy lifestyles. See, e.g., CAL. ST. BD. OF EDUC., *Physical Education Model Content Standards for California Public Schools: Kindergarten through Grade Twelve*, at vi (2006), available at <http://www.cde.ca.gov/be/st/ss/documents/pestandards.pdf>.

68. See, e.g., Sam Dillon, *Schools Cut Back Subjects to Push Reading and Math*, N.Y. TIMES, Mar. 26, 2006, § 1, at 21.

69. Daniel M. Koretz et al., *The Effects of High-Stakes Testing on Achievement: Preliminary Findings About Generalization Across Tests*, EDUC. RESOURCES INFO. CTR., 20–21 (Apr. 5, 1991), http://www.eric.ed.gov:80/ERICDocs/data/ericdocs2sql/content_storage_01/

performance does not lead to the successful teaching of many valued and useful subjects and skills. The character of education has been transformed by the common metric. The goals of education are now contained by the limits of our ability to measure, rank, and compare, not by the limits of our values for education.⁷⁰

B. “Testing Gains, Not Learning Gains”⁷¹

A different harm from teaching to the test arises because the tests, to be administrable across a large population, must be simplified for the sake of uniformity. Any good educator needs to assess regularly what her students are learning. Those who object to the education reform movement still acknowledge the importance of some assessment to understand what gains students are making. But the current assessments are a tail that wags the dog of educational decision making, because the consequences of disappointing outcomes are so severe and because the metric is universal regardless of school structure, goals, philosophy, resources, and population. The accountability provisions are having their intended effect: they are getting schools to drop everything and focus on a few uniform measures. The tests are not helpful servants of a school curriculum designed to accomplish many goals. Instead, they become the goal. This places greater confidence in the test validity of any one instrument than even test-makers and proponents would claim.

If instead schools were using multiple sources of evidence of learning, each appropriate to one hoped for gain, they might produce a more sophisticated sense of whether they are making gains along all the measures that are hoped for. Prior to the standards movement, schools traditionally used many forms of assessment in that way, by looking at portfolio work, verbal assessment, or instruments intended

0000019b/80/23/82/30.pdf (concluding that performance on tests does not generalize more broadly and that teachers focus on content specific to test).

70. See David L. Berliner, *MCLB (Much Curriculum Left Behind): A U.S. Calamity in the Making*, 73 EDUC. F. 284, 294–95 (2009); Charles Rooney with Bob Schaeffer, *Test Scores Do Not Equal Merit: Enhancing Equity & Excellence in College Admissions by Deemphasizing SAT and ACT Results*, FAIRTEST (Sept. 1998), <http://www.fairtest.org/files/optrept.pdf>.

71. JONATHAN KOZOL, LETTERS TO A YOUNG TEACHER 120 (2007); see also Koretz et al., *supra* note 69, at 1.

to measure very specific and localized learning gains such as content knowledge associated with a science project. By combining a range of assessments, schools could evaluate whether they were achieving gains in student learning defined a number of different ways and tied directly to curriculum units.⁷² But in today's high stakes testing world, success has one measure: the statewide annual uniform test. Like an economic market that allows us to compare numbered prices as full representatives of value, numbered test scores are on a path to becoming the full representatives of value in education.

C. Corrupted Schools

In the commodification literature, markets corrupt values by changing the character of goods and extinguishing non-market valuations.⁷³ In the behavioral economics literature, external rewards and metrics change human motivation by diminishing intrinsic motivation. But in education, a most old-fashioned variety of corruption has shown up with disturbing regularity as individual teachers, and sometimes entire schools, cheat or assist their students in cheating on standardized tests.⁷⁴ This gives a different and more literal spin to the idea of corruption in the commodification and behavioral economic literatures. Teacher cheating reflects the intensity of the disconnection between school personnel and the standards advocates.

An increasing number of schools have been judged "failing" under the new standards regime, and climbing up the metric of test scores by any means attracts some school personnel to a deep betrayal of educational values. Imagine the cynicism toward education experienced by a student whose teacher assists her in cheating on a test, and you can see the three versions of corruption converge.

72. Surely with multiple assessments of different varieties it is more difficult to hold schools accountable for uniform standards. A trade-off has occurred among goods and values rather than a climb up a common vision of quality.

73. See generally RADIN, *supra* note 7; RETHINKING COMMODIFICATION, *supra* note 7.

74. Kris Axtman, *When Tests' Cheaters Are the Teachers*, CHRISTIAN SCI. MONITOR (Jan. 11, 2005), <http://www.csmonitor.com/2005/0111/p01s03-ussc.html>; Shaila Dewan, *Georgia Schools Inquiry Finds Signs of Cheating*, N.Y. TIMES, Feb. 12, 2010, at A14.

But corruption takes a less literal form in schools as well, one closer to the description Michael Sandel elaborates—degradation of goods by diminishment of non-market modes of valuation.⁷⁵ We have seen the way curriculum content is narrowed to meet the uniform metrics, such that science and social studies take a back seat to math and language arts.⁷⁶ But the *skills* portion of education is similarly narrowed to that which can be assessed on a statewide test. Many tests have an open response component, but because every student in every state must take the test, a large portion is multiple choice. This makes it difficult to demonstrate thinking skills, much less creativity. Testing proponents may view this as just a difficulty of instrument design. But in the case of thinking and creativity skills, testing might look for what's inherently uniform in attributes valued in part because they are individual and incomparable in some ways. How can educators genuinely value what's incommensurable in these skills in a regime that defines value in ranked comparison to others?

Consider the value students derive from the experience of performing in a play, and then ask what of that value we can measure in a way that allows us to conclude that one set of students attained more of “it” than another. Public speaking confidence? Appreciation of others by embodying a role? Teamwork? Tackling uncomfortable tasks? Most educators, and probably many non-educators, believe that these skills and experiences have educational value, but that belief itself is threatened by the inability to make uniform assessments in our current standardizing system. It is not simply that the values at stake are not susceptible to measurement, though that may be true. It is that uniformity is not exactly a desired outcome—if the skills at stake could be measured and compared, we would be assuming more is better along a single line. That is how the argument from incommensurability meets the argument from corruption: incommensurability of value should not mean there is no value, but the practice of commensurability/comparison corrupts education such that we no longer value what we cannot measure and also what we cannot compare. This is what Kozol calls the “demeaning anti-human

75. Sandel, *supra* note 16, at 94.

76. See *supra* Part IV.A.

view of childhood⁷⁷ that utilitarianism in education implies: it's an anti-human system if it fails to give meaningful energy to the development of attributes whose value is wrapped up in their incomparability.

Student discovery of content instead of being handed content similarly develops important life skills. Discovery is what happens when the learning process includes time for children and adults alike to experiment and find information. A rushed curriculum, on the other hand, asks us to more efficiently deliver information—to spoon feed what ought to be discovered. This may improve test scores without improving retention. More importantly, children fail to learn the skill of discovery. Many educators have pointed out that attainment of the knowledge to achieve on standardized tests in public K-12 schooling today has not been validated to attainment of anything in particular that a young adult might need.⁷⁸ The ability to find or discover information seems more adapted to a changing and unpredictable future than the information itself. This is why anti-test leader Alfie Kohn argues that improved test scores are ruining schools.⁷⁹

Examples of educational values that are in a similar, untestable zone are easy to find. Schools cannot test team-building behavior, problem-solving, attitude, adaptability, motivation, curiosity, situation sense, flexibility, leadership, ethics, open-mindedness, patience, compromise, conflict-resolution, or self-expression. But many agree that children need to develop these capacities to be happy, good, and successful individuals, citizens, and workers in their adult lives.⁸⁰ That we cannot place those values onto the central metric for measurement diminishes their importance to educational culture today, and the “market” becomes flat, the character of the item corrupted, and genuine value is lost. If the question, “What does

77. Kozol, *supra* note 1, at 22.

78. Larry Cuban, *Why Bad Reforms Won't Give Us Good Schools*, AM. PROSPECT (Jan. 1, 2001), http://www.prospect.org/cs/articles?article=why_bad_reforms_wont_give_us_good_schools; Koretz et al., *supra* note 69, at 1 (concluding that performance on tests does not generalize more broadly and that teachers focus on content specific to test).

79. Kohn, *supra* note 27.

80. Jonathan Kozol, *The Details of Life*, THE NATION, May 22, 2000, at 15, *available at* <http://www.thenation.com/article/details-life>; Berliner, *supra* note 70, at 290–91.

a child need to become a fulfilled adult?” were the same as the question, “What makes a good education?,” education would look different from today’s standards and benchmarks-based institution. This might explain in part the suddenly common newspaper or magazine article that asks whether education is still relevant to success in life.⁸¹

These observations beg a number of questions. What are the proper goals for public education, which values among those goals are being compromised, and why? Additionally, what do markets and testing have in common that make them such powerful modes of valuation such that they seem to threaten other modes of valuation so easily? The next two sections take up these concerns in turn.

V. THE HISTORY OF PLURALISM IN EDUCATION: DOES QUALITY EDUCATION MEAN UNIFORM EDUCATION?

Throughout the history of public education in the United States, there has not been consensus around a single goal for the institution. Rather, a variety of goals and purposes have been held out as justifications for public education. It was Jefferson’s ideal that public schools were necessary to prepare citizens to effectively participate in the new democratic form of government by creating literacy, deliberation, and reasoning skills that would inform voting and prepare some in the newly classless society to run for office.⁸²

Other justifications that have fueled the institutionalization of state-funded education in the United States have included resolving cultural conflict that arises in a diverse society of immigrants through shared experience across family and ethnic backgrounds. Education was promoted by earlier arrivals as a way to create a common American identity with later arrivals. Public education is also supposed to prepare workers to serve the economy and prepare workers to improve their own standing in the idealized classless

81. See, e.g., Rebecca Mead, *Learning by Degrees*, NEW YORKER, June 7, 2010, at 21.

82. CARL F. KAESTLE, PILLARS OF THE REPUBLIC: COMMON SCHOOLS AND AMERICAN SOCIETY, 1780–1860, at 6 (1983).

society. It has been charged with promoting social order, industrious habits, and intelligent citizenship.⁸³

At many points in American history, education has been offered as the basis for a theoretically classless society where individuals can compete for wealth based on skills and abilities cultivated on an equal playing field, and the equality ideals embedded in the American form of government have been thought to depend on the availability of free, adequate public education for all children.⁸⁴

Important to this Essay, no one single justification and purpose for public education has triumphed in either the education or the political discourse. Rather, multiple purposes have co-existed as schools grew out of local communities with only loose oversight at the state level and almost no guidance at the federal level. Pluralism and community have been at the heart of public education, and localism managed particularized community values and goals, preventing the dominance of a singular vision for education.⁸⁵

This is not a claim that any ideal has been achieved in the public education system now or in the past. Indeed, the education system has notoriously failed low-income rural and urban children, immigrants, and children of color in particular.⁸⁶ But requiring identical output from schools with vastly different resources, challenges, and cultures does not improve on that problem. Rather, the problems of inequality in school resources and challenges are minimized by an expectation of equal output on a testing metric.

Some critics of testing focus on unintended consequences such as cheating and compromised curricula.⁸⁷ But other critiques of standardized testing reject the concept of measurement more thoroughly.⁸⁸ For example, a recent article criticizing the standards

83. *Id.* at 7–8, 75–83.

84. Of course this is not a claim that public education has produced a classless society and social mobility. My commentator Kieran Healy focuses in part on the role that education plays in the social justification for unequal opportunities and on education as an authoritative system and relates testing to that function. I have no particular objection to his comments in this regard.

85. KAESTLE, *supra* note 82, *passim*.

86. Sarah Deschenes, Larry Cuban & David Tyack, *Mismatch: Historical Perspectives on Schools and Students Who Don't Fit Them*, 103 *TCHRS. L. REC.* 525, 530–31 (2001).

87. See Axtman, *supra* note 74; Berliner, *supra* note 70, at 294–95.

88. See Kohn, *supra* note 27.

movement is poignantly titled *A Child Is Not a Test Score*,⁸⁹ using the kind of language that has so concerned market commodification critics like Elizabeth Anderson in other contexts when *price* is substituted for *test score*.

These perspectives speak to the lack of consensus on the purposes of public education. The genius, though, of American public education has been its ability to withstand this lack of consensus and to thrive in the notion of pluralism. The uniformity of the standardized test movement threatens this pluralism of purpose with a top-down set of purposes selected from among the many. Public education has always failed its ideals in many respects. But at different times and in different places the multiplicity of ideals has shaped pluralistic educational agendas, failure to achieve success with many of those agendas notwithstanding. The testing trend takes multiple values that have co-existed and reduces them to the one value, which even in its best light can only be expressed as competence in math, reading, and writing, without reference to other necessary skills for a fulfilling life or citizenship. Not only is this a flat choice among the numerous values public education serves, it fails to reach the citizenship values that have long helped to justify the public investment in free education.

VI. COMMON POWER IN THE TWO METRICS: MONEY AND TEST SCORES

Testing does not map perfectly onto literal markets, and we could surely spend time drawing out differences. But the similarities can be uncanny. My question is what the lessons of the comparison might be. Several occur as possibilities. If alternative values are just that—values—why can't they stand up to market norms or testing norms? Why do markets (tests) extinguish plural conceptions of personhood (education)? I offer two possible answers. The first asks whether the power lies in numerosity itself: rank orderings cascade inevitably toward flattening values. The second asks whether the standards movement in schools is linked more literally to economic markets.

89. Neill, *supra* note 48, at 28.

A. Numerosity

Does the power to deflate and flatten plural values stem from the ability to compare along a single metric? We might call a person adventurous and another loyal, recognize that these are different values, struggle to compare them, but in the end recognize that they will not be pressed into an agreeable ordering. But when numeric values are at play, the ordering comes naturally. Perhaps the overpowering ethics of markets and testing arise from numerosity. This point is reflected somewhat in the opening quotation by Jonathan Kozol pointing to the de-humanizing utilitarianism reflected in tests. Uniformity for comparison purposes binds together money/markets and scores/schools.⁹⁰ Incommensurability relies on metrics that cannot standardize, and if incommensurability is necessary to the maintenance of plural values, then testing and markets cannot maintain plural values.

The literature on crowding-out intrinsic motivations with extrinsic ones suggests that context matters. Sometimes extrinsic pay is highly effective. Perhaps there are not robust intrinsic motivations to be displaced, for example, when pay is used to motivate behavior in certain contexts. Sometimes one metric of value is enough. In other contexts, it matters greatly. Although there does not yet seem to be extensive work in the behavioral economics literature on the impact of standardized testing on intrinsic motivations, related and relevant study is sobering. There have been several experiments with paying students for higher test scores. The findings vary somewhat but generally point to either no or small increases in test scores, followed by decline to levels worse than before the incentives were introduced when the extrinsic pay is removed.⁹¹ These studies do not demonstrate that testing *without* pay leads to crowding-out, but they do suggest that there are intrinsic motivations in place in the educational setting capable of being crowded-out such that the theory of crowding out of motivations disrupts a standard carrot-and-stick

90. See generally Michael Winerip, *Our Towns; Never Mind the Inventive Curriculum. One Test Fits All*, N.Y. TIMES, (Nov. 18, 2001), <http://www.nytimes.com/2001/11/18/nyregion/our-towns-never-mind-the-inventive-curriculum-one-test-fits-all.html?n=Top/Reference/Times%20Topics/Organizations/B/Board%20of%20Regents&pagewanted=print>.

91. See generally Bettinger, *supra* note 15; Garbarino & Slonim, *supra* note 15.

model of motivation in schools. More than one set of values drive schools.

If the schools context is full of incommensurable motivating values, if education has robust “plural meanings” in Radin’s terminology, the rankings associated with a single metric may do more damage to education than they do to typical market commodities. Numerosity may flatten out more incommensurable values when there are more to flatten. If education is such a good, than the single numeric scores from tests can behave like prices in the area of other goods with robust plural meanings, such as reproductive labor, friendship, housing, or sex.⁹² These goods with plural meanings suffer when prices purport to exhaust their value.⁹³ If the harm of pricing derives in large part from the numeric ranking itself, testing should do the same to education.

B. Creation of Workers

A different possibility is that markets as we understand them conventionally are embedded in the standards movement. That is, the uniformity of purpose occurs because it relates to a particular market-based idea about the purpose of public education, which is workforce readiness.⁹⁴ In a non-uniform system of schooling, multiple conceptions of the purposes of public education have co-existed, sometimes within one school, and sometimes across different schools and school systems. That is, the original explanation for U.S. public schooling, the Jeffersonian conception that self-government depended on an educated electorate, is a citizenship-based rationale for public schools. Many schools in the United States have embraced that purpose, with preparation for citizenship an explicit goal that also pervades the curriculum. But schools are also, and have long been, purposed with improving the economic prospects of individuals, thus, in theory, de-classing the society. This posits a market-based reason for public education that serves the students;

92. RADIN, *supra* note 7, at 105–07.

93. *Id.*

94. *See, e.g.*, Evan Osborne, *Education Reform as Economic Reform*, 25 CATO J. 297, 297 (2005). The same emphasis can be found in the 1983 *A Nation at Risk* report. *See* A NATION AT RISK, *supra* note 32.

education gives individuals the opportunity to achieve upward class mobility in the economy. But the nebulous concept of “the economy” itself is also served by a prepared workforce, and fears of international competition for the best-educated workforce, while long present, seem to have increased in pitch. Perhaps the similarity between the commodification literature and the standards movement in public education exposes the increasingly tight nexus between markets and public education.

Consider the school choice movement, which is tightly linked to the broader standards movement. The school choice movement seeks to harness explicitly market-based mechanisms of comparison shopping by parents in an effort to improve the quality of schools by making them work to retain their students.⁹⁵ Parents are expected to vote with their feet by comparing test scores among schools. That school choice has not worked particularly well—parents want to preserve their school communities rather than shop among them—suggests that parents themselves want schools to serve a broader purpose for their child than the production of improved test scores.⁹⁶ But the choice movement suggests a link in the minds of education reformers between uniform tests and markets.

In response to this explanation—that markets and standardized testing may have a closer link than numerosity alone—we might ask how the elimination of critical thinking, team-building, or self-awareness, for example, could possibly serve the economy. I have no ready answer to this question, and would note that the paradox may explain some of the recent hand-wringing over whether education still has value.⁹⁷ But perhaps the short-hand of comparing workers that only homogeneity of metrics can deliver may explain why some skills are worth giving up in exchange for the ability to order potential workers.

95. Lynn Bosetti, *Determinants of School Choice: Understanding How Parents Choose Elementary Schools in Alberta*, 19 J. EDUC. POL'Y 387, 388 (2004).

96. *Id.*; John Coldron & Pam Boulton, *'Happiness' as a Criterion of Parents' Choice of School*, 6 J. EDUC. POL'Y, 169, 169 (1991); PAUL TESKE, JODY FITZPATRICK & GABRIEL KAPLAN, *OPENING DOORS: HOW LOW-INCOME PARENTS SEARCH FOR THE RIGHT SCHOOL* 4 (2007), available at http://www.crpe.org/cs/crpe/download/csr_files/pub_crpe_open_jan07.pdf.

97. See, e.g., Mead, *supra* note 81.

VII. CONCLUSION

The comparison between the testing movement and the commodification literature is not perfect, because commodification is used to describe a process of putting something into the market for exchange at a price. But they have in common being counted, measured, and compared along a single metric and being subject to the related process of flattening or thinning out whatever values cannot be reconciled with the numbered and ranked system. From the comparison we draw cautionary notes for the testing movement, areas for further research about motivation in behavioral science, and translation of a philosophical debate into practical policy.