

A Critique of Student Evaluations AKA Student Opinion Surveys **New Edition by Doug Mann, February 2016 – Includes “A Modest Proposal”**

A. The Official Story: The premise of student evaluations is that students will use them to provide fair and objective feedback of professors’ teaching skills and class organization, and that they are thus an objective quantitative measure of the relative merits of various professors at a given institution. I submit that this premise is patently false both because **the logic behind it makes no sense**, and because scholarly studies have shown time and time again that **evaluations measure something else other than student learning or teaching skills**, i.e. the popularity or looks of the professor, the satisfaction or lack thereof of students with their grades, the workload in a course, the time and size of the class, etc. I will show that this premise is false through logic, facts and by reviewing a wide range of scholarly studies on student evaluations.

B. The Facts of the Case

Here are the undisputed facts concerning student evaluations:

- They are done near the end of the class, but before final essays are handed back or final exams graded, and thus are based on an incomplete picture of the grade a student will get (in my own classes, students have about 20-30% of their grade on Evaluation Day AKA E-Day).
- They are done **quickly**, sometimes in less than 5 minutes, thus indicating a lack of profound thought on the part of the evaluator.
- They are **anonymous**, thus allowing students to avoid any personal responsibility for what they say, echoing on-line comments sections.
- Students getting **low grades** will frequently attack the value of a course (we can tell this from comments that explicitly refer to the difficulty of tests etc.).
- The comments frequently violate the **law of non-contradiction**, e.g. “this was my best class!” vs. “this was my worst class!”, or “the quizzes were too hard!” vs. “the quizzes were a good way of keeping students up to date!”, etc. This leads one to conclude that they’re evaluating something other than the professor’s teaching skills. Further, they rarely present actionable critiques, and they rarely (if ever) help professors improve on courses by focusing on specific elements of the course’s contents that could be expanded, refined, or cut back (e.g. the order of topics in the course).
- They are done by people who have never taught a course, marked an essay or exam, or designed an outline. Students are by definition **less expert** in the course material than the professor teaching it, usually considerably so. If this weren’t true, then they would be teaching the class themselves. In that sense, it’s like mechanical engineers judging an ice-skating competition, or avant-garde artists marking a physics exam.

- Undergraduate students are generally 17-21 years old, most of them a year or two out of high school. To get into UWO they must have had an A average in high school, so they see themselves as “excellent” students (despite the decline of standards in our high schools). It’s possible that no one has ever told them they’re only “average”, and certainly not “poor”, students. So a 15-30% grade drop represents a **huge ego deflation** for some. This will depress or anger them.
- The only direct, legal, and anonymous means an angry student has of “getting back” at a professor without suffering any consequences is through evaluations. Admittedly, they can use web sites like www.RateMyProfessors.com to vent anger or express taboo sexual desires with a red hot chili pepper icon.
- A quick look at Western’s data base will show that evaluations numbers *for the same professors* can vary by as much as 4 points (out of 7), with a 1.5-point variance being quite common. Though some of this is explainable by varying classes sizes, such **striking variances** suggests either (a) widespread dissociative identity disorder within the UWO professorate (and thus a mental health crisis of epic proportions), or (b) a high degree of subjectivity in students’ ratings of their professors.
- It’s a given in advertising that most products are sold by varying combinations of **sex, beauty, magic, youthful images, funny stories**, and in some cases **price**. Not surprisingly, the literature finds that the same techniques influence students on SETs, with charisma replacing magic, and high marks replacing low prices. Social psychology has proven, as one can see from any introductory textbook, that beautiful young people are rated more highly on unrelated virtues like intelligence than their plainer comrades. Further, charm and wit beat facts and arguments 9 times out of 10. As Mark Oppenheimer reports in the September 21, 2008 *NY Times*, in 1973 Donald Naftulin did an apocryphal experiment where he hired an actor he dubbed “Dr. Myron L. Fox” to deliver a lecture on mathematical game theory. He did so with charm and wit, but both his lecture and the following discussion were nonsense full of “double talk, non sequiturs, neologisms and contradictory statements.” The fraudulent Dr. Fox got excellent reviews from both graduate students and faculty after all three of his lectures.

C. The Student Weltanschauung

1. *Grades Rule*: Students will do scholarly work **only for grades**. A simple challenge to those who disagree is to perform the following experiment: assign a 5-page essay to a class without giving it any grade value, but with a firm deadline. 99.9% of students at UWO will not write this essay. The vast majority of students apply the gauge of **instrumental reason** to their scholarly work: they do it if and only if they’re being “paid” for it with marks. As Love and Kotchen and other economists have argued, students see higher grades as translating into higher *future wages*, so markers giving them low grades are, in effect, taking money out of their pockets. This “consumer model” is at odds with the traditional notion of institutions of higher education as places

of learning where high grades are measures of excellence that most students strive for, driven by some combination of competition, work ethic and personal integrity. So grades and evaluations are interchangeable forms of **currency**.

2. *Marginal Utilities*: Further, the weaker students in a class will not even do work to improve their current low grades if it's defined as "optional": they will not trade work for what they see as a *marginal utility*. Proof: for about five years I've included in all my classes optional "mini-reports" that I post on Web CT/Owl about once a month that students can write to replace either a low quiz grade or a participation grade. They are 3-4 page papers on lively subjects based on class readings and a bit of Internet-based research designed to be writable in two evenings. I sometimes format them as diaries, newspaper articles or screenplays. Despite the fact that up to 80% of students in my classes would clearly benefit from such assignments, in my MIT classes at UWO only about 10-20% actually do them, usually the students who get a B+ or A in any case. Up to last year, the highest number I ever had was in a Sociology class of about 75 students in 2013 where there was a 27% response rate, in part because two-thirds of the class faced losing their full 10% participation grade because of truancy or silence. Despite this being clear, 8 students still got a ZERO on participation, refusing to do a report which they had five opportunities to write over a three-month period.

3. *The Phenomenology of Leisure*: If one simply uses one's eyes in public areas such as cafeterias and buses where students congregate, one will see that *very few*, if any, are reading anything of substance. Instead, most are either talking to friends, texting, listening to music, or using a social networking web site. It's not unusual to take a bus to or from campus with 20-40 students aboard where *no one* is reading anything at all, yet half are using electronic devices. Exam periods and libraries are only partial exceptions: a quick tour of Weldon Library on most days will show you that at least half of the students there are engaging in electronics-based leisure activities. So once class is over, most students flee scholarly labours like medieval peasants fleeing the plague. Added to this fact that only about 50-75% of students in a typical class even *buy* the course texts (which one can tell from online Bookstore numbers), though a few may download these books, or use library copies. And despite clear guidelines asking them to do so, it's difficult to get weaker students to read even three journal articles or book chapters in preparation for writing their final papers. Many essay writers use only web sites.

4. *Changing Attitudes towards Cheating*: Though we can wonder about the validity of media reports of an epidemic of cheating in higher education, it would seem that student attitudes toward it have changed of late. As reported in the March 7, 2014 UWO *Gazette* article "Profitable Plagiarism", ghost writers help students plagiarize papers. "Sam", one such ghost, defends his practice by noting that plagiarists are just "looking for more time to focus on things they saw as more important". A cheating computer science student, "Kirk", claims that Western is committing fraud by forcing him to "take random credits because I didn't meet some arbitrary number", ergo cheating is acceptable. Cheaters argue that they have "no time" to complete all their work, and they shouldn't be expected to do much work in courses unrelated to their future careers. In short, the ethical dimension of cheating has, for at least of minority of students, faded to black. This speaks to the consumerist, instrumental, "win at any costs" attitude of disengaged students today.

5. *Preliminary Conclusion*: From the previous four points above it is easy to conclude that grades are the prime motivators of student work, that most students are very conscious of them (both in the absolute and in relation to class averages), and that they will resent professors who either give them low grades, or who demand more work for the same grade than professors in comparable classes. For many, rigorous professors are literally **taking money** out of their pockets.

D. Scholarly Studies

1. AMINISTRATIVE DENIAL: Numerous studies on the relation between student evaluations (SETs) and grades are easily accessible via a simple Google search. About 80% support the notion that higher grades lead to higher evaluations, and vice versa, which lowers academic standards. Those which don't fall into two categories. First are **short denials** of any effect of grades on SETs seen in blogs and web pages written by officials in "student learning centers" etc. who speak with the official voice of the university where they work. These are bogus for two reasons: their authors' jobs depend on this evaluative structure being valid (see Karl Marx on ideology), and these blogs never cite quantitative data supporting the idea that high SETs = effective teaching. They are thus just wishful thinking.

2. TEACHER EFFECTIVENESS: A second small group of studies admit that there is **some small correlation** between high grades and high SET scores, but chalk this up to students' appreciation of the better "learning experience" in classes with higher SET numbers, which leads them to work harder. This is referred to in the literature as "**teacher effectiveness theory**". However, these studies fail to prove a causal link between teacher effectiveness and high grades: they either take it on faith, make the circular argument that the higher grades *prove* teacher effectiveness, or make their case by *assuming* that students answer relevant specific questions on the same evaluations honestly, including ones like "was the teacher organized?" or "did the professor provide a stimulating learning environment?", then using these individual numbers to make the general case about SETs being a good measure of teaching skills, instead of looking at how students fare in future classes, or doing in-depth individual interviews.

Yet if one looks over a set of evaluations for a given professor diachronically, one will notice that these individual ratings rise and fall with the overall *gestalt* the students have of a given class, which is itself based partly on grades, instead of some of the more "technical" questions having a **constant rating** from one class to another, which one would expect if SETs were an objective measure of teacher effectiveness. For instance, I have taught classes where I hand back 100% of assignments or quizzes in the next class, i.e. as quickly as humanly possible (short of building a time machine). Yet I get varying numbers on the question "does the professor hand back work promptly?", and have never gotten a perfect 7/7 on this question (though it was *physically impossible* to hand back their work quicker). So these individual questions are just as subjectively answered as the meta-questions about a given professor's worth, and provide no real evidence for the notion that high-evaluation classes objectively reflect teacher effectiveness. In conclusion, these "teacher effectiveness" studies fail to empirically link high SET scores to some objective external proof of this effectiveness.

3. DO YOU WANT TO TANGO?: **Valen E. Johnson** finds in his statistical study “Teacher Course Evaluations and Student Grades: An Academic Tango” from *Chance* vol. 15 (2002), that there was a direct correlation between high grades and high scores on nine specific SET questions such as “was the instructor enthusiastic?” He found a causal effect of grades on SETs, concluding that the analysis of his data “suggests that there is a direct effect of grading policy on SETs beyond that which can be explained by such extraneous factors” such as *actual* teacher effectiveness (16). Johnson finishes by noting that the “ultimate consequence” of grade manipulations “is the degradation of the quality of education in the United States”.

4. NO LEARNING REQUIRED: In an important study by **Scott Carrell** and **James West** in the *Journal of Political Economy* (118.3, 2010) entitled “Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors”, the authors find that although professors who “teach to the test” may get higher grades and higher evaluations out of their students within their own course, students who take courses from more rigorous professors do better in future courses.¹ They used a massive data set of 10,534 students from the 2000-2007 period at the US Air Force Academy. They also had the advantage of random assignments of students to classes and standardized curricula and exams. Their study is one of the few to focus on learning outcomes for students in subsequent classes, finding that high SET numbers do not correlate with high grades in later courses, as teaching effectiveness theory claims. It’s worth quoting this study’s conclusions at length so as not to distort them in any way:

We find that less experienced and less qualified professors produce students who perform significantly better in the contemporaneous course being taught, whereas more experienced and highly qualified professors produce students who perform better in the follow-on related curriculum...

One potential explanation for our results is that the less experienced professors may adhere more strictly to the regimented curriculum being tested, whereas the more experienced professors broaden the curriculum and produce students with a deeper understanding of the material. This **deeper understanding** results in better achievement in the follow on courses. Another potential mechanism is that students may learn (good or bad) study habits depending on the manner in which their introductory course is taught. For example, introductory professors who “teach to the test” may induce students to exert less study effort in follow on related courses.

Our results show that student evaluations reward professors who increase achievement in the contemporaneous course being taught, **not those who increase deep learning...** Since many U.S. colleges and universities use student evaluations as a measurement of teaching quality for academic promotion and tenure decisions, this finding draws into question the value and accuracy of this practice. (430)

In short, students in classes with high marks and evaluations experience a short-term gain (a quick “high”), whereas students in more rigorous low-mark classes become better learners over the long term, though this fact is not captured in individual SETs.

¹ <http://www.econ.ucdavis.edu/faculty/scarrell/profqual2.pdf>

5. SKEWED STATISTICS: Statistician **Philip Stark** and teaching consultant **Richard Freishtat**, both from UC Berkeley, deconstruct the value of SETs in their astute 2014 study “An Evaluation of Teaching Evaluations.”² They argue that SETs do not measure teaching effectiveness, while also committing a number of **statistical sins** (my term). First, **averages** of rating scores should not be computed or compared across instructors or departments. These “ordinal category variables” are in fact labels, not numbers, which it makes no sense to average. They tell the joke of three statisticians going hunting: one fires his rifle a yard to the left of a deer, the second a yard to the right. The third exclaims “we got it!” Scatter matters: frequency tables should be used instead.

Second, it’s important to report survey **response rates**. For one thing, people are more motivated to act by anger than satisfaction. They compare an 8am class with a well prepared professor to one where the professor doesn’t provide any notes to one where the professor is very entertaining, giving hints of what’s on the exam. The first class gets low attendance, the second and third ones higher attendance, yet we shouldn’t assume that they’re more effective teachers. Third, we should not **compare** incommensurables. Ratings in classes of different sizes and types with different curricula taught at different times shouldn’t produce constant, reliable results for a given professor. A lab instrument that gave the same reading when its inputs vary considerably is, needless to say, broken, like an outdoor thermometer that reads a steady 20⁰C all winter long.

Fourth, SETs are observational studies that rarely justify inferences about causes. What we need is a **controlled, randomized experiment**: controlled in the sense that the classes being compared need to have the same content, workload and grading standards, and randomized in the sense that the students in each class are a balanced mix of people from a cross-section of different groups (including levels of academic achievement). Yet very few universities have this, since students choose their own classes and professors grade with different standards.³ They give the example of a professor with the reputation of being an “easy A” getting the same lax students from class to class, who reward him or her every time with high evaluations, seemingly proving that he/she is a consistently excellent teacher.

They conclude: “We do not measure teaching effectiveness. We measure what students say, and **pretend it’s the same thing**. We dress up the responses by taking averages to one or two decimal places, and call it a day.” Part of the problem is that students will tend to answer a different question from that being asked, “regardless of their intentions.”

So what do SETs measure, according to Stark and Freihstat? They mention about a dozen studies with conflicting results. These studies argue that they measure **grade expectations, enjoyment scores, visual first impressions of physical attractiveness** (final ratings can be predicted based on a 30-second pre-class silent video of a prof), **gender, ethnicity** or **age** (numerous studies focus on these last three as key factors).

² Stark summarizes their findings in several easy-to-access blogs online. Here’s the complete article: <http://www.stat.berkeley.edu/~stark/Preprints/evaluations14.pdf>

³ They cite Carrell and West as one of the few true cases of such an experiment.

In the end, they suggest that a **complex web of methods** be used to evaluate teaching: student comments, syllabi, lecture notes, websites, software, videos, assignments, exams, teaching statements, surveys of former students, and most importantly, peer observation of lectures.

6. DECREASED EFFORT: In their study “Grades, Course Evaluations, and Academic Incentives” (*Eastern Economic Journal* 36, 2010), **David Love** and **Matthew Kotchen** find that the link between grades and evaluations “can lead to grade inflation, diminished student effort, and a disconnect between institutional expectations and faculty incentives” (151). Further, “lenient grading offers a low-cost means of boosting course evaluations without sacrificing time for research” (151). In their own survey of the literature, they conclude that most researchers find a strong, positive correlation between grades and student rankings, and that grades are the means by which both students and teachers evaluate each other (153). The empirical evidence shows that grade inflation exists, and influences SETs (154). They quote Merrow (2005) to the effect that students and teachers sign a “**mutual non-aggression pact**”, trading decreased expectations for student work for high evaluations (159). In short, an emphasis on SETs causes **grade inflation**, diminished student effort, and **decreased teacher effort**, notably for a professorate dedicated to research (162).

7. THE ECONOMICS OF GRADES: In another extensive study, “The Impact of Relative Grade Expectations on Student Evaluation of Teaching” (*International Review of Economic Education* 6.2, 2007), **Clifford Nowell** concludes from his survey of the literature that there is “overwhelming evidence” that “a student’s evaluation of his or her instructor is related to their expected grade” (43). In his study he tests the hypothesis that there are limits on the ability of professors to buy high evaluations with high grades since it’s their *relative grade* that students care about, not their absolute one. So if everyone’s grades are inflated, this (according to the hypothesis) won’t make the individual student happy. He finds this “**relative reward**” hypothesis to be **false**. In his study of 716 students in all courses offered by a US Economics department in the fall of 2003, he found that students give **higher evaluations** of teachers both when their individual grades are increased in the **absolute**, as well as when their grades increased **alongside all those of their peers** (55). Thus the limits on lowering grading standards to buy high evaluations are more flexible than some have suggested, and the incentive to lower standards “may actually be greater than has been thought in the past” (55).

8. THREE CYNICS IN OHIO: In a 2007 study of 50,000 enrollments at Ohio State in economics courses, **Weinberg, Fleischer and Hashimoto** found that there is no correlation between the evaluations professors get and the learning that is actually taking place. They did so by look at grades in subsequent classes that relied on what students had learnt in the class being evaluated. Their finding echoes Carrell and West.

9. THE BEAUTY MTYH: In the October 15, 2003 issue of *The Chronicle of Higher Education*, Gabriella Montell describes a study by **Daniel Hamermesh** and **Amy Parker** at the University of Texas that “**attractive professors** consistently outscore their less comely colleagues by a significant margin” on SETs. They asked students to rate the beauty of photos of 94 professors, and then compared these ratings to SET scores. The hottest profs scored a full point higher on their evaluations. Montell also interviews a

number of these “hot” professors for anecdotal evidence that good looks equal good evaluations. Kate Antonovics, an attractive 33-year-old economics prof at U California, describes emails from students asking her for dates, for where she shops for her “cute” outfits, and professions such as “I think you are very very hot.” In short, “looks count.”

A 2006 study by Central Michigan U finance and law professor James B. Felton of 7000 faculty ratings on www.RateMyProfessors.com found a high correlation between “hotness” and “quality”: 0.64/1. The relation between a prof being “easy” and having “quality” was almost as high: 0.62. The hottest 99 profs got an average quality rating of 4.43, while the 102 least attractive got an average of 2.20. He suggests a “Halo Effect” radiating from the hottest profs, making them appear more approachable.

In addition, at least half of the other studies I’ve cited here mention physical beauty as a serious source of distortion when using SETs as a measure of teaching effectiveness.

10. EDITORIALS: There have been many **editorials** written on problems with SETs. *The Chronicle of Higher Education* has featured a number of articles on SETs, notably **Richard Vedder’s** June 19, 2010 article “Student Evaluations, Grade Inflation, and Declining Student Effort”. He sees the main danger of evaluation-inspired grade inflation as the fact that students “do less reading, less studying, even less attending class than two generations ago” because “they don’t have to do more... with relatively little work they can get relatively high grades – say a B or even better”. He notes that while in 1961 students studied on average 40 hours per week, by 2003 this had dropped to only 27 hours. Vedder suggests reducing the impact or eliminating entirely SETs, letting websites like www.RateMyProfessors.com take up the slack.

In his September 21, 2008 report to the *New York Times*, **Mark Oppenheimer** describes the rise and fall of one Annemarie Bean, a “funny, enthusiastic” professor who was devoted to her students and passionate about what she studies. He reports that in effect, “her students fired her”, giving her highly polarized evaluations – though many admired her, some loathed her. He also tells the story of the phony lecturer Dr. Myron L. Fox getting rave reviews for spouting charming and funny nonsense in a lecture on mathematics, and of Clark Glymour’s story of how a newly hired professor at Carnegie Mellon University went from getting the lowest to the highest evaluations in his department, thus saving his job, by making sure that most of his students knew they were going to get an A. He cites a 2007 study by Ohio State economist Bruce Weinberg of ten years’ worth of SETs and grades that found that students who did well in upper-year classes were not likely to give their first-year professors high evaluations – “in other words, when you use performance in higher classes as the measurement of learning in previous classes, the correlation between learning and positive evaluations breaks down.” Oppenheimer concludes that although some benefit from their use, many professors see student evaluations as “choking higher education”.

Myron L. Fox himself (an obvious pseudonym) argues on his web site “Student Evaluations are Worthless” (May 7, 2008) that they are academe’s “**dirty little secret**” that “manifests a cynicism about higher education that tragically lies right at the core of educators’ relationship and interaction with students.” He goes on to state that they subvert academic standards by causing grade inflation and dumbing down courses, and

lead to excellent professors being denied tenure, promotion or even jobs “simply because they failed to pander to the increasing desire of students to be entertained or at least to be relieved of the hard work that *genuine* education requires.” Their place is taken by those willing to cave into pressure from students and administrators, the latter more and more resembling their counterparts in the corporate world. Fox finishes with links to 36 articles critical of SETs, many of them comprehensive studies.

In his oped “Student Evaluation of Teaching: Off With Its Head!”, Miami University professor **Neil B. Marks** argues that in a time of that demands instant gratification, “a professor who inflicts ‘pain’ while striving for benefits to be derived well beyond the semester at hand will be punished.” The only solution is to abolish them completely.

11. THE SANDBOX EXPERIMENT: In his 1996 book *Generation X Goes to College*, **Peter Sacks** (a pseudonym) recounts his experiences of taking his rigorous standards of journalism with him as he switched careers to teach at a small American college full of under-motivated and under-achieving students. Once there, he underwent a culture shock. He found a school full of indifferent and sometimes rude students who weren’t afraid to brow-beat their professors into giving them good grades for minimal work (p. 29), backed by an administration that assumed that the customer (student) was always right (p.23). After a year of grief trying to impose those standards on his students, followed by negative student evaluations that attacked everything from his lecture style to his haircut and advice from his academic colleagues to dumb things down, Sacks had a revelation. He decided to put into play a clever plan he calls “the sandbox experiment”. He treated his college class like a kindergarten teacher: he played games “just to have fun”, gave his students easy assignments and high grades, making them happy at any cost, even that of learning anything (p. 83). He became a teaching superstar, getting kudos from his students and tenure from his colleagues.

12. DISENGAGEMENT: In their 2007 book *Ivory Tower Blues*, **Anton Allahar** and **James Côté** of UWO argue that student disengagement, grade inflation and the dumbing-down of curricula are endemic in higher education. A review by Dave Armishaw in *The College Quarterly* 10.2 (Spring 2007) nicely summarizes their point of view:

Conscientious professors struggle to maintain high academic standards out of personal professionalism and for the sake of students and the institutions. They also contend with the impossibility of accommodating underachievers without neglecting high achieving students. The former group has made an art form out of reducing their workload to a minimum and, because they are well aware of the effect of student ratings on a professor’s career, they know how to leverage faculty to give high marks for little or no effort. They generally lack, it could be added, intellectual curiosity or desire; either while in school or later in life.

Allahar and Cote note that “degree purchasers” who are averse to doing work will get revenge on professors who insist on maintaining reasonable academic standards in their evaluations (p. 39). They make it clear that class evaluations cause grade inflation, and that the “**best predictor of professors’ popularity with students is the grades they give: the higher a student’s grade in a course, the higher his or her evaluation of the professor will be**, net of other factors, and regardless of whether student truly benefits from the course intellectually.” (p. 64). As a whole, this book provides considerable evidence for, and condemns, the instrumental, consumerist

mentality of the partially engaged (40% of the total) and disengaged (40-50%) students who currently rule the roost in our universities.

13. **NARCISSISM: Jean Twenge and W. Keith Campbell** argue in *The Narcissism Epidemic* (Free Press, 2009) that due to helicopter parenting, celebrity culture, the rise of the Web 2.0 and easy credit, narcissism has dramatically increased among young people today, and that this outbreak of self-admiration puts students in possible conflict with the possibility of ego deflation in higher education. Students today are used to being told they are “**special**”. They are conditioned in high school to expect higher grades for less work (p. 237), freely admit to cheating on a massive scale (p. 206), and because of their **sense of entitlement** have no qualms about demanding higher grades even when these aren’t warranted (p. 230). Thus another cause of grade inflation and a decrease in academic standards is the **culture of narcissism**, which preaches the impossible notion that everyone is “better than average”.

14. **WE DON’T NEED NO EDUCATION... OR BOOKS... OR KNOWLEDGE: Mark Bauerlein’s** *cri de couer* from 2008, *The Dumbest Generation* (Tarcher/Penguin), castigates young Americans as a peer-absorbed Net-obsessed group who never read books for pleasure and (to quote the *LA Times* review of July 8, 2008):

...know virtually nothing about history and politics. And no wonder. They have developed a “brazen disregard of books and reading.” [Though] “never have the opportunities for education, learning, political action, and cultural activity been greater,”... the much-ballyhooed advances of this brave new world have not only failed to materialize -- they've actually made us dumber.

The problem is that instead of using the Web to learn about the wide world, young people instead mostly use it to gossip about each other and follow pop culture, relentlessly keeping up with the ever-shifting lingua franca of being cool in school. The two most popular websites by far among students are Facebook and MySpace [*would replace MySpace with Instagram today*]. “Social life is a powerful temptation,” Bauerlein explains, “and most teenagers feel the pain of missing out.”

As he puts it on his web site, technology offered the millennials a false promise:

The dawn of the digital age once aroused our hopes: the Internet, e-mail, blogs, and interactive and ultra-realistic video games promised to yield a generation of sharper, more aware, and intellectually sophisticated children. The terms “information superhighway” and “knowledge economy” entered the lexicon, and we assumed that teens would use their know-how and understanding of technology to form the vanguard of this new, hyper-informed era.

That was the promise. But the enlightenment didn’t happen.

Bauerlein finds in US colleges and universities a generation of entitled multi-taskers, enabled by their high school experiences to expect high grades for a modicum of work. A 2006 survey of 81,499 American high school students found that 90% admit to studying five hours or less per week, with 55% reporting an hour or less of homework. Bauerlein reports in the *NY News* on July 24, 2011 that a study of 200 four-year US colleges and universities showed that 43% of students get As and 33% Bs, and that

“anything less than a B has become a humiliation”. His work defends the reality of grade inflation and an entitled student body hooked on celebrity culture and social networking who, in too many cases, expect a B for just showing up and doing a minimum of work, and will punish teachers who get in the way of this expectation.

We should avoid simplifications of Bauerlein’s position that castigate a whole generation as irretrievably stupid, and take heed of his considerable empirical research. And also remember that even in the worst classes, there will always be a hard core of excellent students who do all the readings, attend all the lecture and try to contribute thoughtfully to the class. But needless to say, the harshest online criticisms of Bauerlein’s thesis are full of spelling and grammar mistakes and the occasional swear word from the very people he’s targeting, not to mention complaints that his book was too “boring” to get all the way through. QED. They fail to understand that he’s not attacking *all young people*, but a statistical average of the “millennial” generation, backed up by reams of empirical studies. The existence of a small minority of hard-working students who read for pleasure is not a refutation of his thesis.

15. *UWO Study*: I did two small studies on the subject myself. In the first, a study of a random sample of 30 courses at UWO dating back to the late 1990s, I found that 52% of students reported expecting an A in their courses, 42% a B, 6% a C, and 0.5% a D. Assuming a low B average as standard, in all 30 cases **grade expectations were inflated**, in 14 cases severely so. In 9 out of 10 cases, a given professor’s **highest grade score** (created by assigning a 7 to expected As, a 5 to Bs, 3 to Cs, and 1 to Ds) correlated with their highest evaluation score, with the tenth case being very close. The average grade score was 5.9/7. Though there were a few anomalies, generally speaking, higher grades bought higher evaluations.

16. *FIMS Study*: In my study of 20 MIT courses at UWO from the 2012-2013 period taught by 20 different professors, all of which were 2000- or 3000-level non-core courses with 40 or fewer students (picked to mirror my own teaching load in FIMS), I found once again that every class showed at least *mild grade inflation*. There was a clear pattern of a minority of expected As and a large mass of expected Bs, with only a single C expected in a few classes. Two or three classes broke this mould by showing the majority expecting an A. This pattern was repeated in the 30 or so other classes taught by these same professors which I looked at. The average grade score was 5.6/7, down somewhat from the general study, though explicable at least in part by students and professors in FIMS being conditioned by grade restrictions in core courses. On average, **31% expected an A, 68% a B, 1% a C**, none a D. The most striking fact his study unearthed is that no one in these courses expected a grade lower than around 62%, while **99% expected at worst a grade of B-**. Though there were fewer expected As than in the general study, there were also considerably fewer Cs and Ds, resulting in a steeper grade spike. Thus the idea of failing a course is unthinkable in MIT.⁴

17. GRAND FINALE: In conclusion, the scholarship shows the following:

- **A. Grade inflation** is a real phenomenon across North America, affecting many

⁴ I excluded my own courses from this study for comparison purposes.

programs here at Western.

- B. Students **work only for grades**, which they connect to **future wages**.
- C. At minimum, professors who give out low grades and/or more work than is usual in comparable courses are **punished on evaluations**. This happens whether the grade is lower *relative to* a student's individual grade average, to the average grade in that class, or is low in the absolute sense.
- D. Professors who give out high pre-evaluation grades and/or dumb down their classes (the sandbox experiment) are **rewarded on their evaluations**. This happens even if everyone in the class is the beneficiary of inflated grades. Further, attractive young professors get higher evaluations than their plainer colleagues, unless they hand out low marks.
- E. SETs are **ineffective instruments** with which to measure the teaching skills of professors or how much students have **learned** in a given class. They may actually cause professors to **put less work** into teaching, and to sign a "mutual non-aggression pact" with their students to buy high evaluations. When we look at grades in subsequent classes, we can see that professors getting mediocre evaluations may actually be **better serving** their student's long-term needs.
- F. Student evaluations are driven by a **consumerist mentality**. They are **harmful to institutions** which value scholarly rigour, since they send students the message that they can **blackmail professors** into relaxing standards and giving out high marks for limited effort. Admittedly, they are good PR for schools that want to brag about the "positive student experience" they offer, since they **camouflage student disengagement**, illiteracy and ignorance of art, philosophy, history and politics with a shiny coat of statistical paint.
- G. Why does a largely cynical professorate still use them? As most critics point out, they **are easy, cheap and fast**, requiring zero labour from senior administrators and professors. Though historically a product of the radical student activism of the late 1960s, this activism has now turned to quiescence in consumer capitalism and its selling techniques. SETs today have more in common with Frederick Taylor's management studies of a century ago than this long-absent campus radicalism, and provide administrators with a way to discipline and punish professors who fail to "provide the goods" to their customers. To echo Neil Postman's critique of technopoly, they measure what cannot be measured (at least with a 5-minute fast-food-style questionnaire) to provide administrators with phony statistics about how effective their employees are at providing intellectual services to their student customers.
- H. Professors who ignore the substantial evidence that SETs are ineffective metrics of teaching effectiveness and learning outcomes do so out of economic self-interest (they benefit from the current system), intellectual sloth ("I've got better things to do!") or wishful thinking ("so many institutions use them: they must work!"). They are very reluctant to debate the issue in public. Yet the science supporting the value of SETs is no better than that supporting global warming deniers or cigarette defenders in the 1950s. Administrators support them to exercise control over contract faculty (see Michel Foucault's works).

A Modest Proposal to Overcome Grade Inflation

1. The basic cause of grade inflation is that even if *some* courses in a given institution have controlled averages, others don't, and the system as a whole has a natural tendency to become an "open market" if there are even a few outliers or free riders. In this market most instructors, especially contract and tenure-track professors eager for student approval, have a strong incentive to inflate grades to keep students happy and therefore get good evaluations.

2. There are two ways to combat the market's tendency to inflate grades, paralleling real world inflation. First, we can regulate all courses grades, just as governments have from time to time frozen wages and prices to fight monetary inflation. But in most institutions this simply can't happen, given departmental autonomy. And even if it could be done, individual professors can still press up against the barrier (which presumably wouldn't be a discreet number, but a range of grades like "65-70%"), or could gerrymander grades so that students get an A average during the term but wind up with the required B average due to tough grading on essays and exams.

3. The other option is to come up with a way of the market regulating itself. I suggest borrowing two statistics from hockey: points (goals + assists) and plus-minus. Each instructor would first grade students using the traditional percentage grades, with whatever grade controls his or her department deems fit. This number would parallel the points stat: so Joan Smith gets an 85% in her Astronomy 100 class, equivalent to a hockey player getting say 70 points in a season.

4. On top of this, using a spreadsheet or an online system such as Owl, the class average of each class would be computed. Then positive or negative differential between the individual student's absolute grade and the class average would be calculated to one decimal point and expressed as a +/- stat, as in hockey. In my example, the average grade in Astronomy 100 is 81.5%, so Joan has a +3.5 in this class. Her transcript would record both numbers, so it would say "Astronomy 100: 85% / +3.5".

5. The advantages of this system are obvious. It defeats the argument from professors who say "I should be able to give everyone in my class an A if they're all brilliant students!" You can. But it also exposes professors who are easy markers to a degree of public shame, which will hopefully cause them to tighten up on their grades more than heavy-handed regulation would. If Professor Jones consistently hands in A-level averages while the rest of the department hands in B averages, Jones might mend his ways without needing a lecture from the chair.

6. This system is also much fairer as a way of determining who gets scholarships or into graduate or professional schools. An example: Joan Smith finishes her degree with an 85% average, with a +3 average overall. She's competing against Harold Cooper, who only has a 78% average, but a +8 average overall. In the old system, Smith beats Cooper by 5%. But in the new system, Cooper beats Smith by 5% since Smith's university has inflated its grades by 12% over Cooper's school. Cooper is the better student since he's beat the average grade in his tougher university by a higher margin.

APPENDIX 1. Anticipated Grades in MIT Courses, 2012-2013

Course	Year	A	B	C	D	A%	B%	C%	Grade Score
2403F	2013	8	5	0	0	62	38	0	6.2
2404F	2012	3	18	0	0	14	86	0	5.3
2510F	2013	8	27	0	0	23	77	0	5.5
2570B	2013	5	13	0	0	28	72	0	5.6
2600A	2012	7	10	0	0	28	72	0	5.8
2901G	2013	13	19	0	0	41	54	0	6.0
2934G	2012	6	11	0	0	42	58	0	5.7
3208F	2012	6	17	0	0	35	65	0	5.6
3209G	2013	5	15	0	0	26	74	0	5.5
3211G	2012	8	18	1	0	30	67	4	5.4
3211G	2012	8	18	1	0	30	67	4	5.4
3371G	2013	2	19	0	0	10	90	0	5.2
3374G	2013	7	10	0	0	41	59	0	5.8
3436G	2012	4	18	0	0	18	82	0	5.4
3438F	2012	5	6	1	0	42	50	8	5.5
3439G	2013	3	8	0	0	27	73	0	5.5
3773F	2012	2	23	0	0	8	92	0	5.2
3829G	2013	11	9	0	0	55	45	0	6.1
3833F	2012	9	17	1	0	33	63	4	5.5
3875F	2012	3	9	0	0	25	75	0	5.5
Totals		123	290	4	0	618	1359	20	111.7
Percents		29.5	69.5	1.0	0	30.9	68.0	1.0	5.6

Methodology: I took the list of MIT courses for the current year, and started doing searches on professor's names for those teaching courses similar to mine – 2000 or 3000-level, non-core courses with enrolments of 40 or less. I only used one course per professor so as not to bias the result too much (so these courses represent 20 different professors, the core of the MIT faculty), and avoided repeating the same course. Most, but not all, are from the academic year 2012-2013, and most of these courses are (as far as I know) taught by contract professors. I exclude the names to provide some measure of anonymity. My purpose was to see if there was a pattern of high grade expectations in these courses, which indeed there was.

Findings: I was quite surprised to see an *extremely* consistent pattern in the grade scores of these courses. In most of the twenty listed above, and in the roughly 30 other courses by these same professors I looked at, I found a minority of students expecting As (31% on average), the majority Bs (68% on average), and in a few courses one (or at most two) students expecting a C (and no Ds or Fs). The only exceptions were the handful of courses where *most* students expected As.

In 16 courses (80%), *no one* expected a C, while only one student did so in the other four. The overall percentages of each expected grade were: 30% As, 70% Bs, and slightly less than 1% Cs. The logical conclusion is that 99% of MIT students are convinced that they'll get at least a low 70 in these courses – i.e. that they are safe from the dreaded C (with a D or F being unimaginable).

The grade score is created by assigning an anticipated A a score of 7, B a 5, and a C or D a 1, then doing a simple average to mirror the evaluation scores (out of 7). A grade score of 5 = a 75% average. Scores of 6 or higher are highly inflated. All twenty classes had scores over 5, averaging 5.6. This translates to an anticipated average of around 79%, which is rather high.