

New Educational Foundations

A Trans-ideological Journal of Criticism, Research, and Review

The Rot Spreads Worldwide: The OECD—Taken In and Taking Sides

Richard P. Phelps

What We Want from Schools (And Why We Are Unlikely to Get It)

Gary K. Clabaugh

Evaluating Teacher Preparation Programs: What Not to Do

M. Suzanne Franco and Martha S. Hendricks

Higher Education Mission Statements: How Vital, How Vacuous?

Bridget Bowers

**Schoolhouse Solutions 1.2:
In Praise of Intolerance**
Wade A. Carpenter

**A Pathological Heuristic:
Dodging Hard Practical Questions**
Edward G. Rozycki

**THE PRACTITIONER'S TOOLBOX
When Mission, Vision, and Belief
Statements Turn Nasty**
Gary K. Clabaugh

**INTERVIEW WITH THE DEAD
Friedrich Nietzsche**

SPRING 2013

New Educational Foundations

A Trans-ideological Journal of Criticism, Research, and Review

COMMENTARY

- Schoolhouse Solutions 1.2: In Praise of Intolerance 1
Wade A. Carpenter
- A Pathological Heuristic: Dodging Hard Practical Questions 7
Edward G. Rozycki

RESEARCH & ANALYSIS

- The Rot Spreads Worldwide: The OECD—Taken In and Taking Sides . . 11
Richard P. Phelps
- What We Want from Schools (And Why We Are Unlikely to Get It) . . 42
Gary K. Clabaugh
- Evaluating Teacher Preparation Programs: What Not to Do 51
M. Suzanne Franco and Martha S. Hendricks
- Higher Education Mission Statements: How Vital, How Vacuous?. . . 59
Bridget Bowers

DEPARTMENTS

- The Practitioner's Toolbox*
- When Mission, Vision, and Belief Statements Turn Nasty 67
Gary K. Clabaugh
- Interview with the Dead*
- Friedrich Nietzsche 76

New Educational Foundations, a refereed online journal of ideas and dialogue, has been established as a forum for independent-minded educators, administrators, and researchers.

We favor no particular ideological bent. We unanimously reject the cultures of complaint and compliance; our audience will be professionals who question conventional thinking and the roar of the crowd.

We encourage you to download and distribute any of the materials in this introductory issue. Potential contributors are encouraged to send a monograph or letter of interest to us as info@newfoundations.com.

Copyright © 2013 by New Educational Foundations, Inc.

Wade A. Carpenter was a high school teacher in the Charlotte-Mecklenburg (NC) Schools for fourteen years, and has taught twenty-four years at the college level. He is Associate Professor of Education (Secondary and Foundations) at Berry College in Rome, Georgia. Email Wade at wcarpenter@berry.edu.



Schoolhouse Solutions 1.2: In Praise of Intolerance

by Wade A. Carpenter

I don't generally get too upset at words. Lifestyles might be approved or disapproved, depending on any number of criteria. I'm generally forgiving toward people, places, and things that I might consider questionable. Certain actions, behaviors, and traits, however, might arouse my more negative passions, especially when they impact the education of children. Tolerance is one of those traits. Come to think of it, I don't like the word much, either.

Words

At those points in the liturgy where optional gender-inclusive language in reference to God or humanity is authorized, I might use it and I might not. It's been quite awhile since God inquired about my genitals, so I don't see why I should inquire about God's. It makes no difference to me. And using "men" in reference to humanity does seem at best half-witted.

I understand why some words might be hurtful to people who suffer from them or who have suffered in the past, so I personally draw the line on some, although I don't usually advocate laws or regulations about them. For instance, the popular black comedian Chris Rock once recorded a monologue in which he boldly (and rightly) distinguished between black people and "n-words." I like Chris Rock, and confess to more than one belly-laugh from that skit. It was funny and he had an admirable message. But I cannot endorse his use of the n-word, even in humor with good intent. Never in history has

that word done any human being any good, and it has done an awful lot of people an awful lot of harm. Therefore, I reckon I can live very satisfactorily without it. Let this banning of that word from my presence be my last word on the subject.¹ BTW, I'm beginning to feel the same way about the "r-word" in reference to rural Southerners.

Likewise, misplaced apostrophes irritate me, and I must admit to my membership in the International Apostrophe Police. (We are rumored to go around late at night furtively redmarking apostrophic misapplications in public places.) I was also recently appalled to learn from an English professor that "oftentimes" is now considered acceptable. No it's not, damn it.

But that brings me back to the subject: I suppose all those years of teaching in an urban high school desensitized me to profanity and vulgarity. Frontiers tend to coarsen people, and the modern American high school is civilization's newest frontier. And I guess all these years of teaching teachers at the college level has desensitized me to what the law calls "fighting words"—I'm accustomed to being gratuitously insulted by journalists and politicians.² So I find it hard to sympathize with political correctness on the left or verbal puritanism on the right. In just about any class in my own cheerfully religion-friendly college there will be one person who sits unaffected by horror stories of educational misconduct, economic injustice, and human tragedy, but who will come unglued if I use the word "hell." After twenty years here, I would almost rather call a fundamentalist a ____ to his/her face than use the h-word even in its technical sense. I do not understand that version of morality, nor do I care to. Our hate-school (the rival college on the other side of town, which shall remain nameless [Shorter University]) is a rigorously fundamentalist school, and I might occasionally refer to them as ____s anyway, so I am certainly not without my own culpabilities. Therefore, I seldom get emotional about words.

Because I'm aware of my own flawed nature, however, one word really does annoy me. That word is "tolerance." It is almost always hypocritical. I *know* I am not universally tolerant, and I know you are not, either, so let's not pretend. In fact, intolerance may be perfectly appropriate when it is of certain words, lifestyles, people, places, things, behaviors, actions, or traits. It is also okay to be intolerant of overly tolerant people. People who think they tolerate everything are deluded—surely nobody's *that* unobservant. People who say they tolerate everything lie. Even people who tolerate *too much*

are spineless, and therefore inherently untrustworthy. Cultural relativism in anything like an absolute sense is simply nonsense. Some traits, habits, and behaviors are simply not desirable, and some are downright wrong. Yes, there is much that was admirable in ancient Athens, and without at least a decent familiarity with Athenian civilization, Western history inescapably becomes a tedious load of memorization instead of a worthwhile and often delightful study of what makes people tick. But it also worth remembering that Socrates probably would have considered Jerry Sandusky normal and the rest of us bigots. Since then, we have come to understand that mature and committed love between adults is better than sexual coercion of children. In fact, one is good and the other is bad. I might even be retro enough to use the word "sin." Likewise, most of us have figured out that discriminating for or against people on the basis of their skin color is morally contemptible, economically unprofitable, socially suicidal, and personally cruel. In short, it is intolerable. But I'd prefer to call it *evil*. "Tolerance" is also a weasel-word, an ambiguity that immediately invalidates whatever precedes it and whatever follows it. Really, think about it: When was the last time you told your sweetie, "You know, dear, I really tolerate you"? I'd be happy to read any accounts you might have of that conversation, if you still have use of your writing hand.

Other words and terms I'm not tolerant of:

- "segue"
- "paradigm shift"
- "continuous improvement"
- "current best practice"
- "the student [singular] . . . they. . ."

Lifestyles

Lifestyles don't interest me much, either. For example, I don't have any problem with the proposed blessing for single-sex partnerships that is causing such a stir in my denomination. I've never considered what two consenting adults do with their privates to be any of my business (as long as they don't do it in my class!), and heaven knows I have asked the Lord's blessing on lots of things that I am uncertain about. I'll let the Almighty work out the details. God's smart enough, even if I'm a bit slow on the uptake.

On the other hand, I do tend to get fatigued when people belabor me with their enthusiasms, especially if they are in a hurry. I don't know which frightens me more: a liberal with a cause, or a Christian-School soccer mom in traffic.

Lifestyles I like to keep at a distance:

- Celebrity. I'm perfectly happy to live a continent away from Hollywood. On the other hand, I cannot get far enough away from D.C.
- Mafia, of whatever ethnicity.
- Rap and hip-hop, especially when applied to mathematics. Please tell me I don't have to listen to that stuff everywhere I go.

People, Places, and Things

Although it is a cliché that is dreadfully overused, it is indeed quite possible to hate the sin without hating the sinner. I don't like meth or murder, so I avoid meth dealers and gang-bangers. That doesn't mean I hate anybody; just that I like to keep far from them. Some places are undesirable—my boss once sent me to a conference in Detroit in February. He still owes me. There are a lot of good things that may be misused and also things that are inherently repulsive. Once when hospitalized I offered to trade the chicken soup for the barium sulfate, ounce for ounce. How can anybody screw up chicken soup so badly?

People, Places, and Things I'm not particularly tolerant of:

- Child molesters in denial.
- "Dynamic" leaders and administrators.
- Folks who beat themselves up for being white (or black, or successful, or [fill in the blank]). Get a life, people!
- Evangelists who cheat.
- Teachers who cheat, and the administrators who force them to.³
- Politicians who try to attract supporters by beating up teachers.
- Dissertation directors who claim to believe in the Thirteenth Amendment.
- People who blow people out of airplanes.

- Nations and corporations who mistreat people and then depend on Uncle Sam to bail them out when the fan gets hit.
- Magazines that proclaim their environmentalism, and then put loose subscription postcards within their covers.
- Nightly news. (If it bleeds, it leads.)
- People who don't fish. Most of humanity's problems are caused by people who don't fish; *all* of humanity's problems are caused by people who don't fish enough.

Behaviors, actions, and traits

Prejudice can be good if the predictable consequences of tolerance are bad enough. Like intolerance, prejudice has become a word we are prejudiced against without having given it adequate thought. Nowadays teachers are uniformly told to "celebrate diversity" by people who should know better. *Sometimes* diversity is to be celebrated, and sometimes unity is to be sought. But perversity and groupthink are neither to be celebrated nor sought. Prejudice, in another sense, is helpful and even necessary: It is perfectly okay to automatically go for the brake pedal when the traffic light turns red. There is no good reason why we shouldn't visit or send "care packages" to lonesome, terrified, clueless immigrants who sit in jail for months waiting for the INS to process their deportation. Why wouldn't you say a quick prayer when you encounter an ambulance in a hurry? You really don't have to spend a lot of time thinking about such things. In 1790 Edmund Burke acclaimed prejudice in this broad sense nicely:

Prejudice is of ready application in the emergency; it previously engages the mind in a steady course of wisdom and virtue, and does not leave the man hesitating in the moment of decision, sceptical, puzzled and unresolved. Prejudice renders a man's virtue his habit; and not a series of unconnected acts. Through past prejudice, his duty becomes part of his nature.⁴

Modern-day Jacobins, take note.⁵

Behaviors, actions, and traits I do assume are bad:

- Bullying—I can't guarantee it won't happen in my class, but I *can* guarantee it won't happen twice.
- Sexual harassment—But just as the courts are now finally recognizing that a "hostile environment" is not produced by

an isolated stupid joke, I hope teachers will continue to recognize the difference between the harassment that dishonors and threatens versus the flirting that ensures the survival of our species.⁶

- Ingratitude—Why stifle one of the greatest and longest-lasting joys of life?

Movements

Movements scare the daylight out of me, largely because I have been burned way too often by them. Probably the worst legacy of the 1960s is the wretched true-believer lie that “If you’re not part of the solution, you’re part of the problem.” That is and always has been the logic of the terrorist. Teachers: for heaven’s sake, please don’t teach that crap to children.

Movements I will not support:

- Any of them.

Notes

1. http://www.youtube.com/watch?v=f3PJF0YE-x4&feature=player_detailpage. To his great credit, Mr. Rock no longer performs that routine.
2. David Schimmel, Leslie Robert Stellman, and Louis Fischer, *Teachers and the Law*. 8th ed. (Boston: Pearson, 2011): 157–158.
3. We were planning to include an insider’s view of the Atlanta cheating scandal in this issue of *New Educational Foundations*, but the principal author was stricken by major health issues while in the midst of his first revision, so we have delayed it for now, and wish him a speedy recovery.
4. J. C. D. Clark, *Edmund Burke: Reflections on the Revolution in France: a Critical Edition* (Stanford, Calif.: Stanford University Press, 2001): 251–52.
5. Again, Burke: “What is Jacobinism? It is the attempt . . . to eradicate prejudice out of the minds of men, for the purpose of putting all power and authority into the hands of persons capable of occasionally enlightening the minds of the people. For this purpose the Jacobins have resolved to destroy the whole frame and fabric of the old societies of the world, and to regenerate them after their fashion. To obtain an army for this purpose, they everywhere engage the poor by holding out to them as a bribe the spoils of the rich.” In Russell Kirk, *Edmund Burke: A Genius Reconsidered* (New Rochelle, N.Y.: Arlington House, 1967): 201.
6. David Schimmel, Louis Fischer, and Leslie R. Stellman, *School Law: What Every Educator Should Know* (Boston: Pearson, 2008): 109.

Edward G. Rozycki, Ed.D., served seventeen years as an associate professor of education at Widener University, Widener, Pennsylvania. He is the webmaster and co-sponsor of the article banks at www.newfoundations.com.



A Pathological Heuristic: Dodging Hard Practical Questions

by *Edward G. Rozycki*

If you can't solve a problem, then there is an easier problem that you can solve: find it.

—George Polya, *How to Solve It*
(cited in Kahnemann, 98)¹

The question-substitution heuristic

Does watching violence on TV make kids more violent? The answers to that question, it appears, could have important practical consequences. But the question itself is problematic: it is multiply ambiguous.

It used to be all too common, although somewhat bizarre, to observe even "professionally trained" people spending a lot of time and energy vehemently debating this question without taking the trouble to first determine whether they all understood the terms to mean the same (or similar) things.

Getting practical answers to vague questions often requires operationalization: that is, the specification of vague terms: e.g., "watching," "violence," by providing replicable, observable procedures for their determination.²

For example:

1. What is "watching TV"? Need a child be paying close attention, or would just having the set on in the background count? How do we determine how much TV a child is watching?
2. What counts as violence? Football? *Mighty Mouse*? A dramatization of an assault? Documentary footage from a war?
3. How are we to determine if kids have become more or less violent? From their play-acting? From their actual fighting? From their arguments or threats?

In most social situations, people would find this detailing process a tedious distraction from the entertainment purposes of "debate," typically a sort-of quasi-intellectual arm wrestling, or a contest in one-upmanship. Easy questions are usually thought "funner" than hard ones if only because they can be appreciated by a wider, "technically challenged" audience.

Consider now another example: the case of the office duplicating machine that occasionally malfunctions. When that happens, the question is not "Why did the machine malfunction?" That might take an expert, a repair technician, to determine. Rather, the question posed is more likely to be "Who was using [standing near, looking at, thinking about] the machine when it broke down?" Those are much easier questions to deal with: they provide those who pose them with comfortable illusions of understanding. In addition, they promise a lead into that most entertaining past-time: assigning blame.

Daniel Kahneman calls this tendency the Question-Substitution Heuristic. In his book *Thinking, Fast and Slow*, he furnishes a chart of comparisons between what he calls Target Questions (hard) and Heuristic Questions (easy).³

Chart 1: abridged from Kahneman (2011), 98–99

Target Question	Heuristic Question
How much would you contribute to save and endangered species?	How much emotion do I feel when I think of dying dolphins?
How popular will the president be six months from now?	How popular is the president right now?
This woman is running for the primary. How far will she go in politics?	Does this woman look like a political winner?

Note that working out answers to the target questions, especially when money is involved, is often a laborious undertaking.

By contrast, heuristic questions can be pretty much answered off the top of the head: they reduce the information base for answering to what the questioning persons can easily access in memory. How likely this procedure is to provide good solutions to real problems one might judge by the example of participants of TV talk shows and no small number of our political and social pundits.

But even professional scientists have been criticized for providing answers to questions they seem to have reformulated as easy.⁴ The congressional response in 1958 to the question "Why did the United States fail to beat the Russians into space?" was reformulated in the easier "who's-to-blame" mode, and ultimately answered with the passage of the National Defense Education Act.⁵

The Theory-to-Practice Gap in Education

Public educational institutions in our democratic society are often under severe pressure to widen participation, little matter whether those to be involved are even minimally informed or not. Consequently, questions upon which the most practiced and learned professionals may disagree are usually decided by the least-practiced and least-knowledgeable of people.

What is mathematics? Professional mathematicians can and do disagree. Likewise professional historians on history, and professional political scientists on political science. But for educational systems, elected local school boards, who must act to distribute tax-funded budgets, decide all those questions.

The operative heuristic for cost-chary school boards dealing with educational questions is an interrogatory that looks somewhat like this:

- a. Are there any foreseeable, imminent, and severe repercussions to our ignoring this question? If not, table it.
- b. Is there any demand for any of these subject matters from influential constituencies (ICs)? If not, table it.
- c. What costs would different programs of, say, mathematics education entail, were we to decide to implement one of them in the schools? Get those estimates. (Don't rush. Table it.)
- d. Will our ICs likely support us in our decisions? If not, "table" it.⁶

Let's mimic Kahneman's chart 1. The target question group, in the left column, will contain the kinds of questions encountered in

teacher preparation courses (or casual public discussions). In the right column we'll place the heuristic questions, the kinds of questions likely to be substituted by school boards or other governance body's committees.

Target Question	Heuristic Question
What are the aims of education?	How can we use our schools to satisfy the demands of different community and political constituent groups?
Should elementary pupils have art (science, math, gym) classes?	Can we make room in the budget for that without threatening the sacred cows of influential people?
Should all K-12 students be prepared to attend college?	Whom can we overlook without raising a din that threatens our tenure as school board members (administrators, teachers, politicians)?
Should Johnny be disciplined for knocking down other pupils in the classroom without provocation?	Does Johnny belong to any group with a protected status?

From chart 2 and the preceding interrogatory we can see why, despite much lip service to the contrary, educational decision-makers will likely continue to keep academic programs in distant second place compared to their political concerns.

Notes

1. Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
2. Rozycki, E. G. *Operationalization* (1994), available at <http://www.new-foundations.com/EGR/Oper.html>.
3. Depending on circumstances, heuristics often work and prove much more economical than alternative processes. See Gigerenzer, G., and Todd, P. (1999). *Simple Heuristics That Make Us Smart*. NewYork: Oxford U Press.
4. See Moyer, M. "Person of the Year" Nomination for Higgs Boson Riddled with Errors *Scientific American* Blog, Nov. 29, 2012, at http://blogs.scientificamerican.com/observations/2012/11/29/person-of-the-year-nomination-for-higgs-boson-riddled-with-errors/?WT_mc_id=SA_CAT_BS_20121130.
5. See Rozycki, E. G. (2008) *Illogic and Dissimulation in School Reform*, available at <http://www.newfoundations.com/EGR/IllogicReform.html>.
6. See Rozycki, E. G. *Developing Interrogatories to Aid Analysis, Problem-Solving and Conflict Resolution*, available at <http://www.newfoundations.com/EGR/Interrogatories.html>

OECD encourages world to adopt failed U.S. education programs

IN THE AFTERMATH OF Hurricane Katrina, one of the United States' largest insurers refused to honor damage claims from Gulf Coast customers who submitted hurricane insurance claims, asserting that their property had been damaged not by hurricane, but by flooding. Only a high-stakes, high-profile class-action lawsuit ultimately pried the insurance payments loose. Currently, this large insurance company, with its own trust issues, is running a series of television commercials poking fun at an institution that

it assumes is trusted by the public even less—the Internet. “They wouldn’t put it on the Internet if it wasn’t true,” says the naïve foil who purchased allegedly inferior insurance after believing the promises in an Internet advertisement, presumably eliciting off-screen laughter in millions of living rooms.

Now suppose that you are responsible for learning the “state of the art” in the research literature on an important, politically sensitive, and hotly contested public policy topic. You can save money by hiring master’s-level public policy students or recent graduates, though none with any particular knowledge or experience in the topic at hand—a highly specialized topic with its own doctoral-level training, occupational specializations, and vocabulary. You give your public policy masters a computer with an Internet browser and ask them to complete their reports within a few months. What would you expect them to produce?

You can see for yourself at the website of the Organisation for Economic Cooperation and Development (OECD).¹ In 2009 the OECD launched the Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Apparently the “Review” has not claimed an acronym. In my own interest, then, I give it one—REAFISO.²

In its own words, REAFISO was created to provide analysis and policy advice to countries on the following overarching policy question:

“How can assessment and evaluation policies work together more effectively to improve student outcomes in primary and secondary schools?”

To answer this question, the OECD intended to

look at the various components of assessment and evaluation frameworks that countries use with the objective of improving the student outcomes produced by schools. . . .

and

extend and add value to the existing body of international work on evaluation and assessment policies.

I was interested in REAFISO’s work for two reasons. First, I once worked for the OECD, on two fixed-length consulting contracts

accumulating to sixteen months. I admired and respected their education research work and thoroughly enjoyed my time outside work hours. (The OECD is based in Paris.) I particularly appreciate the OECD's education (statistical) indicators initiatives.

Second, I have worked myself and on my own time to address the overarching question they pose, ultimately publishing a meta-analysis and research summary of the effect of testing on student achievement. Because I lacked the OECD's considerable resources, it took me some time—a decade, as it turned out—to reach a satisfactory stage of completion. I hedge on the word “completion” because I do not believe it possible for one individual to collect all the studies in this enormous research literature.

To date, after reading more than three thousand studies, I have found about a third of them appropriate for inclusion in a summary of qualitative studies and meta-analyses of quantitative and survey studies. I looked at studies published in English between 1910 and 2010 that I could obtain and review before a self-imposed deadline in 2010. My coverage of the research literature, which is far from complete, includes 244 qualitative studies (e.g., direct observations, site visits, interviews, case studies); 813 individual item-response group combinations from 247 survey studies (e.g., program evaluation surveys, opinion polls); and 640 separate measurements of effects from 177 quantitative research studies (e.g., regression analysis, structural equation modeling, pre-post comparison, experimental design, or interrupted time series design). In total, I analyzed 1,671 separate effects from 668 studies.

A summary has been published in the *International Journal of Testing* (Phelps, 2012b). Source lists can be found at these three virtual locations:

<http://www.nonpartisaneducation.org/Review/Resources/QuantitativeList.htm>

<http://www.nonpartisaneducation.org/Review/Resources/SurveyList.htm>

<http://www.nonpartisaneducation.org/Review/Resources/QualitativeList.htm>

All but a couple of these several hundred sources were available to REAFISO as well. Yet despite having many times the resources at their disposal, they managed to find just a few percent of what I found. Granted, the search parameters (as best I can discern theirs) were not exactly the same, but were far more alike than different. Not surprisingly, a review of a great expanse of the research literature,

rather than just the selective, tiny bit covered by REAFISO, leads to quite different conclusions and policy recommendations.

Deficiencies of the OECD's REAFISO research reviews include:

- overwhelming dependence on U.S. sources;
- overwhelming dependence on inexpensive, easily-found documents;
- overwhelming dependence on the work of economists and education professors;
- wholesale neglect of the relevant literature in psychology (the social science that invented cognitive assessment) and from practicing assessment and measurement professionals; and
- wholesale neglect of the majority of pertinent research.

Moreover, it seems that REAFISO has fully aligned itself with a single faction within the heterogeneous universe of education research—the radical constructivists. Has the OECD joined the U.S. education establishment? One wouldn't think that it had the same (self-) interests. Yet canon by canon by canon, REAFISO's work seems to subscribe to U.S. education establishment dogma. For example, in her report *"Assessment and Innovation in Education,"* Janet Looney writes

Innovation is a key driver of economic and social programs in OECD economies. If absent, innovation growth stalls; economies and communities stagnate. . . . (p. 6)

Teaching and learning approaches considered as innovation, on the other hand, are generally characterized as being "student-centered," or "constructivist." (p. 8)

This report has focused on [the] impact of high-stakes assessment and examinations on educational innovation. It has found significant evidence that such assessments and examinations undermine innovation. (p. 23)

First, Looney equates national economies and school classrooms. Then she adds the famous economist Joseph Schumpeter's definition of innovation as "creative destruction." For radical constructivists, and apparently for Looney, each teacher is a unique craftsman in a unique classroom, and anything done to standardize their work inhibits their potential to guide each unique student in his or her

own unique, natural discovery of knowledge. To radical constructivists, there are no economies of scale or scope in learning.

Whereas innovation is a holy commandment for the U.S. education professoriate, critics charge that it leads to a continuous cycle of fad after fad after fad. After all, if innovation is always good, then any program that has been around for a while must be bad, no matter how successful it might be in improving student achievement. Moreover, if the pace of today's-innovation-replacing-yesterday's-innovation proceeds fast enough, evaluation reports are finished well after one program has been replaced by another, become irrelevant before they are published, and end up unread. Ultimately, in a rapidly innovating environment, we learn nothing about what works. Some critics of the radical constructivists suspect that that chaotic, swirling maelstrom may be their desired equilibrium state.

A case in point is the sad and expensive 1990s saga of the New Standards Project in the United States and the most deliberate effort to implement its assessment formula in practice, the State of Maryland's MSPAP (for Maryland School Performance Assessment Program). REAFISO writer Allison Morris (p. 16) cites Thomas Toch's (2006) erroneous assertion that cost considerations reversed the 1980s-1990s U.S. trend toward more performance testing. Not so, the MSPAP and similar programs (e.g., CLAS [California Learning Assessment System] and KIRIS [Kentucky Instructional Results Information System]) failed because test reliability was so low, test scores were too volatile to be useful, feedback was too late and too vague to be useful, and parents thought it unfair when their children's grades depended on other students' efforts (in collaborative group activities).

Resounding public disgust killed those programs. But ten years is a long time in the ever-innovating world of U.S. education policy, long enough for the young REAFISO writers to be unaware of the fiascos. REAFISO now urges us to return to the glorious past of New Standards, MSPAP, CLAS, and KIRIS, dysfunctional programs that, when implemented, were overwhelmingly rejected by citizens, politicians, and measurement professionals alike.

What happened at the OECD?

REAFISO relies on staff generalists and itinerant workers to compose its most essential reports. I suspect that the REAFISO writers started out unknowing, trusted the research work they found most

easily, and followed in the direction those researchers pointed them. Ultimately, they relied on the most easily and inexpensively gathered document sources.

I believe that REAFISO got caught in a one-way trap or, as others might term it: a bubble, echo chamber, infinite (feedback) loop, or myopia. They began their study with the work of celebrity researchers—dismissive reviewers—researchers who ignore (or declare non-existent) those researchers and that research which contradicts their own (Phelps, 2012a)—and never found their way out. Dismissive reviewers blow bubbles, construct echo chambers, and program infinite loops by acknowledging only research and those researchers they like or agree with.

The research most prominently listed in Internet searches for REAFISO's topics of interest is, as with most topics on the Internet, that produced by groups with the money and power to push theirs ahead of others'. When librarians select materials for library collections, they often make an effort to represent all sides of issues: such is ingrained in their professional ethic. Internet search engines, by contrast, rank materials solely by popularity, with no effort whatsoever to represent a range of evidence or points of view. Moreover, Internet popularity can be purchased. In education research, what is most popular is that which best serves well-organized and wealthy interests.

The research literature on educational assessment and accountability dates back to the late nineteenth century, after Massachusetts' Horace Mann and his Prussian counterparts, earlier in the century, had initiated the practice of administering large-scale versions of classroom examinations across large groups of schools to compare practices and programs (Phelps 2007b). So-called "scientific" assessments were invented around the turn of the century by several innovators, such as Rice and Binet, and their use was already widespread by the 1920s. The research literature on the effects of these and more traditional assessments had already matured by the 1940s. Some assessment and evaluation topics had been researched so heavily in the early and middle decades of the twentieth century that their researcher counterparts (in psychology) in more recent times have felt little compulsion to "re-create the wheel." If one limits one's search to recent research, one may not find the majority of it, nor the most seminal.

Dismissive Reviews Lead into One-Way Traps*

In scholarly terms, a review of the literature or literature review is a summation of the previous research that has addressed a particular topic. With a dismissive literature review, a researcher assures the public that no one has yet studied a topic or that very little has been done on it. A firstness claim is a particular type of dismissive review in which a researcher insists that he is the first to study a topic. Of course, firstness claims and dismissive reviews can be accurate—for example, with genuinely new scientific discoveries or technical inventions. But that does not explain their prevalence in nonscientific, nontechnical fields, such as education, economics, and public policy.

Dismissive reviewers typically ignore or declare nonexistent research that contradicts their own. Ethical considerations aside, there are several strategic advantages:

- first, it is easier to win a debate with no apparent opponent;
- second, declaring information nonexistent discourages efforts to look for it;
- third, because it is non-confrontational, it seems benign and not antagonistic; and
- fourth, there is plausible deniability, i.e., one can simply claim that one did not know about the other research.

When only one side gets to talk, of course, it can say pretty much anything it pleases. With no counterpoint apparent, “facts” can be made up out of thin air, with no evidence required. Solid research supportive of opposing viewpoints is simply ignored, as if it did not exist. It is not mentioned to journalists nor cited in footnotes or reference lists.

Dismissive reviews are not credible to outsiders, however, when contradictory research is widely known to exist. Thus, the research that remains—that which cannot credibly be dismissed as nonexistent—must, instead, be discredited. In such cases, the preference for dismissive reviews must be set aside in favor of an alternate strategy: misrepresent the disliked study and/or impugn the motives or character of its author.

Dismissive reviewing can be effective and profitable. The more that dismissive reviewers cite each other (and neglect to cite others), the higher they rise in academe’s status (and salary) hierarchy. In the scholarly world, acknowledgment is wealth and citations are currency.

By contrast, researchers with contrary evidence whose work is ignored are left in the humiliating position of complaining about being left out. If those responsible for their ostracism can claim higher status—by teaching at more prestigious universities, serving on more prestigious commissions and panels, and receiving larger grants—naïve outsiders will equate the complaints with sour grapes. After all, everything else being equal, an ordinary observer is more likely to trust the research pronouncements of, say, the chemistry professor from Harvard than the chemistry professor from No-name State College. One has faith that the community of chemistry researchers has properly designated its authorities. Is the same faith warranted for professors in U.S. education schools?

*See Phelps 2007a, 2008/2009a, 2012a

I made no extra effort to find older sources in my literature review of several hundred sources on the effect of testing on student achievement. As a result, my search was biased toward more recent work. It is easier to obtain—more likely to be available in electronic form, more likely to be available at no cost, and so on. Still, half of my sources were written prior to 1990.

Of the 900+ references contained in the eight OECD staff and contractor reports I reviewed, only 19 were produced before 1990, and just 112 between 1991 and 2000. More than 800 sources were written after 2000. Why this complete neglect of a century's worth of information in favor of that from just the past decade or so? Does the OECD believe that human nature fundamentally changed around the year 2000? Probably not, but consider this: the World Wide Web came online in the 1990s.

To conduct my literature searches, I spent hundreds of hours inside academic libraries reading microfiche and accessing expensive on-line databases or remote archives. Had I wanted to be more thorough, I would have paid for interlibrary loan access, even international library loan access. As it was, the work was plenty tedious, time-consuming, and expensive. I suspect that OECD researchers eschew doing research that way, and it shows in the myopia of their product.

In fairness to the OECD, one particular assessment method, to my knowledge, was rarely studied prior to the past couple of decades—using student test scores to evaluate teachers. But this was only one of several research literatures REAFISO claims to have mastered. For the others, its claims of thorough coverage are grossly exaggerated.

The OECD on educational testing and accountability

Officially founded in 1948, the Organisation for Economic Co-operation and Development (OECD) is the stepchild of other post-World War II transnational economic unions, such as the International Monetary Fund (IMF) and the World Bank. Formed by eighteen non-Communist European countries, it was originally purposed to manage American and Canadian financial aid for the continent's reconstruction. It has since entrepreneurially shaped itself into a "rich country club" with thirty-four members from all continents (save Africa and Antarctica).

The OECD's growth strategy has by necessity been opportunistic; the IMF, World Bank, and other organizations had already laid claim to the more obvious roles for multinational economic unions.

Relative to other sectors, however, education had been given little attention.^{3,4}

For a quarter-century, OECD has published its now-annual *Education-at-a-Glance* collection of “indicators,” comparing education systems at the country level on a wide variety of statistics related to enrollment, level of educational attainment, finance, and staffing. For the past ten years, the OECD’s Programme for International Student Assessment (PISA) has tested fifteen-year-olds around the world in reading, science, and mathematics and published the results for comparison.

All along, the OECD has also conducted research reviews on various education topics and organized country-level consulting visits. Typically, country-level reviews gather several experts from among OECD staff, the education ministries of member countries, other international organizations, and university faculty to spend a week or two meeting a full range of responsible officials in a single host country. Afterward, a report full of recommendations is written, reviewed, and written again.

Americans can be rather jaded and parochial regarding international organizations. Most countries hosting OECD study teams, however, take them quite seriously. The structure of a country-level review is negotiated between country and OECD and costs are shared. Reviewers are invited in and busy officials are required to give freely of their time and resources to aid the evaluation.

In the OECD’s own words,

The OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes, launched in late 2009, is designed to respond to the strong interest in evaluation and assessment issues evident at national and international levels. It will provide a description of design, implementation and use of assessment and evaluation procedures in countries; analyse strengths and weaknesses of different approaches; and provide recommendations for improvement.

The Review looks at the various components of assessment and evaluation frameworks that countries use with the objective of improving student outcomes. These include student assessment, teacher appraisal, school evaluation, and system evaluation. The analysis focuses on primary and secondary levels of education.⁵

Housed within the OECD's Directorate for Education's Early Childhood and School Division, REAFISO seems typical of OECD efforts—committing only a tiny (N=5) full-time staff, while expecting to leverage expertise and other resources from member country personnel and institutions, or short-term contract workers, as it goes along. Its web pages exude a sense of temporariness; the review was scheduled for 2009 to 2012, but not necessarily any longer than that.

Note that the OECD rather confusingly tags the word "Review" with multiple meanings:

- a particular office and its employees within the OECD's Directorate for Education;
- a research project on the topic of testing and accountability scheduled for 2009 through 2012; and
- a 1- or 2-week visit to study a host country's education testing and accountability programs by OECD-assembled teams of experts.

By the end of 2012, the efforts of the Review on Evaluation and Assessment Frameworks for Improving School Outcomes (REAFISO) had amassed a large cache of documents. There are several research summaries on student standardized testing, teacher and school evaluation, alignment, and "assessment and innovation" (about 50 pages each); a dozen or so country reviews (at 150 to 200 pages); equally long "background reports" written by host countries prior to OECD study team visits; minutes of OECD-wide meetings of national experts and country representatives; an occasional newsletter; and miscellaneous other missives.

I have read everything I could find on the Web on the background and training of the OECD staff and the contractors they hired to write the REAFISO research summaries. As far as I can tell, none of them has had any training or experience working in assessment and measurement. Rather, they are smart people with general training and experience in research and specific knowledge of other education topics. In such situations, the initial literature reviews can determine the direction of the entire project, either keeping it on course or, as it turned out in REAFISO's case, steering it off its intended course. Though it may be the least-interesting (and typically is the most tedious) task in the research process, the literature search and review can be the most important, particularly in hotly contested, bitterly disputed venues such as education policy.

The Internet Has Changed How Some Conduct Research . . . Unfortunately**

One common approach to research, employed by many journalists and university students, is to instigate a World Wide Web search by typing all the relevant keywords that come to mind into a Web search engine (e.g., Yahoo! Search, Bing, Google). One obtains a long reference list from a few clicks of the mouse. Only fuddy-duddies would complain; it's so much more convenient, it must be better than traditional search methods.

What would the fuddy-duddies say? They might say: most information is not available on the Web and that which is, is ranked by "popularity," which often means that those with money and power can push their point of view ahead of others'.

In the old days of library card catalog searches, references on a topic were delivered unordered, burdening the searcher with the responsibility of considering various points of view. The Web has relieved us of this burden by ranking all references. A researcher working on a deadline, confronted with a few thousand sources from a World Wide Web search on a particular topic, is forced by time limitations to choose only some of them. What is a naïve, unknowledgeable researcher to do when presented with a thousand choices? You and I would probably do the same—pick from among the highest-ranking sources and assume that they represent the whole.

In card-catalog days, the choosing method might have been random sampling among all the seemingly equal possibilities. These days, the choices are made for us; search engines rank sources by popularity, which is often, if inappropriately, equated with importance, prestige, and accuracy. Compare, for example, the outcome of the publications of researchers X and Y. Researcher X is an unaffiliated individual who submits his work to a scholarly journal (or journals), does what he is told to do no matter how arbitrary by the anonymous reviewers, and, if successful, gets his work published in a journal—one among thousands of journals—in due course, that being a year to several years after first submitting his manuscript for review.

Meanwhile, Researcher Y is a brand-new Ph.D. graduate from Harvard University, liked by faculty with Washington, D.C., connections. He is hired to work at the Brookings Institution, the American Enterprise Institute, or some other made-for-D.C. institution with ample funding; you pick. Unlike Researcher X's report, forced to tread the often-arbitrary, often-haphazard gauntlet of peer review, and to rely on the penurious benevolence of a scholarly journal's meager budget for publication, his is reviewed only by those who already like him and are sympathetic to his views. The Public Affairs Office of the Think Tank arranges a conference for "the release" of the publication: ordering *hors d'oeuvres*, inviting the media, and broadcasting press releases to the masses. Researcher Y's publication is professionally edited and formatted and freely available for download off the Internet, whereas Researcher X's journal article is available only with an annual subscription or an exorbitant one-time fee.

If the research conclusions of Researcher X and Researcher Y differed diametrically, and you were forced by time and expense limitations to choose, which would you choose to believe? Consider that Researcher Y's publication will rank very high in search engine rankings and Researcher X's will rank very low.

**See Herring 2001; Phelps 2005a, 2007a, 2008/2009a, 2012a; Stevens-Rayburn 1998.

In his literature review of high-stakes testing, Morten Rosenkvist (2010) asks:

What is the empirical evidence on the effects of using student test results for accountability and improvement? (p. 5)

He lists about 165 references and he claims (p. 6):

Empirical evidence has been identified through a broad search in the literature using databases and search engines such as *ScienceDirect*, *Jstore*, *Google Scholar*, *ERIC*, and *SpringerLink*. The search has been conducted in English, the Scandinavian languages, and to a lesser degree German and Spanish.

Sounds thorough, and Rosenkvist's discussion of his search method is the most thorough among the eight REAFISO research summaries. But he did not follow the standard meta-analyst's recipe. We do not know which keywords he used, in which time periods, how he decided which studies were relevant and which were not, how many studies he reviewed, or anything else. He basically says he cast a wide net, and we're supposed to trust that he trawled in a section of the sea representative of the entire research ocean.

Surveying his reference list, I found a dozen sources, among the 165, that pre-date the year 2000, and only one that pre-dates 1990, this for an enormous research literature that dates back to the nineteenth century. Ergo, simply by chronology, hundreds of relevant studies are ignored.

By far the majority of the research on this topic has been conducted by psychologists, the folk who invented cognitive assessment in the first place and, to this day, who continue to produce, manage, administer, and score almost all of them. Yet among his 150-odd sources Rosenkvist includes zero psychology journals. By contrast, 43 references lead to political science and economics journals.

The majority of the REAFISO research summaries' references lead to U.S. sources, and two groups of researchers are cited most frequently—those affiliated with the federally funded Center for Research on Education Standards and Student Testing (CRESST)—a consortium of the education schools at UCLA, U. Colorado, U. Pittsburgh, and the Rand Corporation—and a small group of economists (mostly) affiliated with Republican Party-oriented think tanks, such as the Hoover Institution and the Manhattan Institute. Most of the CRESST researchers have testing and measurement training, but espouse a particular doctrine of assessment policy that, contrary to

their claims, is not shared by most in the profession. This particular group of economists has no practical training or experience in testing and measurement and has, for reasons unknown to me, fallen head-over-heels for CRESST doctrine.

Adding internal references to other OECD documents to those for these two groups, one can account for about half of all the references. At the same time, REAFISO blanks on the majority of the research in the field.⁶

Based on my review of the REAFISO reviews, I conclude the following:

- The OECD conducted arguably the most important aspects of the project—the literature searches and reviews—on the cheap, with smart but young and inexperienced researchers with no particular understanding of the highly technical topic of educational assessment.⁷
- The OECD did not follow standard meta-analysis protocols in structuring its literature searches and reviews. Therefore, it is not possible to know why they chose the research literature they chose and ignored the larger body of relevant research.
- Given that it is impossible to know, at least from reading what they have written, how they conducted their literature searches, it is easy to speculate that their searches were determined by opportunity, convenience, and professional biases (e.g., toward economics, away from psychology).

Six different writers drafted REAFISO's eight research summaries.⁸ One possible advantage to doing it this way might have been to diversify information sourcing. With several different individuals, working independently, the probability of a narrow review of the literature should have diminished.

Perusing the reference lists of the various reports, however, one can see that they largely ladled from the same soup pot. The same references and the same research groups appear frequently across different reports. Likewise, major source omissions, obvious to anyone deeply familiar with the research literature, are common across all.

The eight research summaries collectively contain more than 900 references. As is to be expected, many reference each other, other OECD documents, or the documents of closely related institutions, such as the Education Information Network of the European Commission (EURYDICE) and the United Nations Educational,

Scientific, and Cultural Organisation (UNESCO). I count 79 references of this type. For comparison, I count 144 references to CRESST documents and articles by CRESST-affiliated researchers.⁹ For contrast, I count no references to the substantial research literature that contradicts CRESST doctrine or disputes its evidence and methods.

By time period, the eight reports' references are: 19 pre-1990; 112 from 1991 to 2000; and 803 post-2000. For those references to journal articles, 206 lead to journals in education, 99 to economics journals, 29 to testing and measurement journals, and a paltry 16 to psychology journals.

Still, not all eight of the OECD staff and contractor reports are the same. They range in quality from sort of OK to just awful. For example, Stephanie Dufaux's review of upper-secondary level (i.e., high school) assessment programs and research suffers from some of the same biases as the others, relying too much on education and economics professors' work conducted only in the past decade, and neglecting older work and that conducted by psychologists (though with five, she includes more psychology references than any of her colleagues). Though still dominated by U.S. sources, she at least makes a concerted effort to widen her search geographically. Beyond the slants, however, she more or less just reports what she finds; she doesn't try to overreach with her conclusions.

“Student Standardised Testing: Current Practices in OECD Countries and a Literature Review,” *OECD Education Working Paper No. 65*

Dufaux's colleague Allison Morris is described as a master's graduate (from *Sciences Po*) with a specialty in human security and “research experience in the areas of microfinance, education in emergencies, and economic development” (Morris, p. 3). Her report “aims to synthesise the relevant empirical research on the impact of standardised testing on teaching and learning and to draw out lessons from the literature on aspects of standardised tests that are more effective in improving student outcomes.”

That goal well matches that for my meta-analyses and research summary described earlier (Phelps, 2012b). Of the several hundred studies I found to help answer the question, however, Morris includes exactly three in her review. Her report claims that “key debates concerning standardised testing are identified throughout,” but only one side in those debates seems to be represented in her review.

Disseminating Misinformation

The REAFISO writers cite as solid and unchallenged the conclusions from several studies that are misrepresentations of the evidence at best and frauds at worst:

Boaler (2002)

Jo Boaler conducts quasi-experimental studies comparing student performance among schools she refuses to identify with data she refuses to release. Despite the fact that students in her constructivist classrooms end up performing worse on all standardized tests administered to them, they perform better on a test she designed herself. So, she reasons, constructivist learning must be superior. SOURCES: Bishop, Clopton, and Milgram 2012; Bishop and Milgram 2012; Milgram 2012

Haney (2000)

Haney's famous study allegedly disproving the advertised success of Texas's "education miracle" in the 1990s contains an extraordinary number of misleading analyses: he sometimes uses different numbers than claimed; surreptitiously alters the definitions of common terms; frequently makes calculation errors; misrepresents data; misrepresents laws, procedures, and events; neglects to consider confounding factors; and sometimes just makes things up. I checked dozens of Haney's "evidence-based" assertions and found none that stood up to any scrutiny. Moreover, every one of his "mistakes" led in the same direction, strongly suggesting willfulness. In its number of factual misrepresentations, Haney's book-length study is the most substantial collection of research fraud I have ever studied. SOURCES: Grissmer, Flanagan, Kawata, and Williamson 2000; Phelps 2003, pp. 127-144, Toenjes, Dworkin, Lorence, and Hill 2000.

Hout and Elliot (2011)

Whereas all but a trivial amount of the great mass of relevant research is ignored, the work of U.S. National Research Council (NRC) study committee members is cited liberally. Daniel Koretz wins the prize for the most citations and references. Overall, forty-eight citations and forty references (of two hundred) go to committee members' work. More than thirty references cite CRESST work. The bulk of the rest cite the work of close friends and colleagues, or earlier NRC studies. At the same time, a who's who of the leading researchers in the field, past and present, goes missing.

Also, this NRC report advances its pet theory of "test-score inflation," while excluding the full abundance of counterevidence, thus recommending exactly the wrong policy to address a very serious and very topical problem. SOURCE: Phelps 2012c.

Klein, Hamilton, McCaffrey, and Stecher (2000)

In the "October Surprise" of the 2000 U.S. presidential campaign, these CRESST researchers debunk Texas' gains on the U.S. National Assessment of Educational Progress by disaggregating the data to the lowest level possible, then running a separate hypothesis test on each disaggregation. Their method is obviously invalid, as their conclusion is about Texas' gain scores on all the NAEP segments together—the "pooled" data. Overall, Texas' gains on the NAEP exceeded those of all other U.S. states but one in the 1990s. SOURCE: Phelps 2003, pp. 122-127

Koretz (2005a, 2005b, 2008)

J. J. Cannel's "Lake Wobegon Effect" studies showcased the causes of test-score inflation—educator dishonesty and conflicts of interest, lax security, and outdated norms. Koretz identified high stakes as the main culprit, even though all but one of Cannel's score-inflated tests were national norm-referenced monitoring (i.e., no-stakes) tests. Koretz cites a study he and CRESST colleagues conducted around 1990 in an unidentified school district, with unidentified tests, as evidence that high stakes cause test score inflation. But he controlled for none of the other factors—such as lax security—that could have explained the results. Nor, apparently, did the test genuinely have high-stakes. SOURCES: Fraker 1986/1987, Phelps 2008/2009b, 2010; Staradamskis 2008.

Linn (1998, 2000)

Linn further argued that the pre-post testing requirement (or, Title I Evaluation and Reporting System [TIERS]) of the Title I Compensatory Education [i.e., anti-poverty] program from the late 1970s on offered more evidence of the high-stakes-cause-test-score-inflation theory. His study had no controls, however, and the test involved did not carry any stakes—it was merely a reporting requirement (with no actual consequences). Linn argued that tests administered on a fall-spring schedule (presumably by the same teacher) averaged higher gain scores than those administered on a fall-fall schedule (presumably by different teachers). The average summer learning loss, as found in meta-analyses on the topic, however, entirely explains the difference in average test scores gains. SOURCES: Sinclair and Gutman 1992; Cooper, Nye, Charlton, Lindsay, and Greathouse, 1996; Phelps 2008/2009b, 2010.

Shepard (1989)

Shepard published a table that lists, allegedly, all the possible causes of the Lake Wobegon Effect that research to that date had suggested. Conspicuously absent from the table were Cannel's chief culprits: lax security and educator dishonesty and conflicts of interest. With these causal factors eliminated from consideration, she was free to attribute causation to high-stakes. SOURCES: Phelps 2003 (chapter 4), 2008/2009a,b, 2010, 2011a,b

Morris lists fifty-nine references, but visits and revisits certain individual sources repeatedly in the text. She cites five CRESST researchers ninety-one times.¹⁰ She cites a report from the CRESST satellite National Research Council Board on Testing and Assessment nine times.¹¹ Citations for the cuckolded group of economists allied with CRESST exceed fifty. One must squint to see how Morris synthesizes the relevant empirical research and identifies key debates when she cites a single, sparsely informative book chapter by Figlio and Loeb (2011) *thirty-six times*.

Among the more egregious of Morris's erroneous statements:

While being highly reliable and comparable, multiple choice questions can be limiting in that they do not test critical thinking or problem solving skills and it is argued such

questions encourage surface learning and rote recollection, rather than deep, cognitive processes. Rather than testing thinking skills, multiple choice or other close-ended questions test content only (p. 16). . . . [D]ue to the nature of a standardised test, the tests often cannot test for critical thinking, analytical or problem solving skills (p. 21).

This is dead wrong if a multitude of better-supported studies are to be believed (see, for example, Bridgeman 1991; Feinberg 1990; Rudman 1992; Traub 1993; Powers and Kaufman 2002; Goodman and Hambleton 2005; Roediger and Marsh 2005; Struyven, Dochy, Janssens, Schelfhout, and Gielen 2006), and had the author taken the time to peruse some of the many freely available retired tests online and read their items, she could have seen so for herself.

According to the literature, validity of large-scale, standardised tests—specifically those used to assess program effectiveness—is increased through matrix sampling. (p. 23)

"The literature" turns out to be a single source written by CRESST authors. The obvious question is: valid to whom? With most matrix-sample tests, no results are valid at the student, teacher, classroom, or school level, and so are responsibly not reported at those levels.

[P]lacing a "premium" on student test performance in the form of rewards or sanctions for teachers increases the risk of instruction being reduced to test preparation, which in turn limits the depth of the student experience and reduces the skill needed by teachers. Additionally, incentives such as bonuses can lead to strategic actions by teachers that distort or manipulate data. These include cases of teacher cheating, exclusion of students in assessments, and teaching to the test, all of which are reviewed in greater detail below. (p. 29)

Had REAFISO widened its literature search just a little, it might have learned: when teachers teach standards-based subject matter they are properly teaching to the test (as it is aligned with the standards); when they spend more than a smidgen of time drilling on test format they hurt, not help, their students' scores on the upcoming test; when they see in advance the specific content of an upcoming test, the problem is lax test security, not improper incentives. By the way, test developers know that drilling on test format does not work and discourage it (see, for example, Messick and Jungeblut 1981; DerSimonian and Laird 1983; Kulik, Bangert-Drowns, and Kulik 1984; Fraker 1986/1987; Whitla 1988; Snedecor 1989; Smyth 1990;

Becker 1990; Moore 1991; Powers 1993; Tuckman 1994; Tuckman and Trimble 1997; Powers and Rock 1999; Robb and Ercanbrack 1999; Camara 1999, 2008; Briggs 2001; Palmer 2002; Briggs and Hansen 2004; Crocker 2005; Roediger and Karpicke 2006a, 2006b; Allensworth, Correa, and Ponisciak 2008).

It is the researchers REAFISO has chosen to trust who broadcast the erroneous and destructive suggestion that it works.

Research from the United States has shown that if national tests are considered to be 'high stakes' for teachers and schools, teaching to the test can easily lead to an artificial over-inflation of results and thus render the results useless as a measure of real progress. (p. 37)

If CRESST researchers were correct that high-stakes caused test-score inflation, we should expect to find test-score inflation with all high-stakes tests, such as the hundreds of occupational licensure tests and U.S. college admission tests (e.g., SAT, ACT), but we do not. We do not because these tests are administered with high levels of security and frequent form and item rotation. The source of test-score inflation is lax test security, not high-stakes. (See, for example, Phelps 2010; Staradamskis 2008.)

“Using Student Test Results for Accountability and Improvement: A Literature Review,” *OECD Education Working Paper No. 54*

For his research review, Morten Anstorp Rosenkvist, the Norwegian civil servant on loan to the OECD for a few months in early 2010, read about student testing without stakes for students but sometimes stakes for teachers or schools. More so than the other REAFISO writers, Rosenkvist read surveys and opinion poll reports to better gauge the attitudes and preferences of non-researchers toward testing. This interested me personally since I have been studying the same for a couple decades, ultimately publishing a meta-analysis of 813 individual item-response group combinations from 247 program evaluation surveys and opinion polls conducted between 1960 and 2010 (Phelps 2012b).

Alas, 'tis a pity, Rosenkvist did not happen upon my work. He mentions the results of three surveys of local public officials (p. 16), three of school administrators (p. 16), six of teachers (p. 18), seven of parents (pp. 19–20), and three of students (p. 20). According to

Rosenkvist (p. 20), "Students generally dislike high stakes assessments." But my meta-analysis counted twenty student surveys regarding high-stakes tests that accumulated an average, rather large effect size of +1.03. For the other response groups, Rosenkvist concludes generally positive support for high-stakes testing, roughly matching the results from my meta-analysis. Whereas he bases his conclusions on three, three, six, and seven cases, however, mine emerge from seven, thirty-four, eighty-five, and fifty-five, as well as several from university faculty and hundreds from the general public, and average effect sizes can be calculated precisely for each group. All this information was available to Rosenkvist had he asked for it (Phelps 2005b, 2012b).

I felt similar frustration reading Rosenkvist fumbling around with summarizing the research on the effect of high-stakes tests on student achievement (pp. 22–24) and the cost of assessment (p. 27). He would have encountered far more evidence, and reached more reliable conclusions, had he been willing to search outside the CRESST-U.S. think tank bubble.

A wider search might have smoothed out the inevitable contradictions, too. At several points, Rosenkvist encourages readers always to use a variety of measures and multiple tests, because no one test can be perfect or cover a domain of interest sufficiently. Yet he also recommends Daniel Koretz's method of judging the validity of one test score trend by comparing it with that of another, even with completely different topical coverage. In his final concluding paragraph (p. 35), Rosenkvist asserts, reasonably, that "student test results must be reliable, valid and fair." Then, in the next sentence, he recommends that "several assessments should be used to measure [each] student outcome." Good luck with that.

"Assessment and Innovation in Education," *OECD Education Working Paper No. 24*

The worst REAFISO research in most respects, however, is that conducted by the one American, Janet W. Looney (2009, 2011, 2013), the most prone to overreach. Unfortunately, whereas her colleagues wrote one report each, Looney wrote three. Looney is such an ardent, devout missionary for CRESST doctrine, she could be mistaken as their publicist.¹² Using the same comparison metric as before, Looney's references include twenty-two to OECD and related sources, and more than three times as many—a whopping seventy—references to

CRESST and CREST-affiliate sources (and zero references to sources that dispute CRESST's evidence or methods).

While only a small number of studies have focused on the validity of test score gains in high-stakes testing, they have usually found evidence of score inflation. (p. 30)

This statement is wrong on both counts. There exist a large number of studies, and they did not find score inflation where test security was tight and form and item rotation frequent. (See, for example, Becker 1990; Moore 1991; Tuckman 1994; Tuckman and Trimble 1997; Powers and Rock 1999; Briggs 2001; Palmer 2002; Crocker 2005; Roediger and Karpicke 2006a, 2006b; Camara 2008.) The studies conducted by Koretz and Linn that, allegedly, found evidence of score inflation involved low-stakes tests (erroneously labeled as high-stakes). Neither study employed any controls. Linn's result is easily explained away by the summer-learning-loss effect. Koretz's study remains shrouded in secrecy two decades later, with a still-unidentified school district with still-unidentified tests. (See, for example, Phelps 2010; Staradamskis 2008.)

As mentioned earlier, Looney aggressively promotes innovation—specifically the type espoused by U.S. radical constructivists and CRESST researchers. It seems not to matter that the public has expressed little interest, or that the programs have failed miserably when tried. She writes (p. 17):

Most new programmes experience an "implementation dip"—that is, student performance gets worse before improving. Improvements in student achievement may take as long as five years in primary schools, and longer in secondary schools. Teachers working in innovative programmes will need extra support to understand where they may need to adjust practices.

Looney advocates sticking with programs she favors even if the evidence of their effectiveness is negative for five years (with primary graders) or even longer (with older students). But experience shows that most innovations fail; certainly they do not succeed simply for the fact that they are innovative. We should rest steadfast through five, six, seven, or more years of negative results before even considering pulling the plug on a program? And what of the children whose education has been stunted in the meantime? Do they matter?¹³

Continuing, Looney writes:

Teachers may find it impossible to balance the pressures of implementing new and innovative programs and high-stakes tests." (p. 18) Moreover, tight alignment . . . tends to undermine innovative programmes. . . . (p. 20)

The implication is, of course, that it is the standards and tests that must be at fault. They should either be pushed aside to make way for the innovations, or be radically reconstructed to fit the new order.

Summary

My judgment of REAFISO's efforts should be apparent at this point. But REAFISO's efforts should be judged unfavorably even by its own standards. In the *Design and Implementation Plan for the Review* (OECD, 2009), REAFISO promised to, among other goals:

. . . extend and add value to the existing body of international work on evaluation and assessment policies. (p. 5)

Synthesise research-based evidence on the impact of evaluation and assessment strategies and disseminate this knowledge among countries. Identify policy options for policy makers to consider. (p. 4)

. . . take stock of the existing knowledge base within the OECD and member countries as well as academic research on the relationship between assessment and evaluation procedures and performance of students, teachers and schools. It will look at the quantitative and qualitative evidence available on the different approaches used to evaluate and assess educational practice and performance. (p. 16)

To the contrary, REAFISO has not synthesized the existing body of research-based evidence on evaluation and assessment policies, much less extended it. By telling the world that a small proportion of the existing body of research is all that exists, they have instead hidden from the world most of the useful and relevant information (or implied that it is not worth considering).

The ordinary Citizen Joe knows that one shouldn't trust everything one finds on the Internet, nor assume that Internet search engines rank documents according to their accuracy. So naturally, scholarly researchers who are trained to be skeptical, systematic, thorough, aware of biases, and facile with statistical sampling methods would be too. After all, scholarly researchers have spent several

more years in school, often prestigious schools. They should “know how to know” as well or better than the average citizen.

Yet REAFISO’s reviews repeatedly offer one or a few examples of research from their favored sources to summarize topics, even though thorough reviews of dozens, hundreds, or thousands of sources were to be found had they simply looked widely enough. In some cases, REAFISO writers conclude a policy recommendation on the basis of one or a few studies, when a reading of the whole of the research literature on the topic would suggest exactly the opposite policy.

In its document, *Evaluation and Assessment Frameworks for Improving School Outcomes: Common Policy Challenges* (2011), written two years after the *Design and Implementation Plan*, REAFISO claims to have completed “a thorough analysis of the evidence on evaluation and assessment.”

Ironies

Ideas matter, as do their censorship and suppression. For all their hawking of recipes from the radical constructivist cookbook—promoting “critical” and “higher-order” thinking, discovery learning, and innovation for its own sake —the REAFISO writers neither construct their own knowledge, discover knowledge with unique learning styles, nor evaluate what they read critically. The six writers read and came to believe the same, unquestioningly parroting a single dogma.

For all of REAFISO’s adulation of innovation, there’s little to be found in their reviews. Apparently, they started with the most accessible and heavily promoted reports from the U.S., CRESST, and CRESST’s cuckolded cabal of economists, and followed their well-worn path like a workhorse with side blinders. REAFISO’s writers looked where they were told to look and conspicuously avoided looking in any of the directions they were not told to.

REAFISO criticizes traditional tests for “narrowing the curriculum” and “teaching to the test.” But REAFISO itself narrowed its focus in the relevant research literature to a tiny aperture, reducing its search to that seen within the perimeter. Then, over several hundred pages, REAFISO repeatedly, relentlessly drills its confirmation bias into its readers.

As a result, the OECD now recommends to all its members the wisdom of U.S. education research, certainly the world’s least effective and perhaps the world’s most corrupt—responsible for producing

one of the world's least successful education systems (as measured by outcomes over inputs). U.S. students continue to underperform on the OECD's own PISA assessment, despite U.S. taxpayers spending more on education per capita than all but a few other countries. The U.S. public and politicians see their education system in a perpetual state of crisis, as having largely failed. The OECD now suggests the rest of the world copy it.

In 2011, REAFISO wrote:

The effectiveness of evaluation and assessment relies to a great extent on ensuring that both those who design and undertake evaluation activities as well as those who use their results possess the proper skills and competencies. This is crucial to provide the necessary legitimacy to those responsible for evaluation and assessment.

If only they had practiced what they preach.

References

- Allensworth E., Correa, M., Ponisciak, S. (2008, May). *From High School to the Future: ACT Preparation—Too Much, Too Late: Why ACT Scores Are Low in Chicago and What It Means for Schools*. Chicago: Consortium on Chicago School Research at the University of Chicago.
- Becker, B. J. (1990, Fall). Coaching for the Scholastic Aptitude Test: Further synthesis and appraisal. *Review of Educational Research*, 60(3), 373–417.
- Bishop, W., Clopton, P., and Milgram, R. J. (2012). A Close Examination of Jo Boaler's Railside Report. *Nonpartisan Education Review / Articles*, 8(1). <http://www.nonpartisaneducation.org/Review/Articles/v8n1.pdf>
- Bishop, W., and Milgram, R. J. (2012). A Response to Some of the Points of: "When Academic Disagreement Becomes Harassment and Persecution". *Nonpartisan Education Review / Essays*, 8(4). <http://www.nonpartisaneducation.org/Review/Essays/v8n4.htm>
- Bridgeman, B. (1991, June). Essays and multiple-choice tests as predictors of college freshman GPA. *Research in Higher Education*, 32(2), 319–332.
- Briggs, D. C. (2001, Winter). The effect of admissions test preparation. *Chance*.
- Briggs, D., and Hansen, B. (2004, May). Evaluating SAT test preparation: Gains, effects, and self-selection. Paper presented at the Educational Testing Service, Princeton, N.J.

- Boaler, J. (2002). *Experiencing School Mathematics: Traditional and Reform Approaches to Teaching and their Impact on Student Learning*, Lawrence Erlbaum Associates, Mahwah, N.J.
- Camara, W. (1999, Fall). Is commercial coaching for the SAT I worth the money? *College Counseling Connections* 1(1). New York: The College Board.
- . (2008). College admission testing: Myths and realities. In R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing*, Washington, D.C.: American Psychological Association.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., and Greathouse, S. (1996, Fall). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*. 66(3).
- Crocker, L. (2005). Teaching FOR the test: How and why test preparation is appropriate. In R. P. Phelps (Ed.), *Defending Standardized Testing* (pp. 159–174). Mahwah, N.J.: Lawrence Erlbaum.
- DerSimonian and Laird. (1983). Evaluating the effect of coaching on SAT scores: A meta-analysis, *Harvard Educational Review* 53, 1-5.
- Dufaux, S. (2012), "Assessment for Qualification and Certification in Upper Secondary Education: A Review of Country Practices and Research Evidence," *OECD Education Working Papers*, No. 83, OECD Publishing: Paris. <http://dx.doi.org/10.1787/5k92zp1cshvb-en>
- Faubert, V. (2009). "School Evaluation: Current Practices in OECD Countries and a Literature Review". *OECD Education Working Papers*, No. 42, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&cote=edu/wkp\(2009\)21](http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&cote=edu/wkp(2009)21)
- Feinberg, L. (1990, Fall). Multiple-choice and its critics: Are the 'alternatives' any better? *The College Board Review*, 157, 13–17, 30–31.
- Figlio, D., and Loeb, S. (2011). "School Accountability," in E. Hanushek, S. Machin and L. Woessman (eds.), *Handbooks in Economics*, Vol. 3, North-Holland, The Netherlands, pp. 383–421.
- Fraker, G. A. (1986–1987, Winter). The Princeton Review reviewed. *The Newsletter*. Deerfield, Mass.: Deerfield Academy, Winter.
- Goodman, D., and Hambleton, R. K. (2005). Some misconceptions about large-scale educational assessments. In R. P. Phelps, (Ed.), *Defending standardized testing* (pp. 91–110). Mahwah, N.J.: Lawrence Erlbaum.
- Grissmer, D. W., Flanagan, A., Kawata, J., Williamson, S. (2000, July). *Improving Student Achievement: What NAEP State Test Scores Tell Us*, Rand Corporation.
- Haney, W. (2000). The Myth of the Texas Miracle in Education, *Education Policy Analysis Archives*, 8(41). <http://epaa.asu.edu/ojs/article/view/432>

- Hout, M., and Elliott, S. (eds.) (2011). *Incentives and Test-Based Accountability in Education*, National Research Council, The National Academies Press, Washington, D.C. <http://www.nap.edu/catalog.php?record id=12521>
- Herring, M. Y. (2001, April). Ten reasons why the Internet is no substitute for a library. *American Libraries*.
- Isoré, M. (2009). "Teacher Evaluation: Current Practices in OECD Countries and a Literature Review". *OECD Education Working Papers*, No. 23, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&cote=edu/wkp\(2009\)2](http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&cote=edu/wkp(2009)2)
- Klein, S., Hamilton, L., McCaffrey, D., and Stecher, B. (2000). *What do test scores in Texas tell us?* Issue paper, Rand Education.
- Koretz, D. (2005a). *Alignment, High Stakes, and the Inflation of Test Scores*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.
- . (2005b). *Alignment, High Stakes, and the Inflation of Test Scores*, in J. Herman and E. Haertel (eds.), *Uses and Misuses of Data in Accountability Testing*, *Yearbook of the National Society for the Study of Education*, 104, Part 2, Malden, MA: Blackwell Publishing.
- . (2008). *Measuring up: What Educational Testing Really Tells Us*, Cambridge, Mass.: Harvard University Press.
- Kulik, J. A., Bangert-Drowns, R. L., and Kulik, C-L. C. (1984). Effectiveness of coaching for aptitude tests, *Psychological Bulletin* 95, 179-188.
- Linn, R. (1998). *Assessments and Accountability*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.
- Linn, R. L. (2000). Assessments and Accountability, *Educational Researcher*, 29, pp. 4– 16.
- Looney, J. W. (2009). "Assessment and Innovation in Education." *OECD Education Working Papers*, No. 24, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp\(2011\)9&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp(2011)9&doclanguage=en)
- . (2011). "Integrating Formative and Summative Assessment: Progress Toward a Seamless System?" *OECD Education Working Papers*, No. 58, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp\(2011\)4&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp(2011)4&doclanguage=en)
- . (2013). "Alignment in Complex Education Systems: Achieving Balance and Coherence". *OECD Education Working Papers*, No. 64, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp\(2011\)9&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp(2011)9&doclanguage=en)

- Messick, S., and A. Jungeblut (1981). Time and method in coaching for the SAT, *Psychological Bulletin* 89, 191–216.
- Milgram, R. J. (2012). Private Data—The Real Story: A Huge Problem with Education Research, *Nonpartisan Education Review / Essays*, 8(5). <http://www.nonpartisaneducation.org/Review/Essays/v8n5.htm>
- Moore, W. P. (1991). Relationships among teacher test performance pressures, perceived testing benefits, test preparation strategies, and student test performance. Ph.D. dissertation, University of Kansas, Lawrence.
- Morris, A. (2011). "Student Standardised Testing: Current Practices in OECD Countries and a Literature Review." *OECD Education Working Papers*, No. 65, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocument/?cote=EDU/WKP\(2011\)10&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocument/?cote=EDU/WKP(2011)10&doclanguage=en)
- OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. (2009, October). *Design and Implementation Plan for the Review*. Education Policy Committee, OECD Directorate for Education, Organisation for Cooperation and Development, Paris. <http://www.oecd.org/edu/preschoolandschool/44568070.pdf>
- OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. (2011). *Common Policy Challenges*. Education Policy Committee, Directorate for Education, Organisation for Co-operation and Development, Paris. <http://www.oecd.org/edu/preschoolandschool/46927511.pdf>
- Palmer, J. S. (2002). Performance incentives, teachers, and students: Ph.D. dissertation. Columbus: The Ohio State University.
- Phelps, R. P. (2003). *Kill the messenger*. New Brunswick, N.J.: Transaction Publishers.
- . (2005a, February). Educational testing policy: Stuck between two political parties, *Yale Politic*. <http://www.nonpartisaneducation.org/Foundation/YalePoliticArticle.htm>
- . (2005b). Persistently positive. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 1–22). Mahwah, N.J.: Lawrence Erlbaum.
- . (2007a, Summer). The dissolution of education knowledge. *Educational Horizons*. 85(4), 232–247. <http://www.nonpartisaneducation.org/Foundation/DissolutionOfKnowledge.pdf>
- . (2007b). *Standardized testing primer*. New York: Peter Lang.
- . (2008/2009a). Educational achievement testing: Critiques and rebuttals. In R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing*, Washington, D.C.: American Psychological Association.

- . (2008/2009b). The rocky score-line of Lake Wobegon. Appendix C in R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing*, Washington, D.C.: American Psychological Association. <http://supp.apa.org/books/Correcting-Fallacies/appendix-c.pdf>
- . (2010, July). The source of Lake Wobegon [updated]. *Nonpartisan Education Review / Articles*, 6(3). Available at: <http://www.nonpartisaneducation.org/Review/Articles/v6n3.htm> <http://www.nonpartisaneducation.org/Review/Articles/v6n3.htm> <http://www.nonpartisaneducation.org/Review/Articles/v6n3.htm>
- . (2011a). Extended Comments on the Draft Standards for Educational and Psychological Testing (But, in particular, Draft Chapters 9, 12, and 13). *Nonpartisan Education Review/Essays*, 7(3). Available at: <http://www.nonpartisaneducation.org/Review/Essays/v7n3.htm>
- . (2011b). Educators cheating on tests is nothing new; Doing something about it would be. *Nonpartisan Education Review/Essays*, 7(5). Available at: <http://www.nonpartisaneducation.org/Review/Essays/v7n5.htm>
- . (2012a). Dismissive reviews: Academe's Memory Hole. *Academic Questions*, Summer. <http://www.nas.org/articles/d dismissive reviews academes memory hole>
- . (2012b). The effect of testing on student achievement, 1910–2010, *International Journal of Testing*, 12(1), 21–43, International Test Commission. <http://www.tandfonline.com/doi/abs/10.1080/15305058.2011.602920#preview>
- . (2012c, Summer). The Rot Festers: Another National Research Council Report on Testing, *New Educational Foundations*, 1(1), pp. 30–52. <http://www.newfoundations.com/NEFpubs/NewEduFdnsv1n1Announce.html>
- Powers, D. E. (1993). Coaching for the SAT: A summary of the summaries and an update. *Educational Measurement: Issues and Practice*, 39, 24–30.
- Powers, D. E., and Rock, D. A. (1999). Effects of coaching on SAT I: Reasoning test scores. *Journal of Educational Measurement*, 36(2), 93–118.
- Powers, D. E., and Kaufman, J. C. (2002). Do standardized multiple-choice tests penalize deep-thinking or creative students? *Research Report RR-02-15*. Princeton, N.J.: ETS.
- Robb, T.N., and Ercanbrack, J. (1999, January). A study of the effect of direct test preparation on the TOEIC scores of Japanese university students. *Teaching English as a Second or Foreign Language*, 3(4).
- Roediger, H. L., and Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31, 1155–1159.

- Roediger, H. L. and Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210.
- . (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255.
- Rosenkvist, M. A. (2010). "Using Student Test Results for Accountability and Improvement: A Literature Review". *OECD Education Working Papers*, No. 54, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocument/?cote=EDU/WKP\(2010\)17&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocument/?cote=EDU/WKP(2010)17&doclanguage=en)
- Rudman, H. C. (1992, Fall). *Testing for learning* (book review). *Educational Measurement: Issues and Practice*, 31–32.
- Shepard, L. A. (1989, March). Inflated test score gains: Is it old norms or teaching the test? Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Sinclair, B. and Gutman, B. (1992). *A Summary of State Chapter 1 Participation and Achievement Information for 1989–90*. Prepared for the U.S. Department of Education, Office of Policy and Planning, 1992.
- Smyth, F. L. (1990). SAT coaching: What really happens to scores and how we are led to expect more. *The Journal of College Admissions*, 129, 7–16.
- Snedecor, P. J. (1989). Coaching: Does it pay—revisited. *The Journal of College Admissions*, 125, 15–18.
- Staradamskis, P. (2008, Fall). Measuring up: What educational testing really tells us. Book review, *Educational Horizons*, 87(1). Available at: <http://www.nonpartisaneducation.org/Foundation/KoretzReview.htm>
- Stevens-Rayburn, S. (1998). If it's not on the Web, it doesn't exist at all: Electronic information resources—myth and reality. *Library and Information Services in Astronomy III, ASP Conference Series V*, 153.
- Struyven, K., Dochy, F., Janssens, S., Schelfhout, W., and Gielen, S. (2006). The overall effects of end-of-course assessment on student performance: A comparison between multiple choice testing, peer assessment, case-based assessment and portfolio assessment. *Studies in Educational Evaluation*, 32, 202–222.
- Toenjes, L. A., Dworkin, A. G., Lorence, J., and Hill, A. N. (2000, August). *The Lone Star Gamble: High Stakes Testing, Accountability, and Student Achievement in Texas and Houston*. Department of Sociology, University of Houston.
- Traub, R. E. (1993). On the equivalence of the traits assessed by multiple choice and constructed-response tests. In R. E. Bennett and W. C. Ward (Eds.), *Construction versus choice in cognitive measurement*. Hillsdale, N.J.: Lawrence Erlbaum.

- Tuckman, B. W. (1994, April). Comparing incentive motivation to meta-cognitive strategy in its effect on achievement, Paper presented at the annual meeting of the American Educational Research Association, New Orleans (ERIC ED368790).
- Tuckman, B. W. and Trimble, S. (1997, August). Using tests as a performance incentive to motivate eighth-graders to study, Paper presented at the annual meeting of the American Psychological Association, Chicago. (ERIC: ED418785).
- Whitla, D. K. (1988). Coaching: Does it pay? Not for Harvard students. *The College Board Review*. 148, 32–35.

Notes

1. The OECD's English-language publications use British, rather than American, spellings.
2. The OECD is on the Web at www.oecd.org www.oecd.org www.oecd.org and REAFISO, <http://www.oecd.org/edu/preschoolandschool/oecdreviewo-nevaluationandassessmentframeworksforimproving-schooloutcomes.htm>.
3. By the turn of the century, the World Bank had conducted seemingly thousands of "returns to education" studies that calculated the long-term economic effect of varying amounts of time spent in school. As we all know, however, time spent in school can vary quite a lot in efficacy. To know what works inside the schoolhouse, more information was needed.
4. The United Nations Educational, Scientific, and Cultural Organization (UNESCO) enables cooperation among member countries' education ministries. Economists, and economic analysis, have never occupied more than a small proportion of its activities.
5. See project home page home page home page home page.
6. Names missing from all the REAFISO reports include: Roddy Roediger, Frank Schmidt, W.J. Haynie, Harold Wenglinsky, Linda Winfield, C. C. Ross, E. H. Jones, Mike McDaniel, Lorin Anderson, J. R. Nation, J. H. Block, Carol Parke, S. F. Stager, Arlen Gullickson, Lynn Fuchs, Douglas Fuchs, Kathy Green, Max Eckstein, Harold Noah, Jeffrey Karpicke, Michael Beck, Stephen Heynemann, William D. Schafer, Francine Hultgren, Willis Hawley, James H. McMillan, Elizabeth Marsh, Susan Brookhart, Gene Bottoms, Gordon Cawelti, Mike Smoker, David Grissmer, Arthur Powell, Harold Stevenson, Hunter Boylan, Elana Shohamy, Aletta Grisay, Chris Whetton, Steve Ferrara, Glynn Ligon, Micheline Perrin, Thomas Fischer, A. Graham Down, Nigel Brooke, John Oxenham, Arthur Hughes, D. Pennycuick, John Poggio, Anthony Somerset, John O. Anderson, Noel McGinn, Anne Anastasi, Nick Theobald, David Miller, Nancy Protheroe, Floraline Stevens, Ted Britton,

Senta Raizen, Thomas Corcoran, Clement Stone, Frank Dempster, and state agencies in Massachusetts, Florida, and South Carolina.

Those are just names of some folks who have conducted one or more individual studies. Others have summarized batches of several to many studies in meta-analyses or literature reviews, for example (in chronological order): Panlasigui (1928); Ross (1942); Kirkland (1971); Proger and Mann (1973); Jones (1974); Bjork (1975); Peckham and Roe (1977); Wildemuth (1977); Jackson and Battiste (1978); Kulik, Kulik, Bangert-Drowns, and Schwalb (1983–1991); Natriello and Dornbusch (1984); Dawson and Dawson (1985); Levine (1985); Resnick and Resnick (1985); Guskey and Gates (1986); Hembree (1987); Bangert-Drowns, Kulik, and Kulik (1991); Dempster (1991); Adams and Chapman (2002); Locke and Latham (2002); Roediger and Karpicke (2006); and Basol and Johanson (2009).

7. Janet W. Looney, responsible for three reports, has a master's degree in public administration (U. Washington) and apparently once worked at the OECD. She now works as a freelance writer and editor. Morten Anstorp Rosenkvist worked for a few months at the OECD on *secondment* (essentially, temporary loan) from the Norwegian Ministry of Education and Research. Aside from his OECD report, he has written "Mobility of Teachers across Education Sectors in Norway," but I could find nothing else about him on the Web or in the Norwegian Ministry of Education website. Allison Morris, Violaine Faubert, and Marlène Isoré are identified by the OECD as graduate students at the *Institut d'Etudes Politiques de Paris* (*Sciences Po*). Two of the three have apparently written economics papers on financial accounting and international trade, but I could find nothing else about them through Web searches. Stephanie Dufaux works at the OECD as a "Carlo Schmid Fellow," and that's all I could find on her.

8. Dufaux, S. (2012), "Assessment for Qualification and Certification in Upper Secondary Education: A Review of Country Practices and Research Evidence," *OECD Education Working Papers*, No. 83, OECD Publishing: Paris. <http://dx.doi.org/10.1787/5k92zplcshvb-en>

Faubert, V. (2009). "School Evaluation: Current Practices in OECD Countries and a Literature Review." *OECD Education Working Papers*, No. 42, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&cote=edu/wkp\(2009\)21](http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&cote=edu/wkp(2009)21)

Isoré, M. (2009). "Teacher Evaluation: Current Practices in OECD Countries and a Literature Review". *OECD Education Working Papers*, No. 23, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&cote=edu/wkp\(2009\)2](http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&cote=edu/wkp(2009)2)

Looney, J. W. (2009). "Assessment and Innovation in Education". *OECD Education Working Papers*, No. 24, OECD Publishing: Paris.

[http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp\(2011\)9&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp(2011)9&doclanguage=en)

Looney, J. W. (2011). "Integrating Formative and Summative Assessment: Progress Toward a Seamless System?" *OECD Education Working Papers*, No. 58, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp\(2011\)4&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp(2011)4&doclanguage=en)

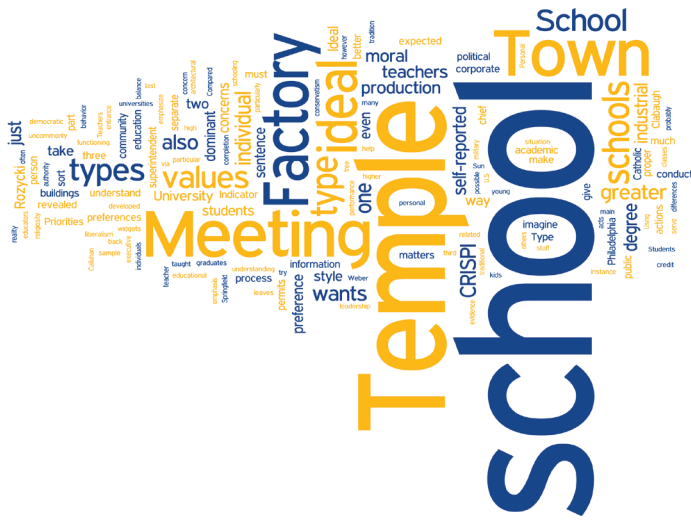
Looney, J. W. (2013). "Alignment in Complex Education Systems: Achieving Balance and Coherence." *OECD Education Working Papers*, No. 64, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp\(2011\)9&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/wkp(2011)9&doclanguage=en)

Morris, A. (2011). "Student Standardised Testing: Current Practices in OECD Countries and a Literature Review." *OECD Education Working Papers*, No. 65, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocument/?cote=EDU/WKP\(2011\)10&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocument/?cote=EDU/WKP(2011)10&doclanguage=en)

Rosenkvist, M. A. (2010). "Using Student Test Results for Accountability and Improvement: A Literature Review." *OECD Education Working Papers*, No. 54, OECD Publishing: Paris. [http://www.oecd.org/officialdocuments/displaydocument/?cote=EDU/WKP\(2010\)17&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocument/?cote=EDU/WKP(2010)17&doclanguage=en)

9. Included in the CRESST counts are those for the US National Research Council's Board on Testing and Assessment which has, essentially, been run as a CRESST satellite since the late 1980s (see Phelps, 2012c).
10. CRESST researcher (number of citations): Laura Hamilton (32); Brian Stecher (17); Stephen Klein (17); Daniel Koretz (15); Robert Linn (7).
11. Hout and Elliot, 2011.
12. The Center for Research in Education Standards and Student Testing, a federally funded consortium including the Rand Corporation, and the Schools of Education at UCLA, U. Colorado, and U. Pittsburgh (and formerly Arizona State U., as well).
13. Looney is a graduate of a master's in public administration program. Presumably, she studied project planning, discounting, returns on investment, and the like. She should know that programs guaranteed to have negative returns for five or more years seldom pay off. Moreover, she says nothing about the size of the assumed future payoff of these innovative programs.

Richard P. Phelps is a member of the *New Educational Foundations* editorial board, the founder of the *Nonpartisan Education Review* (nonpartisaneducation.org) and co-author and editor of *Correcting Fallacies about Educational and Psychological Testing* (APA Books, 2008/9).



What We Want From Schools

(And Why We Are Unlikely to Get It)

by Gary K. Clabaugh

AMERICANS SELDOM AGREE in depth on much of anything. Moreover, what any particular American wants changes with their circumstances. Such chronic dissensus and persistent unsettledness leaves educators trying to straddle a too-tall fence. Not only must they figure out who wants what; they must also try to **balance** mutually exclusive demands. Typically they try to satisfy the loudest complainer and hope for the best. Happily, there is a better way. Using Max Weber's (1864–1920) "ideal types" we can more clearly understand what is going on and better determine what to do.

Weber developed ideal types as an analytic tool to better understand complicated social situations. They are not meant to be ideal in the sense that they are perfect or even admirable. In fact, they involve deliberate simplification and amplification of particular aspects of—in this case—schooling. But they do so in a way that

permits an uncommonly concise and logically coherent understanding of that reality.

Three Ideal School Types

When we review history and present conditions, three ideal school types seem to emerge: the school as Temple, the school as Factory, and the school as Town Meeting. No school embodies just one of these types. Every school is an edgy and fluctuating mixture of all three. In most schools, however, one type dominates, another is secondary, and the third is tertiary.

The School as Temple Ideal Type

The school as Temple dates all the way back some six thousand years to the primal beginning of schools. These first ventures were conducted by priests and actually housed in temples.

With the Temple ideal type the school constitutes a sort of moral community—a temple of learning. Its distinctive functions are maintaining values, shaping “proper” behavior, and upholding tradition. In addition to functioning as an executive, the chief administrator is expected to serve as a moral leader—a sort of high priest. Teachers are front-line clergy fulfilling a special calling. Students are initiates being inducted into a consecrated community.



At the United States Military Academy the school as Temple plays a primary role. Photo credit: U.S. Army

School rules are heavily invested with moral authority. Success is acceptance as a “properly educated” person, a kind of character formation. Infractions are moral evils, a kind of sin. Teachers are expected to be moral exemplars. Their public character, behavior, and demeanor are expected to reflect, if sometimes hazily, the school’s professed ideals. That is why the recently revealed priestly pedophilia in some Roman Catholic schools has proved particularly troubling.

The architectural style of Temple dominant schools often resembles churches or cathedrals. The academic use of Latin phraseology, quasi-religious ceremonies such as commencement, and academic costumes resembling vestments also evokes the Temple. Even the names of buildings or institutions can convey a religious image. The University of Pittsburgh’s Cathedral of Learning and Temple University are examples.

Evidence for the school as Temple is widespread. Reporting on the qualities resident’s are looking for in a new school superintendent, for instance, the *Springfield Sun*, a suburban Philadelphia newspaper, quotes a resident as saying she wants “someone who pays close attention to the sometimes intangible qualities—morality and the ability to know what is right and what is wrong.”¹ In short, she wants a high priest for the school as Temple.

The School as Factory Ideal Type

Compared to the Temple, the school as Factory is a much more recent development, dating back only to the beginning of the industrial age. Here a school operates essentially as an industrial plant that transforms a conglomeration of youngsters into tolerable young adults much as widgets are manufactured.² Factory-dominant institutions of higher education, such as large state universities, take some of these young adults and make them even more finished products.

Like the Temple, the school as Factory permits no questioning of basic authority. But while the main concerns of the Temple are proper conduct and tradition, the chief concerns of the school as Factory are cost effectiveness and efficient production. That’s how, in higher education, we can end up with freshmen classes of 200 or more, many classes taught by adjuncts, distance education, and even for-profit universities that spend more on advertising than faculty.

Goals and production quotas in the school as Factory are set at the top. An infraction is whatever interferes with production. The superintendent or university president acts as a CEO serving a

corporate board. Principals and deans work as foremen supervising the workers (teachers) manning the production line. Students are containers to be filled with information and skills, quality tested, graded, labeled and shipped as efficiently as possible. In a separate context, they also are customers, with all that this entails.

Older school buildings often bear a striking architectural resemblance to the industrial buildings of yesteryear, while modern Factory style schools can easily disappear into the corporate sameness of industrial or corporate office parks.

I once taught in an older industrial style high school. The front entrance was inscribed with "Enter to Learn," and the main exit with "Leave to Serve." I used to imagine the little darlings shuffling in the entrance, being transformed into cheerful robots, then marching mechanically out the exit to serve their masters.



This Philadelphia school as Factory looks the part. One can imagine it manufacturing widgets.

Remember Raymond Callahan's classic *Education and the Cult of Efficiency*?³ Published in the 1960s, Callahan's book offers compelling everyday evidence—talks given at educator's confabs, excerpts from school leadership texts, and teacher correspondence—of how the business/industrial model captured public schooling in the early twentieth century—and has yet to give it up.

As with the Temple, we can find evidence of the school as factory in the media. In that previously quoted *Springfield Sun* article, for instance, we find a citizen who wants a superintendent “with a proven track record whose style of management eliminates a staff’s tendency to become pathetic and unwilling to take responsibility for their own actions, or lack of performance.” She adds, “The staff should have no built-in excuse for a sub-standard performance.” School as Factory, anyone?

The School as Town Meeting Ideal Type

Our third and last ideal type, the school as Town Meeting, is a political marketplace, as economists understand that term. What matters most in this type school are connections, access to information and resources, *skillful* bargaining, and sufficient civility to prevent outright conflicts. Individual differences matter too—particularly if those differences coalesce into constituencies of concern.



Stressing values and proper conduct, the traditional one-room school favored the Temple; but there also was an emphasis on the school as Town Meeting. Photo credit: G. K. Clabaugh

Except for discourteousness, dogmatism, and unwillingness to compromise, there are few infractions, only occasions for renegotiation.⁴ Competing concerns are addressed via compromise and give-and-take. Appeals to morality or efficiency are only one part of this process.

Besides performing routine administrative tasks, the chief executive of a Town Meeting dominant school acts as a sort of mediator or referee, balancing the competing concerns of individual teachers, parents, students, and others. At the same time he or she represents the school to outsiders.

Schools based on personal freedom and democratic principles, such as A. S. Neal's famed Summerhill, emphasize the school as Town Meeting. So do many Society of Friends schools and a growing number of U.S. "free schools" such as those found in Albany, N.Y., and Philadelphia. A key emphasis in these schools is respect for the individual as an individual and participation by all in decision-making.

That *Springfield Sun* article quotes a parent as saying she wants a district superintendent who will "recognize students as individuals that require undivided attention." Another wants a "team builder" who "reacts positively constructively to input from all parties and who will help the community feel that it is a functioning part of the educational process. Both unknowingly favor the school as Town Meeting.

Preferences Shift as Personal Circumstances Change

In the abstract, individuals typically prefer one ideal type to the others; but we should not imagine that any individual's preferences are unchanging. People's inclinations vary with their circumstances. Imagine a teacher in a school, for example, who wants to take her kids on a school-is-in-session vacation without academic penalty. In this circumstance her preference for the Factory is likely to give way to the individual consideration characteristic of the Town Meeting. Such shifting personal preferences make educator's jobs even more difficult.

The Clabaugh / Rozycki School Priorities Indicator

This question arose while developing these ideal types. Is the Town Meeting a separate category, or just a variety of the school as Temple? The Town Meeting is, after all, focused on values—a hallmark of the Temple. Yes, the Town Meeting's values are very different from those of, say, a military academy or old-fashioned Catholic school, both of which are Temple dominant. Still, the Town Meeting's focus is on values.

To investigate this matter and to preliminarily examine how ideal type school preferences differ from person to person and situation to situation, my colleague Edward Rozycki and I developed the CRiSPI.⁵ This instrument uses a forced-choice, sentence completion format to measure an individual's school related priorities.

Here are five sample items of the original thirty employed in the CRiSPI. The ideal type associated with each sentence completion is indicated: *T* = Temple, *F* = Factory and *TM* = Town Meeting.

Clabaugh / Rozycki School Priorities Indicator (CRiSPI)

These sentence completions explore your priorities concerning schools and teachers. There are no right or wrong answers. Please rate each of the three possible endings for each incomplete sentence. No duplicate ratings or ties, please.

1st Choice endings are worth 3 points; 2nd Choice = 2 points; 3rd Choice = 1 point

1. School policy should focus chiefly on . . .

- A. consensus and participant involvement. TM
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- B. instructional effectiveness and efficiency. F
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- C. values and proper behavior. T
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice

2. The very best schools first insist on . . .

- A. structure and discipline. T
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- B. shared decision making. TM
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- C. measurable outcomes. F
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice

3. Educators should pay especially close attention to student . . .

- A. conduct and character. T
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- B. knowledge and skills. F
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- C. interests and talents. TM
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice

4. The school I would least like to attend would . . .

- A. disregard individual interests. TM
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- B. tolerate ineffective instruction. F
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- C. ignore improper conduct. T
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice

5. In outstanding schools . . .

- A. research guides practice and outcomes matter. F
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- B. everyone has a voice and personal interests matter. TM
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice
- C. perennial values are emphasized and tradition matters. T
☐ 1st Choice ☐ 2nd Choice ☐ 3rd Choice

The CRiSPI and a Personal Data Form (gathering information on such matters as type of school attended, self-reported degree of religiosity, self-reported degree of conservatism or liberalism) were administered to a random sample of 214 undergraduate and graduate students at La Salle University in Philadelphia and Radford University in Virginia. Results were tallied, flawed surveys discarded, and the results statistically analyzed. Here are the findings:

- The CRiSPI sentence completions clustered into three types as intended.
- The Town Meeting was the most popular ideal type, the Temple was second, and the Factory last.
- Compared to public school graduates, Catholic school graduates more strongly preferred the Temple.
- The greater the degree of self-reported religiosity, the greater the preference for the Temple.
- The greater the degree of self-reported political conservatism the greater the preference for the Temple.
- The greater the degree of self-reported political liberalism the greater the preference for the Town Meeting.

The most significant finding, however, was this. A factor analysis revealed just two dynamics at work in the survey: the school as Factory and the school as Temple. The Town Meeting was revealed to be a Temple subtype.

Originally, the values of the Temple and Town Meeting were thought to probably be too dissimilar to combine into one Temple ideal type. Compare a traditional spit and polish military school, for instance, with a free school where a council of students and teachers decide matters via group process and democratic dialogue. Both emphasize values, but the values are starkly different.

Perhaps more ably worded CRiSPI sentence completions could restore the Town Meeting to separate vitality. But for now the school as Town Meeting must be considered a sub-type of the school as Temple. That leaves us with just two ideal types: the school as Factory and the school as Temple.

Using ideal types simplify and amplify school realities permits an uncommonly clear understanding of educational developments. Those pushing No Child Left Behind, Race to the Top, charter schools,

for profit school management and the ever-growing movement to hold teachers “accountable” just want to make America more “competitive” and help kids escape “failing schools.” Maybe so, but they also are dramatically increasing the dominance of the school as Factory at the expense of the school as Temple.

These “reformers” probably have no idea that their actions are promoting this transformation. They are politicians, after all, not visionaries. But that does not alter the negative consequences of their actions.

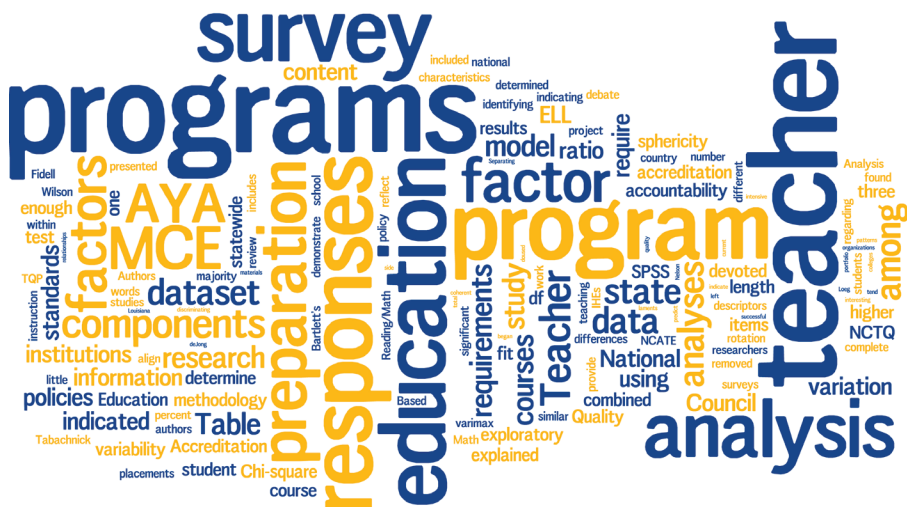
The school as Temple has its own problems—not least of which is possible abuse of power and a related absence of leadership accountability. We have also seen that the Temple can be further broken down into two sub-types.

A strong school requires that a dynamic tension be maintained between the Factory and the Temple. Traditions die, conduct becomes a concern only when production is impeded, values are neglected and teachers are malformed into mere technicians when the Temple is weakened excessively. The Temple, on the other hand, can ossify and lose touch with reality. For educators, the trick is to strike the right balance between these two ideal types.

Notes

1. Robert Ahart, “District residents outline qualities for superintendent,” *Springfield Sun*, Montgomery County, Pa., p. 1, date of publication unavailable.
2. For a classic treatment of the school as factory see Raymond E. Callahan, *Education and the Cult of Efficiency*, University of Chicago Press, 1964.
3. Callahan, *Education and the Cult of Efficiency*.
4. It’s interesting to speculate whether there are degenerate ideal types. A degenerate factory ideal type school, for example, might maintain industrial style procedures long after its actual efficiency has vanished. A degenerate Temple might retain ceremonial forms, but without the ethical underpinnings and moral tone. A degenerate Town Meeting might maintain the fiction of democratic community amidst hopelessly gridlocked partisan bickering.
5. Edward Rozycki, Ed.D., is a retired Associate Professor of Education at Widener University and founding publisher of the newfoundations.com website,

Gary K. Clabaugh is a Professor of Education at La Salle University and a founding partner of the popular education website, newfoundations.com.



Evaluating Teacher Preparation Programs: What Not to Do

M. Suzanne Franco and Martha S. Hendricks

Abstract

The National Council on Teacher Quality intends to rate all teacher education programs in the country and publish its findings in *US News and World Report*. The rating is to be determined by a survey of program characteristics. This paper presents the state-wide results of a multi-state, collaboratively developed survey of teacher education program characteristics. Exploratory factor analysis ($n=86$) determined that the programs were more similar than not. Reasons for the similarity could lie in accreditation requirements. The authors warn that implementation of the proposed NCTQ methodology will have similar results. Evaluation of teacher education programs should include observational, interview and other contextualized data.

Introduction

Evaluating teacher preparation programs has been part of the national debate on improving the American education system since the 1983 release of *A Nation At Risk*. Wilson and Young (2005), panel

members who reviewed the research on accountability in teacher education for Cochran-Smith and Zeichner's book (2005) *Studying Teacher Education*, concluded "given the impassioned debates around accountability in teacher education, it is both surprising and troubling that there is so little relevant empirical research" (p. 616).

Within the last year, the debate has been doused with fuel by the National Council on Teacher Quality (NCTQ), an organization that pronounced on its website, "It's never been done. We're going to do it" (National Council on Teacher Quality, 2010). The NCTQ intends to rate all teacher education programs in the country and publish its findings in *US News and World Report*. The rating is to be determined by a survey that seeks information about such criteria as the number of professional education courses offered, policies for student teaching, and surveys of alumni. Not surprisingly many institutions of higher education with teacher education programs are refusing to participate, citing issues with methodology, specifically the "input" or survey model of accountability, which the National Council for Accreditation of Teacher Education (NCATE) rejected over a decade ago (National Council for Accreditation of Teacher Education, 2008). Collecting data about the program characteristics does not necessarily capture the quality of a teacher education program.

The authors of this study have additional reasons for rejecting NCTQ's proposed methodology. Based on a statewide study we conducted, we found that the survey method of identifying policies and practices does not work for discriminating among teacher education programs. The Teacher Quality Partnership (TQP), a statewide collaborative of all institutions of higher education with teacher preparation programs, state agencies, and other educational associations, intended to determine the impact of teacher education on student achievement through a comprehensive, longitudinal study (Authors, 2010). In 2007 TQP was in partnership with researchers in New York, Florida and Louisiana, all of which were implementing statewide teacher education studies. A collaboratively designed survey gathered information about program structure, content requirements, and field experiences. Although TQP was unable to continue with the multi-state project, we did administer the survey to Adolescent/Young Adult (AYA) Math and Middle Childhood Educator (MCE) Reading/Math Programs in the research state. Teacher preparation programs for middle school grades were the focus. The survey is included in the technical report (Authors, 2010).

Data Analysis

Forty-three (43) public and private IHEs responded to the online survey about AYA Math and MCE Reading/Math programs. Some notable program descriptors are included in Table 1.

Do teacher preparation program course requirements include?	Percent Reporting YES
General course sequence for education (aside from General Ed)	93%
Content knowledge completed in the College of Arts and Sciences.	80%
Designated course in special education.	85%
Designated course in technology.	80%
Designated course in diversity.	76%
Designated course on classroom management.	40%
Designated course in assessment.	40%
Designated course about English Language Learning (ELL).	10%

Table 1: Notable Descriptors about Teacher Preparation Course Requirements

Regarding student teaching requirements among the 43 responding IHEs, 95% require a capstone project, and of those, 40% require a portfolio. Ninety percent have a mid-point benchmark; 90% require students to complete placements at more than one school setting; and 80% require students to complete placements in more than one grade level.

Given the variety of responses received, we elected to implement exploratory factor analysis to determine if there were sets of responses that naturally grouped teacher preparation programs together. In other words, are there patterns of correlation among the responses? Factor analysis is appropriate when researchers are trying to determine which items reflect coherent subsets (Tabachnick and Fidell, 2001).

Analysis of the program survey data began with data cleaning. From the fifty-five survey questions on the survey, 45 items were selected for factor analysis study. A review of the responses indicated that 24 had little to no variability (i.e., semester/quarter; graduate/undergraduate; program length; program type; content source for instruction; descriptions of entry/exit and minimum requirements, etc.). This left 19 factors for the factor analysis, using the 86 total responses (43 AYA and 43 MCE). The ratio of factors to responses is 19:86, or approximately 1:4. This ratio is on the low side for employing factor analysis, but does not negate the use of the procedure to explore the items' relationships (Tabachnick and Fidell, 2001).

Using SPSS, exploratory factor analysis with 19 factors and 86 responses was completed on three groups of data: MCE and AYA

combined, MCE only, and AYA only. Table 2 displays the results of MCE and the AYA program responses (n = 86). The resultant components using varimax rotation explained 66% of the variation in the dataset. The Bartlett's test of sphericity (Chi-square = 318.947, df = 171, p = .0000) indicated that the dataset does fit the model. The SPSS output identifying the components for all three analyses can be obtained from the authors.

Component	% of Variance	Cumulative %
Math Requirements	12.262	12.262
Diversity	10.562	22.824
ELL Requirements	9.831	32.655
Program Characteristics	9.714	42.369
Math/ELL Entry	8.801	51.169
Math/ELL Exit	7.883	59.052
Student Teaching	7.435	66.487

Table 2: Components for the MCE and AYA responses combined

Each component in this analysis explained between 12% and 7% of the variation, indicating that none of the components was particularly indicative of variation among the combined AYA and MCE preparation program responses.

Separating the AYA and MCE responses provided the following constructs using SPSS and the varimax rotation. For the AYA only analysis (n = 43), all factors related to English Language Learners (ELL) content areas were removed for the factor analysis, reducing the number of factors used in the analysis to 13 (Table 3). The 1:3 ratio of factors to responses is less than desirable for factor analysis, but not contraindicated for exploratory analyses. Bartlett's test of sphericity (Chi-square = 158.75, df = 78, p = .000) indicated that the dataset does fit the model.

Component	% of Variance	Cumulative %
Diversity	17	17
Math Requirements	16.6	33.6
Math Program Specifics	15	48.6
Student Teaching	11.5	60.1
Unexplained construct	11	71.1

Table 3: Components for AYA-Math responses (ELL responses removed)

Since the MCE license includes math and ELL, the ELL factors were not removed for the MCE analysis. There were 18 factors and $n = 43$. The 1:2 ratio of factors to responses failed to provide enough information for factor analyses: the Barlett's test of sphericity (Chi-square = 171.01, $df = 153$, $sig. = .15$) indicated that the dataset does not fit the model. There are not enough cases for each factor to have confidence in the model presented in Table 4.

Component	% of Variance	Cumulative %
Entry Math or ELL	13.2	13.2
Diversity	12.5	25.7
Math Program specifics	10.3	36.0
Program Characteristics	10.2	46.2
ELL Program specifics	9.7	55.9
Math/ELL Exit	8.0	63.9
Student Teaching	7.8	71.7

Table 4: Components for MCE program responses

The fact that no major components represented a substantially higher proportion of the variability in any of the three analyses presented here indicates that state teacher preparation programs are relatively homogeneous based on survey responses. The majority of institutions locate content course work in arts and sciences colleges; clinical assessments are standardized. It is interesting to note that length of the program did not load in any of the significant components; in other words, there was not enough variability in program length across the dataset to be significant.

To demonstrate differences captured in the survey, Table 5 includes information about program differences regarding devoted courses within programs. One hundred percent of the programs have a devoted course to learning development, whereas 40% have a devoted course to assessment. It is suspected that the reason there are fewer programs with dedicated courses in Assessment, Classroom Management, or ELL is that such instruction is integrated or embedded into other courses within the program (Harper and deJong, 2009; Mahon, Bryant, Brown, and Kim, 2010; Nelson, 2006).

Devoted Courses in Programs?	Yes	No
Learning Development	100%	
Assessment	40%	60%
Technology	80%	20%
Diversity	77%	23%
Special Education	85%	15%
ELL	10%	90%
Classroom Management	40%	60%

Table 5: Devoted Courses in Teacher Preparation Programs

Discussion and Conclusions

The analyses demonstrate that the 2007–2008 survey responses regarding Ohio teacher preparation program elements were not different. The components identified in the factor analysis of the program descriptors were easily explained, indicating that the dataset did discriminate well among the programs. However, no components accounted for a majority of the variation among programs; the programs were more alike than they were different.

Possible explanations could exist because of national accreditation standards and state policy requirements. State policies and accreditation protocols provide standards for the conceptual framework of a teacher preparation program. As a result, programs align courses to address the standards. Accreditation organizations such as NCATE or TEAC review the teacher preparation programs to ascertain that all standards are being met. Allington (2005) laments that accreditation requirements tend to “homogenize teacher preparation” (p. 199); these analyses support his concerns. Moreover, Goodlad (as cited in Wilson and Youngs, 2005), in an intensive study of 29 teacher education institutions, found heads of programs to be resigned to accepting that their curricula would be largely determined by state policy. The results reflect that the teacher education programs in the research state do align with the state’s policies; survey responses indicate they are very similar.

Based on this research, we predict that if the NCTQ is successful in gathering data on all teacher preparation programs in the country using the current survey methodology, the results will be similar to ours. Such studies as the one documented by Boyd, Grossman, Langford, Loeg, and Wyckoff (2009), which combined data collected

from surveys, documents, and other materials with interviews, observations, and other contextualized sources, is much more revealing about the qualities and characteristics of effective teacher education programs.

We would like to thank our funders:

Carnegie Corporation of New York, Joyce Foundation, Proctor and Gamble Fund, Martha Holden Jennings, George Gund Foundation, Ohio Board of Regents, Ohio Department of Education, Bank One, and Student Loan Association.

References

- Allington, R. (2005). Ignoring the policy makers to improve teacher preparation. *Journal of Teacher Education*, 56(3), 199-204. doi: 10.1177/0022487105275845
- Boyd, D., Grossman, P. L., Lankford, H., Loeg, S., and Wyckoff, J. (2009). *Teacher preparation and student achievement. Educational Evaluation and Policy Analysis*, 31(4), 416-440.
- Cochran-Smith, M., and Zeichner, K. M. (Eds.). (2005). *Studying teacher education: The report of the AERA panel on research and teacher education*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Harper, C. A., and de Jong, E. J. (2009). English language teacher expertise: The elephant in the room. *Language and Education* 23(2), 137-151.
- Authors. (2010). Teacher quality partnership novice teacher studies technical report NTS 10-01. Research Brief 2007-2009. ERIC (ED511661).
- Mahon, J., Bryant, B., Brown, B., and Kim, M. (2010). Using Second Life to enhance classroom management practice in teacher education, *Educational Media International*, doi: 10.1080/09523987.2010.492677
- National Commission of Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Retrieved from <http://www2.ed.gov/pubs/NatAtRisk/title.html>
- National Council of Teacher Quality. (2010). National review of education schools. Retrieved from <http://www.nctq.org/edschoolreports/national/>
- National Council for Accreditation of Teacher Education. (2008). Professional standards for the accreditation of teacher preparation institutions. National Council for Accreditation of Teacher Education, Washington, D.C. 2008.
- Nelson, B. (2006). On your mark, get set, wait! Are your teacher candidates prepared to embed assistive technology in teaching and learning? *College Student Journal*, 40(3), 485-94.

- Tabachnick, B. G., and Fidell, L. S. (2001). *Using Multivariate Statistics*. Boston: Allyn and Bacon.
- Wilson, S., and Youngs, P. (2005). Research in accountability processes in teacher education. In M. Cochran-Smith and Kenneth M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 591–643). Mahway, N.J.: Lawrence Erlbaum Associates.
- Wilson, S. M., Floden, R. E., and Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps and recommendations: A research report prepared for the U. S. Department of Education*. Seattle: Center for the Study of Teaching and Policy.
- Zeichner, K. M., and Conklin, H. G. (2005). Teacher education programs. In M. Cochran-Smith and Kenneth M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 645–736). Mahway, N.J.: Lawrence Erlbaum Associates.

M. Suzanne Franco is an associate professor in the Department of Educational Leadership at Wright State University, Dayton, Ohio.

Martha S. Hendricks is an assistant professor of education at the Wilmington College of Ohio, Wilmington, Ohio.



Higher Education Mission Statements: How Vital, How Vacuous?

by Bridget Bowers

Some of us have perhaps grown tired of mission statements. We've seen them everywhere it seems: in business advertising and correspondence, academic brochures and annual reports, and even in our church bulletins. Perhaps you know a family that has one too. Sometimes catchy and succinct, and sometimes not, these proclamations of vision and goals can either inspire us or make us feel cynical and weary.

—Van Gaalen, 2000

THE 1970S AND 1980S featured a “furor over mission statements” in corporate America; a buzz that “inevitably, but belatedly, reached the academy” (Morphew and Hartley, 2006). However, a study of liberal arts colleges and their missions by Delucchi (1997) found that mission statements still were a “neglected” aspect of higher education in the late 1990s.

More recently, however, unprecedented scrutiny of America’s colleges and universities has arisen. Margaret Spellings’s Commission on Higher Education is the most noteworthy. Similarly, accreditation associations have begun evaluating institutional focus on mission and goals as a part of their approval process.

Today’s schools must not only create a mission statement, but provide evidence of its achievement. The question remains, however, if this hand-me-down from business practice simply linguistic “smoke and mirrors,” or if it is actually having a positive impact on organizational practices and planning (Bartkus et al., 2000, 1).

Some experts think the impact is quite positive. Kotler and Murphy describe institutional mission as “an invisible hand that guides a college or university’s diverse personnel to work independently and yet collectively toward the realization of the organization’s goals” (1981, 245). Similarly minded experts in higher education (and organizational planning generally) maintain that mission statements are a crucial part of any organization in terms of maintaining legitimacy, effectiveness, and strategic planning processes.

Less optimistic appraisers, however, question whether mission statements are anything more than “a collection of stock phrases that are either excessively vague, unrealistically aspirational or both” (Morphew and Hartley, 2006).

Whichever side one chooses, the fact remains that mission statements now are commonplace in higher education, and much time and effort goes into their development and/or modification. What is more, strategic planning is “predicated on their formulation” (Morphew and Hartley, 2006). The purpose of this paper is to begin examining the literature in reference to the role of mission and/or vision in higher education.

Background

In the business world, a "company's mission and vision statements have long been regarded as a way to express a company's 'corporate character' or 'personality' (Chun, 2001, 317; Bartkus et al., 2000). Often several years are spent on their development and honing. Mission statements are often argued to "aid in strategy formulation and implementation."

Ireland and Hitt (1992) define the purpose of mission statements as "publicly declaring] the purpose, goals, products, markets, and philosophical views of the organization." Other theorists maintain that the mission statement should combine purpose with vision, while still others argue that "obligations to stakeholders" [in the case of higher education its students, faculty, staff] and notions of "competitive advantage" should be included. Many theorists claim that a mission or vision statement does more than just define the company but unifies employees, leadership, and stakeholders around a certain task (Tarnow, 2001).

In Mazza's research the role of mission statements as 'legitimizing' is examined (1999). According to Mazza, mission statements are "institutional communication tools, increasingly popular within large and small organizations, diffused among profit and not profit organizations, and made fashionable by consulting firms and business periodicals" (Mazza 1999, 3). Also of importance, however, is Mazza's observation that mission statements serve the purpose of legitimization" (Mazza 1999, 3).

The language used in said mission statements arguably has a key role in organizational identity. However, empirical evidence is mixed on whether mission statements truly provide guiding principles or merely consist of rhetorical, broad, catch-all phrasing. Bartkus, Glassman and McAfee, in "Mission Statements: Are They Smoke and Mirrors?" describe the problems that arise in the use of mission statements (2000). They claim that these general statements of purpose are: a) often redundant and b) either too narrow or too broad in their scope, either limiting change and adaptation or not providing substantial boundaries.

Mission Statements in Higher Education

According to Morphew and Hartley (2006), of the few studies that have empirically examined university mission statements, most researchers conclude that mission statements "ultimately fail to follow

through on, or convey, any noteworthy sense of an institution's current identity" (p. 457). Rather than focus and direction, they "provide means to an uncertain end"; "evoke an all-purpose purpose"; and "maximize institutional flexibility." Some experts even describe mission statements as "vague and vapid" (p. 458). If mission statements are purely rhetoric and/or a manipulation of semantics to please stakeholders, yet strategic planning is predicated upon them, where does this leave higher education in terms of organizational development?

Morphew and Hartley (2006), in their study of three hundred mission statements of U.S. four-year colleges and universities, examined the following question: "How do college and university mission statements differ in content, and are any differences reflective of recognized differences between institutional types?" (p. 460). Their examination yielded 118 unique mission elements across the institutions that were then broken down by Carnegie Classification. Interestingly, they found that across-the-board mission statements generally seemed to lack an aspirational quality or plan.

In the end, Morphew and Hartley (2006) seem to raise more questions than they answer as they explore the complexity of university mission statements: In the end, while there is evidence that mission statements are used to signal and symbolize, it also seems that institutions are using these documents to communicate their willingness to serve in terms that are both normative and politically apt (p. 469).

If one argues on the side of theorists who see missions as "symbolic artifacts," Morphew and Hartley's research seems to support the notion that missions do not speak to vision, but are used to "signal key constituencies that the institution in question shares these groups' values and goals" (p. 466). Further, they found that mission statements seem to reflect the environment of that institution, i.e., public versus private expectations and demands. Overall, "institutions include in their mission what their benefactors value" and these, then, reflect the various values across institutions (Morphew and Hartley, 2006, 466).

In Delucchi's study of liberal arts colleges and their "myth of uniqueness," he states "colleges and university in the United States make many claims about what they do for students" (Delucchi 1997, 415). He continues on to say "the current proliferation of popular college guides suggests that these claims are powerful symbolic devices that administrators and consumers think is meaningful" (Delucchi 1997, 415).

Interestingly, Delucchi's study, which examined the relationship between mission statements and curricula of 327 U.S. liberal arts colleges, found that inclusion of liberal arts in the mission statement was often (and ironically) coupled with a curriculum "dominated" by "professional education"—a seeming mismatch.

While the aforementioned research seems to point to the use of rhetoric and semantics in mission statements to dilute their specificity, Morphew and Hartley (2006) did note that what may seem variants of similar (and perhaps clichéd) terms are actually in some cases purposefully crafted. For example, the notion of "service" in a mission statement may refer to very different actualities across campuses, regional versus global etc. Additionally, the mention of liberal arts education may be either vague or specific depending on the external stakeholders of the institution and political underpinnings.

The question remains, however, why these notions of uniqueness are not further delineated in mission statements and how they come into being if a mission statement does not clearly delineate their definition.

The Influence of External Forces

In "Mission Possible?: Enabling Good Work In Higher Education," Berg, Csikszentmihalyi, and Nakamura (2003) speak to the stakeholders and external demands mentioned by Morphew and Hartley. They describe, however, a focus on creating a "balance of continuity and alterations in response to the changing environment," which, they state, does not mean "submitting willy-nilly to external forces. Rather it means that these forces be integrated with the institution's internal vision of a better reality" (Berg et al., 2003, 41).

Berg et al. (2003) states, "[T]o survive and prosper, and to enable the good work of the people who work there, a school must live up to a set of ethical guidelines embodied in a mission that expresses the spirit of a community. . . ." (p. 42).

Berg divides external forces influencing institutions into three categories:

- I. The field of higher education, which includes models, curriculum, and pedagogies in addition to accrediting agencies;
- II. External stakeholders, which includes alumni, funding sources, employers, prospective students, etc.;
- III. Social and cultural norms and their implications.

As far back as 1994 a survey by the Association of American Colleges showed that 80 percent of institutions were revising their mission statements. Berg et al. would likely regard these revisions as a consequence of changes within, and shifting balances among, these three categories.

When all these categories are in alignment, it is far easier to determine the direction and vision of an institution and to follow through strategically. In fact, Berg et al. maintain that "the mission is both a result of alignment and its cause" (Berg et al. 2003, 43). Disharmony or "misalignment" of forces and mission make it difficult to continue with "good work" and requires either a reaffirmation or revamping of that mission (Berg et al. 2003, 43).

Given ever-changing changing political, cultural, and social environments, Berg et al. recommend these six key question be asked with regard to any reexamination of mission:

1. What kind of school?
2. To whom are we responsible?
3. What are our strengths?
4. Whom should we hire?
5. Who shall lead?
6. When should we change? (pp. 45-46)

This process, they note, should not be one of semantics, rhetoric, and poetic statements, but an "organic" one that "involves the entire community," both internal and external. This inclusive process is crucial, they say, because having a clear and detailed, agreed-upon mission and vision can streamline programs, save precious financial resources that might be otherwise spent on misaligned goals, divide resources fairly, and save time and energy in planning.

Delucchi's previously mentioned study of mission statements and the changing curricula of liberal arts colleges provides an excellent example of external forces influencing mission statements. As Delucchi observes:

The nearly universal shift from liberal arts to professional education was a dominant strategy followed by hundreds of colleges in the past two decades. Nonetheless, for many colleges, the liberal arts claim remains. Why? Because retaining a liberal arts claim strengthens attachments and loyalties to the institution and its public image. (Delucchi 1997, 421)

In other words, so-called liberal arts colleges have not been “immune to pressures,” mostly financial and admission related, that are forcing a so-called mission-creep. However, they are reluctant to acknowledge it for fear of alienating crucial supporters.

Conclusions and Areas for Further Research

What is the role of mission and vision statements, particularly within higher education? Some think they are a compilation of slogans with an admixture of “ambiguous” buzzwords that can mean all sorts of contradictory things depending on the stakeholder’s point of view. In this case their role is little more than ceremonial—though it does provide senior staff with plenty of latitude to do what they want. Others say the “shallow agreement” and “systematic ambiguity” common to mission statements simply reveals how higher education is trying to be all things to all people in order to survive in the market place (Clabaugh and Rozycki 1999). There are others, however, who claim that these statements are an indispensable first step for action.

Research to date on the impact of higher education mission statements has produced mixed results. And there are many unanswered questions. For instance, how do similar mission statements end up serving vastly different institutional “personalities”? What role do mission statements play in institutional “impression management”? How do definitions vary among institutions for such commonplace phrases as “educational excellence,” and “service to others”?

With the spotlight now focused on higher education in terms of accreditation, performance, and output—particularly since the Spellings Commission report on higher education (2006)—questions such as these require further exploration.

Works Cited

- A Report of the Commission Appointed by Secretary of Education Margaret Spellings. (2006). *A Test of Leadership: Charting the Future of U.S. Higher Education*, Education Publications Center, U.S. Department of Education, 2006.
- Bartkus, B., Glassman, M, and McAfee, B. (2000). Mission statements: Are they smoke and mirrors? *Business Horizons*, 43(6), 23-29.
- Berg, G., Csikszentmihalyi, M., and Nakamura, J. (2003). Mission possible? Enabling good work in higher education. *Change*, 35(5), 40-47.

- Clabaugh, G., and Rozycki, E. (1999). Slogans in education. New Foundations. Retrieved January 23, 2008, from: <http://www.newfoundations.com/Slogans.html>
- Delucchi, M. (1997). "Liberal arts" colleges and the myth of uniqueness. *The Journal of Higher Education*, 68(4), 414-427. Retrieved November 17, 2006, from <http://pi'oquest.umi.com/pqdweb?index=0&did^>
- Ireland, R. D., and Hitt, M. A. (1992). Mission statements: Importance, challenge and recommendations for development. *Business Horizons*, 35, 34-42.
- Kotler, P., and P. E. Murphy. 1981. Strategic planning for higher education. *Journal of Higher Education* 52(5): 470-89.
- Mazza, C. (1999). *Claim, intent, and persuasion: Organizational legitimacy and the rhetoric of corporate mission statements*. Boston: Kluwer Academic Publishers.
- Morphew, C., and Hartley, M. (2006). Mission statements: A thematic analysis of rhetoric across institutional type. *The Journal of Higher Education*, 77(3), 456-471.
- Tarnow, E. (2001). A recipe for mission and vision statements. *IEEE Transactions on Professional Communication*, 44(2) Institutional Mission, 10.
- Van Gaalen, N. (2000). Development of a mission statement. Christian Engineering Education Conference. Retrieved on March 10, 2008, from <http://www.calvin.edu/academic/engineering/ces/ceec/vangaalen.htm>.

Bridget Bowers is Director, Alumni Career Services, at Georgetown University.

THE PRACTITIONER'S TOOLBOX

by Gary K. Clabaugh

When Mission, Vision, and Belief Statements Turn Nasty

Mission, vision, and belief statements seem harmless enough—just rhetorical fur balls coughed up by committees. In reality, though, they are far from harmless. As a matter of fact, these seemingly innocuous potions of purr words can easily turn nasty.

The Humpty Dumpty Principle

Consider this chosen at random mission statement from the Cumberland Valley School District in Mechanicsburg, Pennsylvania.

The Cumberland Valley School District, in collaboration with students, educators, parents and the community, is committed to developing 21st century learning and thinking skills through a rigorous, relevant, and comprehensive curriculum, while preparing students to be innovative, productive citizens in an interconnected world.

How can such vague, happy talk ever turn nasty? Lewis Carroll provides a clue. In *Through the Looking Glass*, a worldly Humpty Dumpty imperiously informs Alice: “When I use a word, it means just what I choose it to mean—neither more nor less.” Puzzled, Alice replies, “The question is whether you can make words mean so many different things.” Humpty Dumpty condescendingly responds, “The question is which is to be master—that’s all.”

The Deciders

How is that connected to mission, vision, or belief statements? The characteristic vagueness of their key terms permits senior staffers to

be Humpty Dumpty and decide which of the meanings that could be assigned are to be master. In the case of the Cumberland Valley mission statement, for example, they get to decide:

- What counts as "collaboration,"
- Which parents represent all "parents,"
- Which residents represent "the community,"
- What counts as "21st Century learning and thinking skills," and
- What counts as a "rigorous, relevant and comprehensive curriculum."

They even get to decide the characteristics of "a productive citizen in an interconnected world." How do they get this much power? Well, who else is in a position to successfully challenge the meanings they assign?

This is precisely how mission, vision, and belief statements can turn nasty. They provide senior school staff or board politicians with way too much arbitrary authority. Inadvertently granting them this much control can cause a world of harm—particularly if those in authority are stupid, irresponsible, venal, incompetent, or cowardly.

As a matter of fact, granting that much discretion to competent, responsible, and otherwise reliable leaders is unwise because some of them cannot be counted on to define things to their own disadvantage even though they should.

The general rule, then, is this. To the extent that mission, vision, or belief statements include vague key terms, to that extent decision-makers have too much power.

When Districts Come Alive

There are other dangerously vague terms sprinkled throughout the Cumberland Valley School District's belief statements. They say, for example, that the "District's belief statements . . . *guide its thinking and planning for the future.*"

School districts do not have beliefs, and they certainly do not "think" or "plan for the future." School districts are not living things. They are bureaucratic contrivances where beliefs, thoughts, and plans are the creatures of ordinary mortals. So the "beliefs" that "guide" here are not those of the Cumberland Valley School District,

but particular persons. And whenever it matters, key officials can assign their own meanings.

Superintendent Pollyanna

To illustrate this point let's consider another "Cumberland Valley" belief:

"Everyone can learn and succeed."

No, *everyone* cannot learn and succeed. That is why, for example, some public school students are categorized as "severely or profoundly impaired." But those exceptions aside, the key issues here are: learn what, succeed at what, and to what extent? Those questions are not addressed, much less resolved, by this belief claim. The result is that those in power can decide the answers to suit themselves.

It isn't hard to predict who will be blamed if a youngster's learning and/or success fails to match a key manager's or politician's self-defense and public relations needs. Odds are some hapless frontline educator will be strung up for this alleged failure.

High Expectations

How about this so-called "Cumberland Valley" belief?

"High expectations yield educational excellence."

This is rubbish. Even a person who regularly shines suede shoes realizes that high expectations do not necessarily yield excellent results. One can have a high expectation that a pet monkey will skillfully play a Chopin nocturne on the piano, for instance, but it is unlikely that this will result in an excellent performance.

Once again, the problem is the characteristic vagueness of the key terms—in this case "high expectations" and "excellence." Suppose a superintendent's definitions of "high expectations" and "excellence" are much too optimistic. Frontline educators still have to live with the resultant failures and blame.

Responsibility

Let's consider another "Cumberland Valley belief":

"Individuals are ultimately responsible for their own actions."

But some individuals clearly are not responsible for their own actions. That is precisely why "not guilty by reason of insanity" is a legal option.

It is also why civil law recognizes the importance of contributory negligence—when people are injured because of their own carelessness rather than the carelessness of another. Any worthwhile justice system also considers mitigating and extenuating circumstances before assigning responsibility to an individual.

How responsible are individuals, for instance, if superiors have dictated all, or at least some, of their actions? Suppose others choose a teacher's textbooks and other teaching materials. And further suppose that the lessons he or she teaches have been scripted and imposed from above. Is the teacher still responsible if those measures fail?

If "the district" believes that "individuals are ultimately responsible for their own actions," school officials need not bother with any of these niceties. They can just assign blame to whichever of their subordinates are least able to defend themselves.

Conclusion

Mission, vision, and belief statements can turn nasty whenever key words are sloganistically vague. Then those in power can decide on a definition most convenient for them.

The only way to check such arbitrary power is to insist that these statement's key terms are carefully defined. Unfortunately, it is virtually impossible to reach consensus on definitive definitions in a diverse community such as a school district or university. That is the chief reason that mission, vision, and belief statements are characteristically vague to begin with.

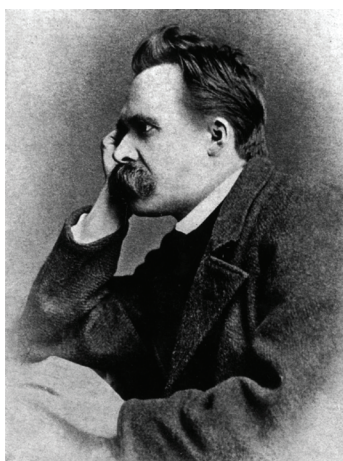
Allowing the meaning of key terms to remain fuzzy is typically the only way one can generate an admittedly superficial consensus. Those involved in the formulation process unwittingly assign their own meanings to key terms and, thus, can be brought to "agree."

Understanding the dangers of vague mission, vision, and belief statements is an essential aspect of educator self-defense. In the hands of frightened, cowardly, incompetent, or unprincipled power holders they can have really nasty consequences.

Besides, do we really need these things? What evidence is there that they in fact positively influence people's day-to-day, on the job behavior?

INTERVIEW WITH **THE DEAD**

Friedrich Nietzsche



Nietzsche, 1882

Interviewer: Herr Nietzsche, how long since you left us?

Nietzsche: Biologically, I died in the late summer of 1900; but my strength of mind died in early January 1889. I was in Turin and saw a coachman flogging a horse. I rushed to the beast and collapsed with my arms around its neck. They had to carry me home. That was the very moment when I lost the last vestiges of my sanity. All that remained was diving ever deeper into madness for ten long years. One should die proudly when it is no longer possible to live proudly. For me that proved impossible.

Interviewer: You must have suffered.

Nietzsche: Yes, although what really raises one's indignation against suffering is not suffering intrinsically, but the senselessness of it. We would like our suffering, and our entire existence, to have a point. But human life is inexplicable, and without meaning: a fool may decide its fate. Death and the stillness of death are the only things certain and common to all.

Interviewer: It was long thought that your madness and death were caused by syphilis. Medical experts now agree this was not the cause.

Nietzsche: Now that my enemies have to change their opinion, they will charge my account heavily for the inconvenience I have caused them.

Interviewer: Some experts think that your illness caused the extraordinary spurt of creativity you experienced shortly before your breakdown.

Nietzsche: Quite possibly it did; and that illustrates why you have to be careful in casting out your demon. You might exorcise the best thing within you.

Interviewer: The prime reason I requested this interview was to have you comment on the present state of education—particularly in the United States. Where would you like to begin?

Nietzsche: Let's begin with the elemental fact that there are two different types of people in the world: those who want to know, and those who want to believe. There is no point in trying to educate the latter, though educators keep trying.

Interviewer: But beliefs are the basis of faith.

Nietzsche: Faith is not wanting to know what the truth is.

Interviewer: What about the curriculum? Do you agree with what is being taught these days?

Nietzsche: Of course not. Haven't you noticed, for example, that when money runs short music is typically sacrificed? How ridiculous! Life without music is a mistake.

Love too has to be learned. Yet that isn't in the curriculum. And where one can no longer love, there one should pass by.

Then there is dancing. Dancing in all its forms cannot be excluded from the curriculum of all noble education; dancing with the feet, with ideas, with words, and need I add that one must also be able to dance with the pen?

None of that is really encouraged in school, which means school days are lost days. We all should consider every day lost on which we have not danced at least once. To dance is to be out of yourself—larger, more beautiful, and more powerful. That is power, that is glory on earth, and it is yours for the taking.

A real education facilitates that kind of taking and helps people regain the seriousness that they had as a child at play.

Then there is the penultimate goal of education, which should be to teach people how to fly. Anyone who would learn to fly one day must first learn to walk and run and climb and dance; one cannot fly into flying. Those you cannot teach to fly, teach to fall faster.

Interviewer: Don't schools at least strengthen community?

Nietzsche: No, they just strengthen tribalism, and the individual should always struggle to keep from being overwhelmed by the tribe.

Anyone who tries that will be lonely often, and sometimes frightened. But no price is too high to pay for the privilege of owning yourself.

Interviewer: Is that the kind of truth that kids should be exposed to?

Nietzsche: Truths are illusions that we have forgotten are illusions.

Interviewer: But earlier you said that believers do not want to know the truth, which presupposes that truth exists. Aren't you contradicting yourself?

Nietzsche: Only idiots fail to contradict themselves three times a day. Anyway, what is truth but a lie agreed upon or illusions that we have forgotten are illusions? That, my friend, is the *essential* truth.

Interviewer: Business people are frequently consulted on the way U.S. schools are run. Some business billionaires have even spent staggering sums to shape schooling to their liking. What do you have to say to them?

Nietzsche: I would say this. Business people—your business is your greatest prejudice. Diligent in business, but indolent in spirit.

Content with your inadequacy, and with the cloak of duty hung over this contentment. That is how you live, and that is how you want our children to live!

Interviewer: What about the way U.S. schools are organized and operated?

Nietzsche: Here is what I have to say about that:

- First, most schools are run by the State, and everything the State says is a lie. Everything it has it has stolen.
- Second, schools depend upon labels. First grade, second grade, emotionally disturbed, honors, and so forth. What labels us negates us.
- Third, schooling takes up too much of an individual's time. Whoever does not have two-thirds of the day for himself is a slave, whatever he may be: a statesman, a businessman, an official, or a student.
- Fourth, schools require doing what others tell you to do. Nothing destroys a person more quickly than to work, think, and feel without inner necessity, without any deep personal desire, without pleasure—as a mere automaton of duty. It is here that we learn to labor at our daily work more ardently and thoughtlessly than is necessary to sustain our life, because it is even more necessary not to have leisure to stop and think. Haste is universal because everyone is in flight from himself.
- Fifth, digressions, objections, delight in mockery, carefree mistrust are signs of vitality and health. Disobedience also is healthy—it is the nobility of slaves. Yet it is precisely those things that get you in trouble in school. He who obeys does not listen to himself!
- Sixth, scholarship is overrated. Behind a remarkable scholar we not infrequently find an average human being, and behind an average artist we often find a very remarkable human being. Art is essentially the affirmation, the blessing, and the deification of existence, and that is precisely why it is given short shrift in school.

Interviewer: From what you just said about government you obviously aren't a fan of public schools. What, then, do you make of the rapid growth of Christian schools?

Nietzsche: In truth, there was only one Christian and he died on the cross. What is more, in Christianity neither morality nor religion comes into contact with reality at any point. Finally, there is not enough love and goodness in the world to permit giving any of it away to imaginary beings.

Interviewer: For several decades politicians have been trying to reform schooling. What do you make of that?

Nietzsche: A politician divides mankind into two classes: tools and enemies. That is what their school reforms are based on. There are terrible people who, instead of solving a problem, bungle it and make it more difficult for all who come after. Whoever can't hit the nail on the head should, please, not hit at all. Do you think these would-be school reformers are hitting the nail on the head?

Interviewer: What of the chase after ever-higher standardized test scores?

Nietzsche: Chasing after test scores readies students to chase after money and material possessions. Living in a constant chase after gain compels people to expend their spirit to the point of exhaustion in continual pretense and overreaching and anticipating others. Virtue has come to consist of doing something in less time than someone else.

They remind me of frantic monkeys. Watch them clamber over one another and push one another into the mud. They all want to get to the throne: that is their madness—as if happiness sat on the throne. Often, mud sits on the throne. Mad they all appear to me, these over-ardent and clambering monkeys. Foul smells their idol, the cold monster. Foul, they smell to me altogether, these idolaters.

Interviewer: So much for materialism!

We are nearly out of time, so allow me this last bold question. How would you sum up your life?

Nietzsche: Was that a life? Well then, once more!

This "interview" was constructed of actual quotes taken from a variety of sources. While minor modifications were made to fit the venue, Nietzsche's thoughts and sentiments remain intact.

*I never considered a
difference of opinion
in politics, in religion,
in philosophy, as cause
for withdrawing from
a friend.*

—Thomas Jefferson

